Searching for beyond the Standard Model physics using direct and indirect methods at LHCb

Samuel C. P. Hall

High Energy Physics Group
Department of Physics
Imperial College London

A thesis to be submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

It is known that the Standard Model of particle physics is incomplete in its description of nature at a fundamental level. For example, the Standard Model can neither incorporate dark matter nor explain the matter dominated nature of the Universe. This thesis presents three analyses undertaken using data collected by the LHCb detector. Each analysis searches for indications of physics beyond the Standard Model in different decays of B mesons, using different techniques. Notably, two analyses look for indications of new physics using indirect methods, and one uses a direct approach.

The first analysis shows evidence for the rare decay $B^+ \to D_s^+ \phi$ with greater than $3\,\sigma$ significance; this also constitutes the first evidence for a fully-hadronic annihilation-type decay of a B^+ meson. A measurement of the branching fraction of the decay $B^+ \to D_s^+ \phi$ is seen to be higher than, but still compatible with, Standard Model predictions. The CP-asymmetry of the decay is also measured, and its value is precisely in line with the Standard Model expectations.

The second analysis claims the first observations of the decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$ which are both flavour changing neutral currents, forbidden at leading order in the Standard Model. Branching fractions of both these decays are measured, and for the high statistics channel $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ the differential branching fraction, as a function of the invariant mass squared of the dimuon system, is also presented.

These first two analyses both constitute indirect searches for physics beyond the scope of the Standard Model, where the observables are sensitive to contributions from new physics entering at loop-level. In contrast, the third analysis presented in this thesis involves the direct search for a new dark boson, χ, which is a messenger particle between a dark sector and the Standard Model particles. Using a frequentist technique, the dimuon component of candidates consistent with the decay $B^0 \to K^*(892)^0\mu^+\mu^-$ is searched for an excess indicative of $\chi \to \mu^+\mu^-$.

Declaration

The work presented in this thesis is the result of a work undertaken by the LHCb collaboration and myself between January 2012 and June 2015. Three analyses, published in Refs. [1–3], are described in this thesis. Significant proportions of all these analyses were done by the author. This thesis has not been submitted for any other qualification.

Samuel C. P. Hall
November, 2015

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work (Imperial College London PhD regulations).

Acknowledgements

I would like to begin by expressing sincere gratitude towards my supervisor, Prof. Andrey Goutvin, for the guidance and help he bestowed on me throughout my PhD. Other people who offered invaluable help and suffered me gladly included Mike Williams, Sebastian Neubert, Mitesh Patel, and Kostas Petridis. Without these individuals, this thesis would not have been possible.

I must also thank my pals, Sam Cunliffe, Robyn Lucas, Patrick Owen and Eluned Smith for being my islands of sanity while on long term attachment to CERN. They were also constant sources of entertainment throughout my PhD.

To list the normal, non-physicist, friends that have popped up and been of immeasurable value over the last few years would not be a little cheesy and, frankly, make a tedious read. You know who you are.

Lastly, I must thank my family, for their relentless encouragement, support, and unwavering belief in me throughout my life.

Contents

List of Figures viii

List of Tables xi

1 Introduction 1

2 The Standard Model and beyond 3
- 2.1 The Standard Model . . . 3
  - 2.1.1 The CKM matrix and Unitarity Triangle . . . 9
- 2.2 Physics beyond the Standard Model . . . 11
  - 2.2.1 Failures and inconsistencies of the Standard Model . . . 12
- 2.3 Operator Product Expansion . . . 16
- 2.4 Dealing with QCD . . . 17

3 The LHCb experiment 18
- 3.1 The LHC . . . 18
- 3.2 The LHCb detector . . . 18
  - 3.2.1 Tracking . . . 20
  - 3.2.2 RICH detectors . . . 23
  - 3.2.3 Calorimetry system . . . 25
  - 3.2.4 Muon systems . . . 27
- 3.3 Trigger . . . 29
- 3.4 Data, stripping, and simulation . . . 32

4 Multivariate selection techniques 34
- 4.1 The bagging algorithm . . . 36
- 4.2 The AdaBoost algorithm . . . 37
- 4.3 The uBoost algorithm . . . 38
- 4.4 Summary . . . 39

5 Search for the decay $B^+ \to D_s^+ \phi$ 40
- 5.1 Introduction . . . 40
  - 5.1.1 Other annihilation-type hadronic decays . . . 42

5.2 Data selection . . . 43
5.2.1 Suppression of combinatorial background . . . 45
5.2.2 Vetoes of D^+ and Λ_c^+ deacys . . . 47
5.3 Calculation of the branching fraction . . . 48
5.3.1 Mass fits . . . 50
5.3.2 Modelling components in the mass fit . . . 52
5.3.3 Constraining the mass fit . . . 54
5.3.4 Efficiency calculations . . . 55
5.4 Systematic uncertainties . . . 57
5.5 Direct *CP* asymmetry . . . 59
5.6 Summary . . . 60

6 Search for the decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$ 61
6.1 Introduction . . . 61
6.2 Selection . . . 65
6.2.1 Background contributions . . . 66
6.2.2 Multivariate selection . . . 68
6.3 Efficiency calculations . . . 70
6.4 Differential branching fraction of the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ . . . 74
6.4.1 Systematic uncertainties . . . 79
6.5 Branching fraction of the decay $B^+ \to \phi K^+\mu^+\mu^-$ . . . 80
6.5.1 Systematic uncertainties . . . 82
6.6 Summary . . . 83

7 Search for the decay of a dark sector particle $\chi \to \mu^+\mu^-$ in $B^0 \to K^{*0}\mu^+\mu^-$ 85
7.1 Introduction . . . 85
7.2 Analysis strategy . . . 89
7.2.1 Searching in the mass dimension . . . 89
7.2.2 Searching in the lifetime dimension . . . 90
7.2.3 Calculation of a global p-value . . . 91
7.2.4 Limit setting . . . 92
7.2.5 Resonant backgrounds . . . 93
7.3 Selection . . . 95
7.3.1 Preselection . . . 96
7.3.2 Other backgrounds from particle misidentification . . . 98
7.3.3 Possible contamination from the $X(1070)$ . . . 100
7.3.4 Combinatorial background and multivariate selection . . . 101
7.4 Evolution of parameters as a function of mass . . . 103
7.4.1 Mass resolution of the χ candidate . . . 103

7.4.2 Lifetime resolution of the χ candidate . . . 104
7.4.3 Parameterizing the efficiency of the χ selection . . . 106
7.5 Unblinding results . . . 109
7.5.1 Calculation of the p-value . . . 112
7.5.2 Systematic uncertainties . . . 112
7.6 Summary . . . 115
7.6.1 The published LHCb analysis . . . 115

8 Conclusions **118**

Glossary **121**

Bibliography **123**

A License **131**

List of Figures

2.1 Shape of the Higgs potential . . . 7
2.2 Schematic diagram of the Unitarity Triangle . . . 11
2.3 Unitarity triangle and current constraints . . . 14

3.1 Simulated production of $b\bar{b}$ pairs and the LHCb acceptance . . . 19
3.2 Diagram of the LHCb detector . . . 20
3.3 Diagram of the LHCb Vertex Locator . . . 21
3.4 PV resolution as a function of track multiplicity . . . 22
3.5 IP resolution as a function of momentum . . . 22
3.6 Diagram of the LHCb Tracker Turicensis . . . 23
3.7 Diagram of the LHCb tracking stations T1–3 . . . 24
3.8 Diagram of the LHCb RICH detectors . . . 25
3.9 Particle identification and Cherenkov angles . . . 26
3.10 Diagram of the LHCb hadron and electron calorimeters . . . 27
3.11 Side-on diagram of the LHCb muon system . . . 28
3.12 Front-on diagram of the LHCb muon system . . . 29
3.13 LHCb trigger sequence in 2012 . . . 30
3.14 Impact parameter resolution as a function of inverse transverse momentum comparing data and simulation . . . 33

4.1 An example single DT, taken from the analysis in Chap. 7, showing the root node and decendant nodes. At each node the cut value is shown along with the purity of the node. The variables Flight Distance (FD) and vertex χ^2 are defined in later chapters. Signal nodes have a higher purity, p, and are split to the left, such that candidates landing on leaf 1, 2, or 4(3, 5, or 6) are classified as belonging to signal-like(background-like) events. . . . 35

5.1 Feynman diagram for the decay $B^+ \to D_s^+ \phi$ . . . 40
5.2 Example distribution of $\mathrm{BDT}_{D_s^+} \times \mathrm{BDT}_{D^0}$ from candidates for the decay $B^+ \to D_s^+ \overline{D}^0$ . . . 45
5.3 Optimisation of BDT cut for selection of $B^+ \to D_s^+ \phi$ candidates . . . 47
5.4 Vetoing D_s^+ contamination from Λ_c^+ and D^+ decays . . . 48
5.5 Selected D_s^+ and ϕ candidates . . . 49

5.6 Fit to the normalisation channel $B^+ \to D_s^+ \overline{D}^0$ . . . 49
5.7 Simulated distribution of the helicity angle of the ϕ . . . 50
5.8 Contribution to the $B^+ \to D_s^+ \phi$ fit from the $B^+ \to D_s^{*+} \phi$ background . . . 52
5.9 Shapes of the $\overline{B}_s^0 \to D_s^+ K^{*0} K^-$ background . . . 53
5.10 Shapes of the $\overline{B}_s^0 \to D_s^{*+} K^{*0} K^-$ background . . . 53
5.11 Fits to $B^+ \to D_s^+ \phi$ data . . . 57
5.12 Feldman-Cousins coverage for $\mathcal{A}_{CP}$ . . . 60

6.1 Schematic Feynman diagrams for FCNCs . . . 61
6.2 Feynman diagrams for $B^+ \to K^+ \pi^+ \pi^- \mu^+ \mu^-$ and $B^+ \to \phi K^+ \mu^+ \mu^-$ . . . 63
6.3 Theoretical q^2 distribution for $B^+ \to K_1^+ \mu^+ \mu^-$ . . . 64
6.4 Charmonium vetoes is in $B^+ \to K^+ \pi^+ \pi^- \mu^+ \mu^-$ and $B^+ \to \phi K^+ \mu^+ \mu^-$ . . . 67
6.5 Backgrounds from misidentified charmonia . . . 67
6.6 Simulated background from cascade decays . . . 68
6.7 Comparison of track multiplicity between data and simulation . . . 71
6.8 Effect of resampling PID variables in simulation . . . 72
6.9 Corrections for `isMuon` and tracking efficiency in simulation . . . 73
6.10 Distributions of BDT response in data and simulation . . . 74
6.11 Efficiency as a function of q^2 for $B^+ \to K^+ \pi^+ \pi^- \mu^+ \mu^-$ and $B^+ \to \phi K^+ \mu^+ \mu^-$ 74
6.12 Fits to $B^+ \to J/\psi K^+ \pi^+ \pi^-$ and $B^+ \to \psi(2S) K^+$ candidates . . . 75
6.13 Fits to signal decay $B^+ \to K^+ \pi^+ \pi^- \mu^+ \mu^-$ . . . 77
6.14 Differential branching fractions of $B^+ \to K^+ \pi^+ \pi^- \mu^+ \mu^-$ . . . 78
6.15 Invariant mass distributions of $K^+ \pi^+ \pi^-$ in data . . . 78
6.16 Fits to $B^+ \to \phi K^+ \mu^+ \mu^-$ and $B^+ \to J/\psi \phi K^+$ candidates . . . 81
6.17 Invariant mass distributions of ϕK^+ in data . . . 82

7.1 Parameter space for a model including an inflaton . . . 87
7.2 Feynman diagram for the decay $B^0 \to K^{*0} \chi$ . . . 87
7.3 Decay rates for decays of the form $B \to K \chi$ . . . 89
7.4 Theoretical distributions of $c\bar{c}$ resonances interfering with a non-resonant component . . . 94
7.5 Track definitions in the LHCb detector . . . 96
7.6 Effect of the double misidentification veto . . . 98
7.7 Contamination from the decay $\Lambda_b^0 \to p K^- \mu^+ \mu^-$ . . . 100
7.8 Invariant mass of the $\mu^+ \mu^-$ distribution under the $K^+ K^-$ mass hypotheses 100
7.9 Analysis of the $K_S^0 \to \pi^+ \pi^-$ background under the $K^+ K^-$ mass hypothesis 101
7.10 Signal efficiency of a uBDT cut as a function of mass and lifetime . . . 102
7.11 Evolution of mass resolution with m_χ . . . 103
7.12 Fits to the lifetime resolution for simulated χs . . . 104
7.13 Opening angle for low mass dark bosons . . . 105

7.14 Opening angle of the $K^+\pi^-$ system close to threshold . . . 105
7.15 Parameterisation of lifetime resolution as a function of χ mass . . . 106
7.16 Decay time distributions for various dark boson masses . . . 107
7.17 Parameterisation of decay time acceptance as a function of mass . . . 108
7.18 Relative efficiency map . . . 108
7.19 Fit to the normalisation channel, $B^0 \to K^{*0}\mu^+\mu^-$ for $1.1 < q^2 < 6.0\,\text{GeV}^2$ 109
7.20 Efficiency ratios . . . 110
7.21 Estimation of combinatorial background contribution . . . 111
7.22 dimuon mass distribution and p-value calculation . . . 113
7.23 Validation of efficiency ratio at long lifetimes . . . 114
7.24 Projected sensitivity in an inflaton search . . . 116
7.25 Projected sensitivity in an inflaton search . . . 117

List of Tables

2.1 Fundamental, force-mediating, gauge bosons . . . 4
2.2 Fundamental fermions of matter . . . 5

3.1 Conditions to satisfy the `isMuon` criteria . . . 29
3.2 Level one trigger thresholds . . . 31

5.1 Selection criteria applied to $B^+ \to D_s^+ \phi$ candidates . . . 44
5.2 Training variables for the D_s^+ and ϕ BDTs . . . 46
5.3 Definitions of fit regions . . . 51
5.4 Constraints applied to the fit to $B^+ \to D_s^+ \phi$ data . . . 55
5.5 Efficiencies for calculating $\mathcal{B}\big(B^+ \to D_s^+ \phi\big)$ . . . 56
5.6 Systematic uncertainties . . . 59

6.1 Selection criteria applied to candidates in the LHC beauty (LHCb) stripping phase. The definitions of χ^2 values can be found in the text in this section or in Sec. 5.2. . . . 66
6.2 BDT training variables . . . 69
6.3 Differential branching fractions for the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ . . . 78
6.4 Systematic uncertainties on the differential branching fraction of $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ . . . 80
6.5 Systematic uncertainties on the branching fraction of $B^+ \to \phi K^+\mu^+\mu^-$ . . . 83

7.1 Summary of dark boson portals . . . 86
7.2 Total branching fractions for decays with the final state $B^0 \to K^{*0}\mu^+\mu^-$ . . . 93
7.3 Summary of veto regions for $q\bar{q}$ states . . . 95
7.4 Samples of simulated $B^0 \to K^{*0}\chi$ generated for the analysis . . . 95
7.5 Stripping criteria applied to the dark boson candidates . . . 97
7.6 Preselection criteria . . . 98
7.7 Veto conditions to suppress backgrounds from misidentifications . . . 99

"Fiction is obliged to stick to possibilities. Truth isn't."
Mark Twain

Chapter 1

Introduction

Scientific endeavour has long been a quest to understand the Universe in which we live. From the vastness of space, to the incomprehensible world of quantum mechanics, increased understanding comes in leaps and bounds in a feedback loop of scientific discovery. Still the most ancient questions remain unanswered, including: why does the Universe the way it is? The subject of high energy particle physics sits at the forefront of modern physics and seeks to answer, arguably, this most basic question.

Our current understanding of the interactions governing the behaviour of fundamental particles is encapsulated in the Standard Model (SM). This is a remarkably successful theory which results in no significant deviations between theory and experimental measurements. That being said, there are a plethora of phenomena which remain completely unexplained by the SM. The SM is introduced fully in Chap. 2.

The Large Hadron Collider (LHC) is currently the world's most energetic particle accelerator. It is host to a number of experiments probing high energy proton-proton (pp) collisions for evidence of New Physics (NP) which might go towards resolving the inadequacies of the SM. Located on the LHC are four main particle detectors. One of these is the LHCb detector, which specialises in precision measurements of beauty flavoured hadrons in the hunt for NP, this is introduced in Chap. 3.

High Energy Physics (HEP) experiments can broadly be split into two camps: ones that search for NP *directly* and those that search *indirectly*. These names are somewhat misleading, because both can be clear indicators of the existence of NP. Direct searches rely on observing a particle that is not encompassed by the SM, and seeing its immediate manifestation, such as a peak in an invariant mass spectrum. Indirect searches use precisely measured observables and compare them with theoretical predictions; discrepancies between the two will indicate either some flaw in the theoretical understanding, or new and unknown physics processes. This can be an advantageous search method, because virtual particles can contribute to processes and therefore indirect searches have access to much higher

mass scales than direct ones.

This thesis describes three analyses: in Chap. 5, Chap. 6, and Chap. 7, performed on data collected by the LHCb detector. The first two of these analyses search for indications of NP indirectly, and the third is a direct search for a new particle.

Chapter 2

The Standard Model and beyond

"Before beginning a Hunt, it is wise to ask someone what you are looking for before you begin looking for it."

Winnie the Pooh, A.A. Milne

This thesis contains the work undertaken in three analyses; each of which concerns a different area of interest in HEP. The following chapter aims to motivate each analysis in turn after introducing the Standard Model of particle physics.

Firstly, the formulation of the SM will be outlined, with particular detail paid to the flavour sector. Various successes of the SM will then be discussed before going on to identify its shortcomings using arguments from both experiment and theory. These shortcomings will then be used to motivate the three analyses: a search for the decay $B^+ \to D_s^+ \phi$ (Chap. 5); a search for the two related decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$ (Chap. 6); a search for dark sector particles in $B^0 \to K^*(892)^0$ (Chap. 7). Theory specific to each of these analyses will be detailed in the relevant chapter.

2.1 The Standard Model

The behaviour of fundamental particles and forces are described by the SM of particle physics [4–10][1]. The SM was concocted in the 1970s, when the Higgs mechanism was incorporated into Glashow's electroweak theory by Salam and Weinberg. The theory prescribes a treatment as to how fundamental particles interact via three of the four fundamental forces, namely: the strong, weak and electromagnetic forces.

Mathematically, the SM is a locally gauge invariant quantum field theory. It inhabits a space-time with a global Poincaré symmetry that obeys a local $SU_C(3) \otimes SU_L(2) \otimes U_Y(1)$

[1] This section was written with reference to Refs. [11–13].

Force	Particle	Mass	Charge
Electromagnetic	γ	0	0
Weak	$W^{\pm}$	80.4 GeV	± 1
	Z	90.2 GeV	0
Strong	G^a	0	0

Table 2.1: Fundamental, force-mediating, gauge bosons in the SM. All values are taken from Ref. [14].

symmetry[2]. The $SU_L(2) \otimes U_Y(1)$ gauge group contains the electroweak formalism, and the $SU_C(3)$ group contains that of the strong force. Generators for each group correspond to the vector bosons which mediate interactions — the $3 + 1$ electroweak gauge bosons (Z, $W^{\pm}$ and the photon), and the eight gluons of the strong force. A summary of gauge bosons in the SM is given in Table 2.1.

Symmetries are fundamental to the dynamics of particle physics. It was shown by Emmy Noether that for each symmetry in the action of a physical system there is a conserved quantity [15]. This is Noether's theorem. For any sensible theory of physics, it is necessary that the laws remain the same independent of time and space, these symmetries lead to the conservations of momentum and energy, respectively. Discrete symmetries of important are Time reversal (T) and Parity (P), so that physics occurs the same regardless of the direction of time, and under reflections in space.

Spin-$\frac{1}{2}$ particles in the SM are known as fermions, and are described by spinor fields, ψ, and obey the Dirac equation:

$$\big(i\gamma^{\mu}\partial_{\mu} - m\big)\psi = 0. \tag{2.1}$$

The fermions of the SM can be broadly categorised into those which couple to the strong force, *quarks*, and those which do not, *leptons*. There are six quarks: up, down, charm, strange, top, and bottom (u, d, c, s, t and b); and six leptons: the electronic leptons (ν_e, e^-), the muonic leptons (ν_{μ}, μ^-), and the tauonic leptons (ν_{τ}, τ^-). Due to the properties of the strong force, quarks can only be observed as colour-neutral bound states, usually these are *mesons* (quark-antiquark bound states) and *baryons* (bound states of three quarks). All fermions are organised into pairs forming three generations. For each fermion there is a corresponding antiparticle with the same mass and opposite charge — where charge is the conserved quantity resulting from the global electroweak $U_Y(1)$ gauge invariance, by Noether's theorem. A summary of all fermions and some of their properties is given in Table 2.2. There is also a single scalar field in the SM: the Higgs boson.

[2] The convention of natural units is used throughout. Other conventions are that the indices μ and ν are used for four vectors, and generators for the $SU_L(2)$ and $SU_C(2)$ are denoted by $\{i, j, k\}$ and $\{a, b, c\}$, respectively.

		Leptons			Quarks	
Generation	Particle	Mass	Charge	Particle	Mass	Charge
1	e^+	0.511 MeV	-1	u	2.3 MeV	$+\frac{2}{3}$
	ν_e		0	d	4.8 MeV	$-\frac{1}{3}$
2	μ^+	0.105 GeV	-1	c	1.275 GeV	$+\frac{2}{3}$
	ν_μ		0	s	95.0 MeV	$-\frac{1}{3}$
3	τ^+	1.777 GeV	-1	t	173 GeV	$+\frac{2}{3}$
	ν_τ		0	b	4.18 GeV	$-\frac{1}{3}$

Table 2.2: Fundamental fermions of matter. In the SM each has a corresponding anti-particle of opposite charge. All values are taken from Ref. [14], where the quark masses are calculated according to the $\overline{\text{MS}}$ renormalisation scheme. It is known that neutrinos have non-zero masses, but their precise masses are unknown.

The SM Lagrangian can be expressed as a sum of components:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{V}} + \mathcal{L}_{\ell} + \mathcal{L}_{q} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yuk}}. \tag{2.2}$$

These Lagrangians describe the: strong force interactions between colour carrying particles in the theory of Quantum Chromodynamics (QCD) ($\mathcal{L}_{\text{QCD}}$); weak vector self-interactions ($\mathcal{L}_{\text{V}}$); electroweak behaviour of leptons ($\mathcal{L}_{\ell}$); electroweak behaviour of quarks ($\mathcal{L}_{q}$); Higgs interaction ($\mathcal{L}_{\text{Higgs}}$); and Yukawa couplings ($\mathcal{L}_{\text{Yuk}}$).

As well as the discrete symmetries of P and T, there is also the Charge conjugation (C) symmetry. The violation of these symmetries are of fundamental interest to modern particle physics. Violation of the combined Charge-Parity (CP) symmetry, and flavour, arise in the Cabibbo-Kobayashi-Maskawa (CKM) matrix of the SM, which emerges after the Higgs mechanism breaks the local electroweak symmetry. The important terms for this are: $\mathcal{L}_{q}$, $\mathcal{L}_{\text{Higgs}}$, and $\mathcal{L}_{\text{Yuk}}$ from Eq. 2.2.

The Higgs doublet, Φ, is defined to be

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.3}$$

where each ϕ_i is a real field. The Lagrangian of the Higgs field is:

$$\begin{aligned} \mathcal{L}_{\text{Higgs}} &= \big(D_\mu \Phi\big)^\dagger \big(D^\mu \Phi\big) - V\big(\Phi\big) \\ &= \big(D_\mu \Phi\big)^\dagger \big(D^\mu \Phi\big) - \mu^2 \big(\Phi^\dagger \Phi\big) + \lambda \big(\Phi^\dagger \Phi\big)^2, \end{aligned} \tag{2.4}$$

where μ and λ are constants, and D_μ is the covariant derivative. Figure 2.1 shows that taking $\mu^2 < 0$ and $\lambda > 0$ shifts the ground state of the vacuum of $V(\Phi)$ away from zero. When the system collapses in to the ground state a direction is chosen, breaking the

symmetry of the system. The amount by which the ground state shifts with respect to the origin is

$$v = \sqrt{\frac{\mu^2}{\lambda}}. \tag{2.5}$$

At this point the Higgs field gets a Vacuum Expectation Value (VEV) of $\langle\phi\rangle = v/\sqrt{2}$. The direction of the VEV from the origin is arbitrary, but the choice of

$$\begin{aligned} \langle 0|\phi_1|0\rangle = \langle 0|\phi_2|0\rangle = \langle 0|\phi_4|0\rangle = 0 \\ \langle 0|\phi_3|0\rangle = v, \end{aligned} \tag{2.6}$$

is convenient, and changes the Higgs doublet in Eq. 2.3 to

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \eta_1 + i\eta_2 \\ v + i\eta_4 \end{pmatrix}. \tag{2.7}$$

Here, η_1, η_2, and η_4, are Goldstone bosons which become the longitudinal components of the weak vector bosons when an appropriate gauge is chosen Thus, Φ simplifies to

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \tag{2.8}$$

where H is the physical Higgs boson. Inserting Eq. 2.8 into Eq. 2.4 gives:

$$\mathcal{L}_{\text{Higgs}} = \frac{1}{2}\big(\partial_\mu H\big)\big(\partial^\mu H\big) + \frac{m_H^2}{2}H^2 + \left(m_W^2 W_\mu^+ W^{-\mu} + \frac{m_Z^2}{2} Z_\mu Z^\mu\right)\left(1 + \frac{H}{v}\right)^2. \tag{2.9}$$

Thus, there is Spontaneous Symmetry Breaking (SSB) of the local $U_Y(1)$ gauge group. A result of this is that weak gauge bosons become massive, while photons remain massless: as is consistent with observations.

All fermions (excepting neutrinos) also acquire mass after SSB. The Dirac mass term for a chiral field has the form:

$$\mathcal{L}_{\text{mass}} = -m_\psi\big(\overline{\psi}_R\psi_L + \overline{\psi}_L\psi_R\big), \tag{2.10}$$

but the left- and right-handed fields (ψ_L and ψ_R) have different $U_Y(1)$ charges and so transform differently. Therefore, using Eq. 2.10 to give masses to fermions would break gauge invariance. Instead, masses are generated through the Yukawa couplings, which describe interactions between all fermionic fields and the Higgs doublet. This can be written:

$$\mathcal{L}_{\text{Yuk}} = \sum_\ell \left(\mathcal{L}_{\text{Yuk}}^\ell\right) + \mathcal{L}_{\text{Yuk}}^q, \tag{2.11}$$

where terms encapsulate lepton and quark interactions, respectively. Each lepton term

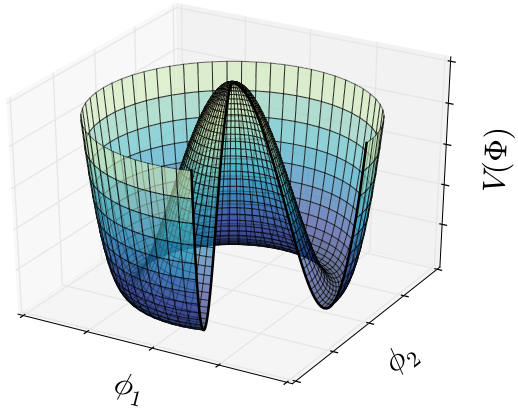


Figure 2.1: The shape of the Higgs potential, $V(\Phi)$, for the simple case of $\Phi = \phi_1 + i\phi_2$ and $\mu^2 < 0$ and $\lambda > 0$. Spontaneous symmetry breaking occurs when the vacuum settles in a minimum, and this choice of direction breaks the symmetry of the gauge. A section of the potential is not shown, to convey the shape of the potential.

describes the interaction between the Higgs boson and the chiral fields ℓ_R and the spinor

$$\chi_L = \begin{pmatrix} \nu_L \\ \ell_L \end{pmatrix}, \tag{2.12}$$

via

$$\mathcal{L}^{\ell}_{\text{Yuk}} = -g_\ell\big(\overline{\chi}_L \Phi \ell_R + \bar{\ell}_R \Phi^\dagger \chi_L\big), \tag{2.13}$$

where each g_ℓ is a coupling constant. After SSB the Lagrangian becomes

$$\mathcal{L}^{\ell}_{\text{Yuk}} = -m_\ell\big(\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L\big)\left(1 + \frac{H}{v}\right) \tag{2.14}$$

and the charged lepton masses

$$m_\ell = \frac{v}{\sqrt{2}} g_\ell \tag{2.15}$$

are dependent on the fundamental parameters g_ℓ and v.

Yukawa interactions for quarks involve the right-handed chiral operators of the up- and down-type quarks, q^i_R for $q \in \{u, d\}$, and the left-handed doublet

$$Q^i_L = \begin{pmatrix} u^i_L \\ d^i_L \end{pmatrix}. \tag{2.16}$$

Before SSB, the Yukawa Lagrangian is

$$\mathcal{L}^q_{\text{Yuk}} = -y^u_{ij}\overline{Q}^i_L \Phi u^j_R - y^d_{ij}\overline{Q}^i_L \widetilde{\Phi} d^j_R + \text{ h.c.} \tag{2.17}$$

where $\widetilde{\Phi}_i = \varepsilon_{ij}\Phi_k$, there is an implicit sum over the generations i and j, and the shorthand h.c. denotes hermitian conjugate. The coupling constants, y^q, are 3×3 matrices characterizing Yukawa coupling strengths between generations. After SSB, $\mathcal{L}^q_{\text{Yuk}}$ becomes:

$$\mathcal{L}^q_{\text{Yuk}} = -\frac{v}{\sqrt{2}}\left(y^u_{ij}\overline{u}^i_L u^j_R + y^d_{ij}\overline{d}^i_L d^j_R + \text{ h.c.}\right)\left(1 + \frac{H}{v}\right). \tag{2.18}$$

Similar to lepton masses, in Eq. 2.15, quark masses are defined as

$$\begin{aligned} m^u_{ij} &= \frac{v}{\sqrt{2}} y^u_{ij} \\ m^d_{ij} &= \frac{v}{\sqrt{2}} y^d_{ij}. \end{aligned} \tag{2.19}$$

Thus far all calculations have been in the flavour basis. But, it is now more convenient to change to the mass basis in which the matrices $m^{u,d}$ are diagonal. This is done using the rotation matrices V_L and V_R such that

$$\begin{aligned} m^{u\prime}_{il} &= \left(V^{u\dagger}_L\right)_{ij} m^u_{jk} \left(V^u_R\right)_{kl} \\ m^{d\prime}_{il} &= \left(V^{d\dagger}_L\right)_{ij} m^d_{jk} \left(V^d_R\right)_{kl}. \end{aligned} \tag{2.20}$$

The addition of a prime distinguishes the mass basis from the flavour basis. This transformation is exactly equivalent to transforming the up- and down-type chiral quark fields according to:

$$\begin{aligned} q'_L &= \left(V^q_L\right) q_L \\ q'_R &= \left(V^q_R\right) q_R. \end{aligned} \tag{2.21}$$

Applying these transformations to all parts of $\mathcal{L}_{\text{SM}}$ leaves the majority of it unchanged, since $V^{q\dagger}_L V^q_L = V^{q\dagger}_R V^q_R = \mathbb{1}$ by definition. However, this is not the case for $\mathcal{L}_q$.

The Lagrangian $\mathcal{L}_q$ can be decomposed into $\mathcal{L}^{\text{NC}}_q + \mathcal{L}^{\text{CC}}_q$, where the superscripts denote Neutral Current (NC) and charged current Charged Current (CC) components. The NC part of the Lagrangian characterises interactions between quarks and the neutral electroweak vector bosons, while the CC part involves the interactions of quarks wiht the charged $W^{\pm}$ bosons. After changing to the mass basis, $\mathcal{L}_{\text{NC}}$ remains unchanged, whereas

$\mathcal{L}_{\rm CC}$ transforms as:

$$\begin{aligned}\mathcal{L}_q^{\rm CC} &= i\frac{g}{2}\gamma^\mu\left[\bar{u}_L d_L W_\mu^+ + \bar{d}_L u_L W_\mu^-\right] \\ &= i\frac{g}{2}\gamma^\mu\left[\bar{u}'_L\left(V_{uL}V_{dL}^\dagger\right)d'_L W_\mu^+ + \bar{d}'_L\left(V_{dL}V_{uL}^\dagger\right)u'_L W_\mu^-\right] \\ &= i\frac{g}{2}\gamma^\mu\left[\mathbf{V}\bar{u}'_L d'_L W_\mu^+ + \mathbf{V}^\dagger\bar{d}'_L u'_L W_\mu^-\right]. \end{aligned} \tag{2.22}$$

In the final step the matrix $\mathbf{V} = V_{uL}V_{dL}^\dagger$ is defined. This is known as the CKM matrix, and parameterises the couplings between up- and down-type quarks in charged weak currents.

2.1.1 The CKM matrix and Unitarity Triangle

The CKM matrix is defined as:

$$\mathbf{V} = \left(V_{uL}V_{dL}^\dagger\right) = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{2.23}$$

where each $|V_{ij}|$ parameterises the probability of an up-type quark, of generation i, transitioning to down-type quark, of generation j, in a weak interaction. In the SM, it is assumed that the total charged current couplings of up- to down-type quarks is the same as down- to up-type. This assumption means that the CKM matrix is unitary, $\mathbf{V}^\dagger\mathbf{V} = \mathbb{1}$, and therefore it contains only four physical parameters: three angles (θ_{12}, θ_{13} and θ_{23}) and one complex phase (δ). In fact, the observation of *CP* Violation (*CPV*) in kaon mixing [16] led to the prediction of a third generation before its discovery, precisely because a 3×3 matrix is the smallest necessary for a phase to enter a unitary matrix.

There are many ways of representing the CKM matrix. One way is as a product of three rotation matrices, one of which contains the complex phase, this is known as the *standard* parameterisation:

$$\mathbf{V} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{2.24}$$

where s_{ij} and c_{ij} denote $\sin\theta_{ij}$ and $\cos\theta_{ij}$, respectively. A convenient simplification is the *Wolfenstein* parameterisation, which is obtained by defining

$$\begin{aligned}\sin\theta_{12} &= \lambda, \\ \sin\theta_{23} &= A\lambda^2,\end{aligned}$$

and

$$e^{-i\delta}\sin\theta_{13} = A\lambda^3(\rho - i\eta), \tag{2.25}$$

which results in

$$\mathbf{V} \simeq \begin{pmatrix} 1-\frac{1}{2}\lambda & \lambda & A\lambda^3\big(\rho - i\eta\big) \\ -\lambda & 1-\lambda^2 & A\lambda^2 \\ A\lambda^3\big(1-\rho-i\eta\big) & -A\lambda^2 & 1 \end{pmatrix}. \tag{2.26}$$

The values of the Wolfenstein parameters A and λ are [14]:

$$\lambda = 0.22537 \pm 0.00061,$$

and

$$A = 0.814\,^{+0.023}_{-0.024}.$$

So, $\mathbf{V}$ cannot be diagonal because $A \neq 0$ and $\lambda \neq 0$. Therefore Flavour Changing Neutral Currents (FCNCs) are allowed in the SM. However, the diagonal elements are still the largest, meaning that intra-generational interactions are preferred in the weak interaction and the CKM matrix exhibits a strongly hierarchic structure.

It has been asserted that the CKM matrix is unitary, and therefore a unitarity condition can be expressed as $V^*_{\alpha\beta}V_{\beta\gamma} = \delta_{\alpha\gamma}$. When $\delta_{\alpha\gamma} = 0$, this condition gives six equations of the form:

$$\sum_{\beta=1}^{3} V^*_{\alpha\beta}V_{\beta\gamma} = 0, \qquad \sum_{\beta=1}^{3} V_{\alpha\beta}V^*_{\beta\gamma} = 0, \qquad \alpha \neq \gamma; \tag{2.27}$$

each mapping a closed triangle on the complex plane. Two of these triangles have all sides of similar length $\big(\mathcal{O}(\lambda^3)\big)$; one of these is is known as *the* Unitarity Triangle (UT) and is defined by

$$1 + \frac{V_{ud}V^*_{ub}}{V_{cd}V^*_{cb}} + \frac{V_{td}V^*_{tb}}{V_{cd}V^*_{cb}} = 0, \tag{2.28}$$

where the length of the base has been normalised to unity. The apex of the UT is at

$$\begin{aligned} \bar{\rho} + i\bar{\eta} &= \big(1 - \tfrac{1}{2}\lambda^2\big)\big(\rho + i\eta\big) \\ &= \frac{V_{ud}V^*_{ub}}{V_{cd}V^*_{cb}}, \end{aligned} \tag{2.29}$$

and forms the angles

$$\alpha = \arg\left(-\frac{V_{td}V^*_{tb}}{V_{ud}V^*_{ub}}\right), \qquad \beta = \arg\left(-\frac{V_{cd}V^*_{cb}}{V_{td}V^*_{tb}}\right), \qquad \gamma = \arg\left(-\frac{V_{ud}V^*_{ub}}{V_{cd}V^*_{cb}}\right). \tag{2.30}$$

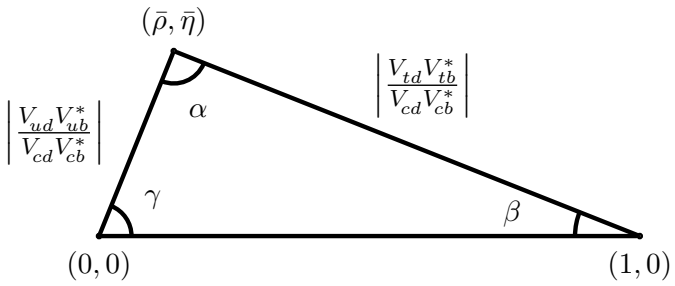


Figure 2.2: Schematic diagram of the UT given in Eq. 2.28 on the complex plane, where the base has been normalised to unit length. The angles α, β, and γ are defined in Eq 2.30.

which define phase differences between edges. Figure 2.2 depicts a schematic diagram of the UT.

Each CKM matrix element is a fundamental parameter in the SM. It is therefore important to measure each of them; particularly because the CKM matrix holds all the information about flavour violation and *CPV* allowed within the framework of the quark sector of the SM. All the measurements relating to the CKM matrix can be shown in the UT.

2.2 Physics beyond the Standard Model

For a long time the completion of the SM was reliant on the discovery of the Higgs boson. Finally, in 2012, the CMS [17] and ATLAS [18] collaborations observed a Higgs-like boson with a mass of $m_H \simeq 125\,\text{GeV}$ [19, 20]. This final piece of the picture has made the SM a remarkably robust theory in which, aside from the major failings described in the next section, there are currently no predictions deviating significantly from experimental observations. Indeed, the theory of Quantum Electrodynamics (QED) — which describes interactions between photons and charged particles in the SM — is one of the most accurate theories yet constructed. The coupling constant in QED is the fine structure constant, α, which has been measured experimentally to be [21]

$$\alpha_{\text{exp}}^{-1} = 137.035\,999\,074\,(44),$$

and predicted theoretically to be [22]

$$\alpha_{\text{th}}^{-1} = 137.035\,999\,073\,(35).$$

These measurements have precisions which are better than one part per billion, and the extent to which they agree is testament to our understanding of QED interactions.

Despite the countless successes of the SM, there are a plethora of indications — both experimental and theoretical — that additional physics exists, Beyond the Standard Model

(BSM).

2.2.1 Failures and inconsistencies of the Standard Model

There are some phenomena that have been observed experimentally which cannot be explained by the SM. The flavour-change of neutrinos in flavour space mean that they must have mass; this is not accounted for the SM framework. Neither are the observations of the Baryon Asymmetry of the Universe (BAU), or Dark Matter (DM).

The SM cannot reconcile the matter-antimatter asymmetry observed in the Universe today. The hypothesised process which caused this asymmetry is known as baryogenesis, the minimum requirements for this process are outlined in the three Sakharov conditions [23]. The first, most obvious, criteria is that baryogenesis must violate baryon number. The second Sakharov condition is that both C and CP are violated. Lastly, baryogenesis must occur out of thermal equilibrium. While the SM does contain some CPV, it is approximately ten orders of magnitude [24, 25] too small to explain the BAU. In Chap. 5 a measurement of the CP-asymmetry of the decay $B^+ \to D_s^+ \phi$ is made in an effort to find CPV processes not predicted by the SM that would go towards explaining the BAU.

It is well known that the vast majority of mass in the Universe is unaccounted for. Luminous matter totals only $\sim$4.9 % of the Universe [14, 26], and the rest is known only as DM ($\sim$26.8 %) and dark energy ($\sim$68.3 %). Dark Matter is an old and well motivated concept with the first evidence found in 1939 by H. W. Babcock in the form of flat galactic rotation curves [27, 28]. Since then, further credence to the existence of DM has come from corroborating evidence supplied by, for example, gravitational lensing around the Bullet Cluster [29] and the Cosmic Microwave Background [30].

Observations of DM are used to motivate NP models which include *dark sectors*. A dark sector is a name for a particle, or group of particles, which is gauged under a different gauge group to the SM particles and therefore cannot interact with them directly. There are a plethora of such models, but generally dark particles can only interact with the SM via weakly interacting messenger particles, which could be either vector or scalar. In generality, these are known as *Dark Bosons*. This thesis documents a search for a dark boson in the dimuon spectrum of $B^0 \to K^{*0}\mu^+\mu^-$ in Chap. 7.

Some excitement was caused by a hint of a dark sector messenger particle from the Hyper-CP experiment [31], which observes three $\Sigma^+ \to p\mu^+\mu^-$ events which survive a stringent selection. These three events also peak in the invariant mass of the dimuon pair. The narrowness of this peak is indicative of a two body decay, consistent with $\Sigma^+ \to pP^0$ and the subsequent decay of the NP particle via $P^0 \to \mu^+\mu^-$, where $m_{P^0} = 214.3 \pm 0.5\,\text{MeV}$ [32]. The P^0 could be a supersymmetric Goldstone boson, or a dark boson from many other theories.

Supersymmetry (SUSY) is a theory which imposes a symmetry relating fermions to bosons, and naturally supplies a DM candidate in the shape of the lightest supersymmetric particle. The lightest SUSY particle is stable, in most models, because the symmetry of R-parity is assumed to be conserved. The definition of R-parity involves the baryon number, lepton number, and spin of a particle; the upshot is that for SM(SUSY) particles R-parity is 1(-1). The Higgs sector in SUSY comprises four Higgs doublets; two are spin-0 and two are spin-$\frac{1}{2}$, and then there are two each for $Y = \pm\frac{1}{2}$. After SUSY is broken there are five Higgs physical scalar particle; two are CP-even (h^0, H^0); one is a CP-odd scalar (A^0) and two are charged ($H^\pm$). Unfortunately, masses of the super-particles are unconstrained, and could be anywhere between a few TeV and the Planck scale.

Particle dynamics can be affected by massive NP particles, like those in SUSY, in lower order processes because at this level virtual particles can contribute. FCNCs are heavily suppressed in the SM. Firstly, they are forbidden at tree-level; secondly, loop-level diagrams are suppressed by factors coming from the CKM matrix. These rare, and background-suppressed processes provide ideal environments in which to search for BSM physics, since new massive off-shell particles can contribute to the loops and cause significant deviations from SM expectations. Chapter 6 details an observation of a high statistics FCNC decay, which could be used for future NP searches.

Theoretical shortcomings of the SM include: its inability to incorporate gravity at the quantum scale and the existence of dark energy. But theoretical arguments are often less tangible, and rather subjective, revolving around the idea of *naturalness*. Naturalness is a concept whereby a theory is deemed to be natural, or more plausible, if it has few free parameters, all of which have a magnitude $\mathcal{O}(1)$. The SM is not a natural theory: having a total of 18 free parameters, 13 of which reside in the flavour sector. Other unnatural features of the flavour sector of the SM are that the CKM matrix is strongly hierarchic, and quark masses vary by four orders of magnitude.

Of all the fundamental parameters in the CKM matrix, V_{ub} is known to the lowest precision, and it is therefore important to accurately measure it. This parameter is particularly interesting because drives the largest discrepancies in global fits to the UT [33]. A determination of V_{ub} can be made using inclusive and exclusive measurements of semi-leptonic $B \to X_u \ell \bar{\nu}_\ell$ decays; where X_u is some meson containing a u quark. Inclusive measurements are made difficult by large $B \to X_c \ell \bar{\nu}_\ell$ backgrounds, while exclusive semi-leptonic modes suffer from theoretical uncertainties. Both these methods are well established, and both rely on non-perturbative QCD calculations. Current inclusive and exclusive measurements of V_{ub} are [14, 34]:

$$|V_{ub}|_{\text{excl}} = \left(3.28 \pm 0.29\right) \times 10^{-3}$$
$$|V_{ub}|_{\text{incl}} = \left(4.41\,^{+0.21}_{-0.23}\right) \times 10^{-3}.$$

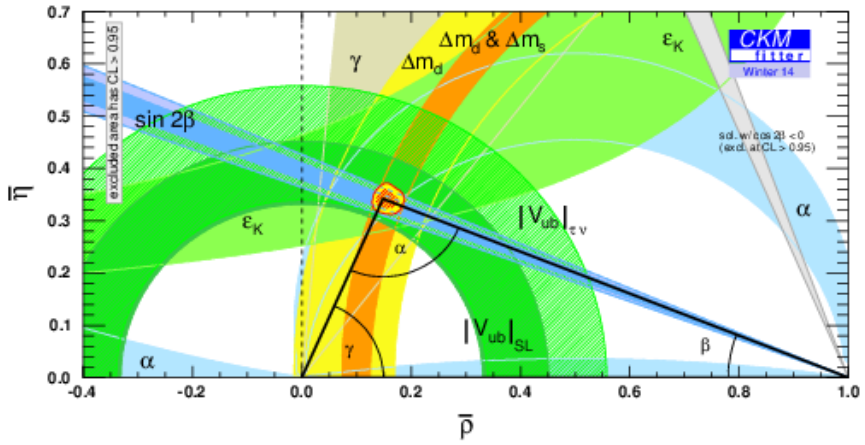


Figure 2.3: Diagram of the UT with coloured bands indicating various constraints on side lengths, angles and position of the apex, which is taken from the CKMfitter group in Ref. [33]. The constraints on V_{ub} from the combination of inclusive and exclusive modes ($|V_{ub}|_{\rm SL}$) is given separately to a value obtained using $\mathcal{B}\,(B^+ \to \tau^+\nu_\tau)$, ($|V_{ub}|_{\tau\nu}$).

Currently, there is no explanation for this discrepancy between inclusive and exclusive measurements. An exclusive measurement of $|V_{ub}| = (3.27 \pm 0.23) \times 10^{-3}$ [35] has also been made by the LHCb collaboration using the baryonic decay $\Lambda_b^0 \to p\mu^-\overline{\nu}_\mu$.

Another method to access the CKM matrix parameter $|V_{ub}|$ is via the annihilation-type decay $B^+ \to \tau^+\nu_\tau$. Measurements from both the BaBar and Belle experiments of $\mathcal{B}\big(B^+ \to \tau^+\nu_\tau\big)$ [36,37] suffer from small statistics, but are found to be in better agreement with values of $|V_{ub}|$ determined using inclusive measurements than exclusive. The lack of statistics for this channel is because with the missing energy from neutrinos, and the finite lifetime of the τ^+, it is very unlikely to form a vertex of good quality. Searching for the decay $B^+ \to \tau^+\nu_\tau$ is not viable at LHCb; instead, decays of the same topologies can be searched for. The decay $B^+ \to D_s^+\phi$ is also an annihilation-type decay in which V_{ub} appears in the amplitude; an analysis of this decay is described in Chap. 5.

Current measurements of angles and side lengths of the UT, from Ref. [33], are shown in Fig. 2.3. This figure also shows global V_{ub} measurements from the semi-leptonic and $B^+ \to \tau^+\nu_\tau$ modes alongside one another.

Unnatural NP models with parameters that differ wildly in magnitude tend to lead to parameters or processes that must cancel to absurdly high precision in order to agree with experimental observations. These precise cancellations are known as *fine tuning*. In the SM, quantum loop corrections to the Higgs mass are of the order 10^{19} for $m_H \simeq 125\,\text{GeV}$ [19,20]. This means that the cancellations required to result in a Higgs mass comparable to the masses of the weak vector bosons must be exact to 17 orders of magnitude. This instance of fine tuning is known as the *hierarchy problem*. A solution for the hierarchy problem is to introduce NP particles, whose contributions to loop level processes reduce the magnitude of

fine tuning required to a level deemed acceptable. SUSY immediately solves the hierarchy problem because for every SM particle that contributes to the Higgs mass, a SUSY particle also contributes, but with the opposite sign. It should be noted, however, that while SUSY does solve a number of problems, it is not natural since it has far more free parameters than the SM.

Fine tuning also appears in QCD. A gauge invariant term that can be added to $\mathcal{L}_{\text{QCD}}$ is

$$\mathcal{L}^{\theta}_{\text{QCD}} = \theta \frac{g^2}{32\pi^2} G^a_{\mu\nu} \widetilde{G}^{\mu\nu}_a, \tag{2.31}$$

where θ is a phase, and g is a constant [38]. The operator $G_{\mu\nu}$ is the gluon field strength tensor, and

$$\widetilde{G}^{\mu\nu}_a = \frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} G^{\rho\sigma}_a. \tag{2.32}$$

Interactions in $\mathcal{L}^{\theta}_{\text{QCD}}$ would conserve C symmetry, but violate both P and T conjugation [38]. Such symmetry violations contradict the observed properties of the strong force, so $\mathcal{L}_{\text{QCD}}$ must either be absent, or heavily suppressed. Bounds placed on the value of the neutron dipole moment, $|d_n| < 2.9 \times 10^{-26}\,\text{ecm}$ (at 90% CL) [39] leads to the constraint that $\theta < 10^{-19}$ [40], when *a priori* it could be in the range $0 < \theta < 2\pi$. This occurrence of fine tuning is referred to as the *strong CP problem*.

Despite the evidence for BSM physics and the list of problems that must be solved, its precise manifestation is unknown. There are numerous theories concerning NP scenarios which seek to solve various problems.

A solution to the strong *CP* problem is to introduce an additional chiral $U(1)$ symmetry, which is known as a Peccei-Quinn (PQ) symmetry [38]. Breaking the PQ symmetry leads to θ, in Eq. 2.31, becoming a field with quanta known as *axions*. These axions could be the messenger particle between a dark and visible sector [38].

Some searches look directly for evidence of NP; this is the case for the analysis detailed in Chap. 7, where a new particle, χ, is searched for in the dimuon invariant mass spectrum of $B^0 \to K^*(892)^0 \mu^+\mu^-$ consistent with $\chi \to \mu^+\mu^-$. This is sensitive to a range of models which predict a light particle with a mass in the range $2m_\mu \lesssim m_\chi \lesssim 4000\,\text{MeV}$, such as the axion model. It is also sensitive to the P^0 that was hinted at by the Hyper-*CP* experiment [32].

Instead of counting on NP to behave in an expected way, it is possible to search in a model independent manner by exploring general physics couplings. To do this it is useful to introduce the Operator Product Expansion (OPE) [41].

2.3 Operator Product Expansion

When describing macroscopic physical systems, it is frequently necessary to simplify the situation by making assumptions about the distance scales involved. One would not, for example, dream of using quantum mechanics to model a ball colliding with a wall despite the treatment being far more proper than a Newtonian approach. An Effective Field Theory (EFT) works in an equivalent way in particle physics by decoupling the short- and long-range interactions and treating them separately. Contributions from particles with very high mass, much greater than some pre-defined energy scale, are suppressed. These simplifications are advantageous as they allow processes to be modelled at a scale relevant to the particles involved without complications from other scales.

Creating an effective field theory for particle physics begins by defining an energy scale, Λ, which separates the long and short range interactions. For the case of a process involving a decaying b quark, with initial state $|i\rangle$ and final state $|f\rangle$, the energies are of order m_b. In the full treatment of the SM, contributions from the t quark and weak bosons — which all have masses $\mathcal{O}(100\,\mathrm{GeV})$ — must be accounted for. Therefore, an appropriate choice for Λ is $\sim m_W$. Heavy fields above Λ are then integrated out and are parameterised by complex numbers, known as Wilson coefficients, C_i. The remaining physics is encapsulated in the long distances operators, $\mathcal{O}_i$, each having its own gauge group defining a particular type of process. Transition matrix elements for the interaction in the effective Hamiltonian are

$$\langle f|\mathcal{H}_{\mathrm{eff}}|i\rangle = \sum_j C_j(\Lambda)\langle f|\mathcal{O}_j^{(d)}|i\rangle\Big|_\Lambda, \tag{2.33}$$

which is simply a weighted sum over the long distance matrix elements $\langle f|\mathcal{O}_j|i\rangle$ which operates in dimension d.

The sum in Eq. **??** runs over an infinite number of operators — this is clearly impractical, and matters are simplified by extracting factors of Λ^{-1} from the Wilson coefficients making each one. This modifies the effective Hamiltonian to be a sum over dimensions:

$$\langle f|\mathcal{H}_{\mathrm{eff}}|i\rangle = \sum_d \frac{1}{\Lambda^d}\sum_j c_j^{(d)}\langle f|\mathcal{O}_j^{(d)}|i\rangle\Big|_\Lambda. \tag{2.34}$$

This is entirely general, and one can calculate the Wilson coefficients to a high degree of accuracy — using perturbative methods — in the SM and many BSM extensions. Calculating the long distance operators is more challenging, however, because they contain QCD processes.

This approach leads to an effective Hamiltonian, with Wilson coefficients which are entirely independent of the underlying physical processes and can be calculated to a good degree of accuracy in a range of physics models. In this way very different observables can be used

to make measurements of Wilson coefficients and compared independently of the actual process. Measurements can be used for predictions of processes, and to check the validity of NP models, enabling experiments to favour or disfavour entire classes of physics BSM.

2.4 Dealing with QCD

For the branching fraction measurements discussed in this thesis, theoretical uncertainties from QCD make predictions difficult[3]. QCD describes the interactions of colour charged particles (quarks and gluons), and exhibits two peculiarities: confinement, and asymptotic freedom. Confinement means that over long distances ($\sim$1 fm) the interaction strength of the strong force does not weaken — unlike all other known forces. This means that as a quark is separated from others, there is enough energy in the gluon field to create new quark-antiquark pairs, where the resulting bound states always have net zero colour charge. Free quarks cannot be seen over macroscopic distances, and are instead observed as mesons, baryons, tetra-quarks [43] or even penta-quarks [44]. Asymptotic freedom means that forces between quarks become asymptotically weaker as the energy of the system increases, and the distance decreases.

Predictions of b-hadron processes involving QCD can also be made using an EFT. Despite the large mass of the b quark with respect to $\Lambda_{\rm QCD} \simeq 200\,\text{MeV}$, the system can be treated perturbatively since $\alpha_{\rm QCD}(m_b)$ is sufficiently small. This is known as a Heavy Quark Effective Theory (HQET). In contrast to an EFT where the weak fields have been integrated out, in a HQET it is not possible to remove heavy quark contributions entirely because the b quark cannot decay without violating flavour number. Essentially the b-hadron system is treated akin to a hydrogen atom, where the b quark takes the place of the nucleus, allowing for a highly simplified theoretical treatment, with corrections of order m_b^{-1}.

Despite the use of HQET, the fact is that hadrons are inherently non-perturbative objects, and so it is useful to make further assumptions. An important supposition is that of *factorisation*, which assumes that the short-distance, process dependent, QCD effects are separable from hadronization, the long distance effects. Hadronization is very difficult to calculate with QCD; for this reason *form factors* are used to empirically encapsulate the process. Form factors must be measured experimentally and are the dominant source of uncertainty in hadronic B decays.

[3] The following section is based on Ref. [42].

Chapter 3

The LHCb experiment

The following chapter first introduces the LHC and the LHCb detector before briefly describing data collection and processing. Other important features, such as particle identification and triggering, of the LHCb experiment will also be outlined.

3.1 The LHC

The LHC is a superconducting synchrotron which can simultaneously accelerate beams of proton bunches in opposite directions. Physically, the LHC is located at CERN, near Geneva, Switzerland; it is 27 km in circumference and spans the Franco-Swiss border at a depth of about 100 m. Protons are supplied to the LHC from the Super Proton Synchrotron (SPS) with an energy of 450 GeV, they can then be accelerated and collided with a centre-of-mass energy of up to 14 TeV. In the years 2011 and 2012 collisions operated with a centre-of-mass energy was 7 and 8 TeV respectively. Once the desired energy is reached the beams are collided at four interaction points. The LHCb detector is situated at one of them [45]. Collisions of proton bunches occur every 50 ns reaching luminosities of up to $7 \times 10^{32}\,\mathrm{cm^2 s^{-1}}$, however the beams entering LHCb must be luminosity levelled, to $3(4) \times 10^{32}\,\mathrm{cm^2 s^{-1}}$ in 2011(2012), in order to reduce detector occupancy. These high energy collisions produce a vast number of $b\bar{b}$ pairs which subsequently hadronize into b hadrons. It is the prospect of studying these bound states of b (and other heavy flavour) quarks that has motivated the design of the LHCb detector.

3.2 The LHCb detector

Before introducing the LHCb detector, it is helpful to first define the Cartesian coordinate system around which the LHCb detector is built. The z direction is defined by the LHC

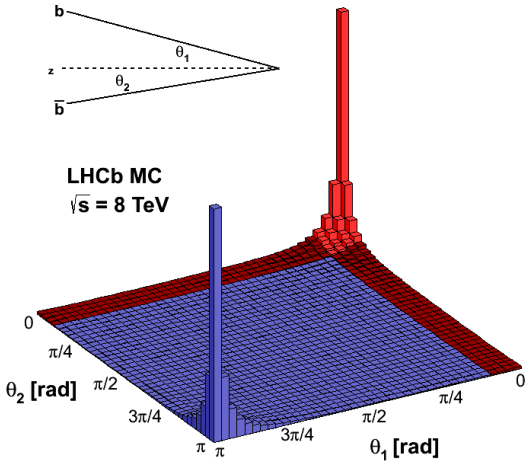


Figure 3.1: Simulation of the production of b and $\bar{b}$ quarks leaving the pp interaction point with angles θ_1 and θ_2, respectively. It is clear that the majority of of these particles have momentum vectors that are close to the beam pipe, which justifies the geometry of the LHCb detector. The dark red indicates the LHCb acceptance region.

beam pipe, where the pp interaction point is located at $z = 0$, and the LHCb detector extends in the positive z direction. Often, the terms *upstream* and *downstream* are used to refer to negative or positive z direction, respectively. An vector originating at the interaction point has an angle θ with respect to the z axis. The x- and y-axes define the horizontal and vertical directions, respectively.

The LHCb detector is a single arm forward spectrometer, reminiscent of a fixed target experiment, with an acceptance of $1.8 < \eta < 4.9$, where pseudorapidity defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]. \tag{3.1}$$

A small angular acceptance is used, because in hadronic interactions $b\bar{b}$ pairs are predominantly produced with momentum vectors aligned close to the beam line. Figure 3.1 shows the θ distribution of the b and $\bar{b}$ quarks produced in simulated LHC collisions, with a centre-of-mass energy of 8 TeV.

A trajectory of a charged particle leaving the interaction region is measured with the tracking system, which consists of a vertex detector and four planar tracking stations. Part of the detector volume is immersed in a magnet field, provided by a large conventional magnet with an integrated field strength of 4 Tm. This bends particles in the x direction, providing information about charge and, in conjunction with the tracking system, momentum. Two Cherenkov detectors are used for offline Particle Identification (PID), allowing LHCb to distinguish between, for example, pions and kaons. Further downstream of the

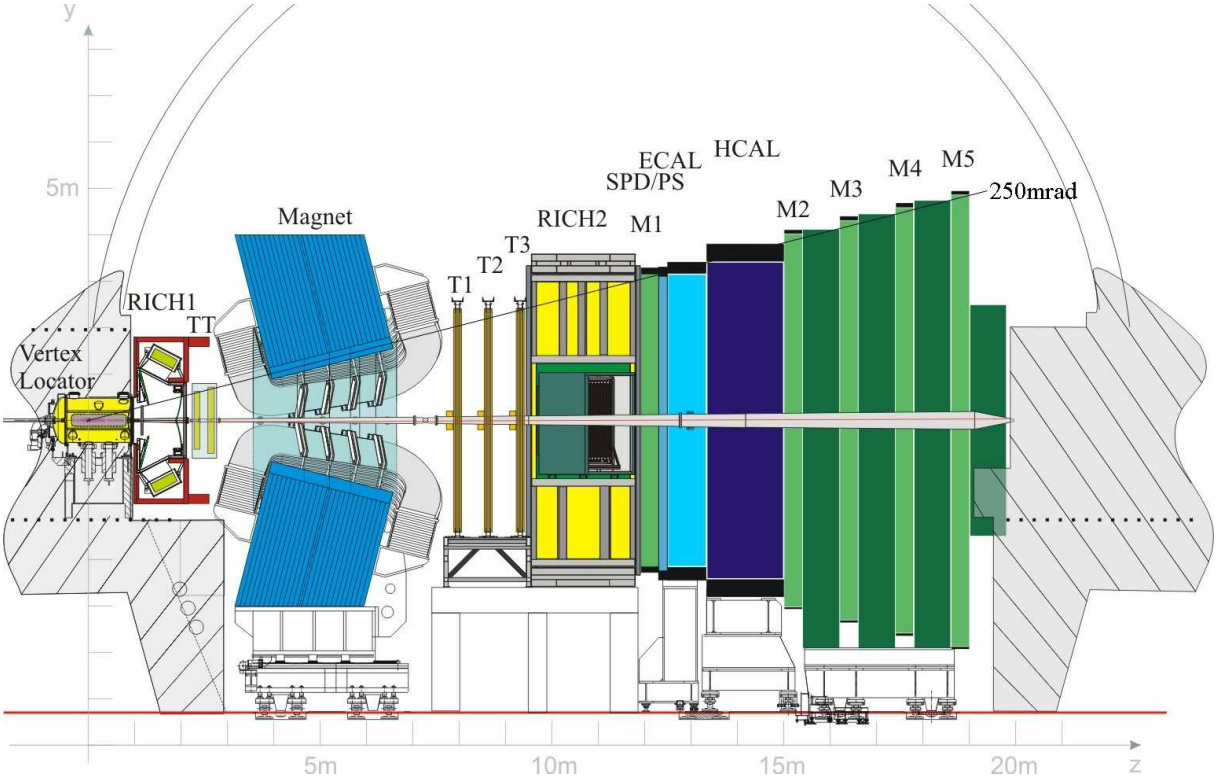


Figure 3.2: Schematic diagram of the LHCb detector with labelled subdetectors and coordinate system, the x direction comes out of the page.

tracking systems is the calorimetry system, which is primarily used for triggering; but is also used for some particle identification. The final subdetector system that a particle might traverse is the muon system; which consists of five stations, and is used for muon tracking, triggering, and PID. Figure 3.2 is a schematic diagram of the LHCb detector, indicating the coordinate system and subdetector modules. The following sections provide further information about LHCb subdetectors.

3.2.1 Tracking

A charged particle produced in a decay of a heavy flavour meson exiting the Primary Vertex (PV) is first detected as *hits* in the Vertex Locator (VELO). The VELO subdetector is made up of 21 modules orientated in the xy-plane, and each module consists of two layers of silicon strip detectors with (r, ϕ) coordinates. The pitch of the silicon strips vary from $\sim$40 µm nearest the centre, where detector occupancy is highest, to $\sim$100 µm at the extremities.

Spacial resolution is vitally important so close to the interaction point. The VELO must be able to resolve all tracks and distinguish PVs coming from pp interactions, and secondary vertices indicative of decaying heavy flavour hadrons. For example, a B^+ meson with a momentum of 100 GeV will travel approximately 1 cm before decaying. This must all be

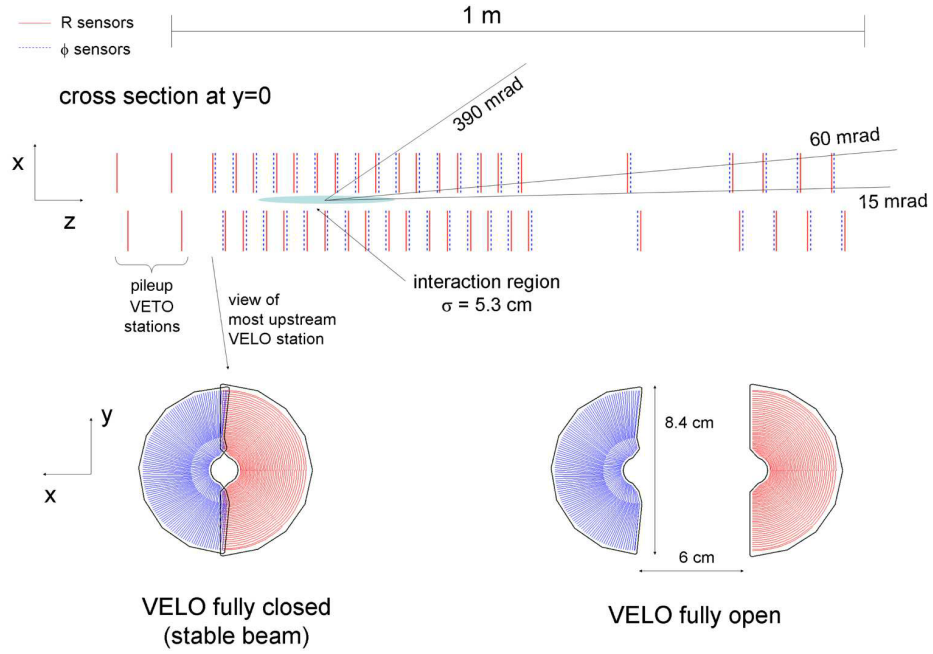


Figure 3.3: The layout of silicon sensors the VELO in showing r sensors in red and ϕ sensors in blue. A cross section at $y = 0$ in the xz-plane is shown while the VELO is closed. Along side these are slices in the xy-plane, with the VELO closed and open.

done in a high track multiplicity environment.

Reconstructed tracks in the VELO are used to construct the PVs, the resolution of each PV is highly dependent upon the number of tracks coming from the vertex. Figure 3.4 shows that for a 25-track vertex, the resolution in the xy-plane is 13 µm, whereas in the z direction it is 71 µm [46]. The resolution of the Impact Parameter (IP) is shown in Fig. 3.5, where it is seen that the performance in the x and y directions are almost identical and exhibit an inverse power law, tending asymptotically to around 15 µm at high momentum.

Hits recorded in the tracking system are fitted to tracks, and in order to decrease the distance of extrapolation of tracks to vertices and increase the spacial resolution of vertices, the active area of the VELO starts 8 mm from the beam line. This is made possible because each module is split into two halves which are retracted when the LHC beam is being injected, and then closed when the beam is declared stable and data taking can begin. Its design leads to a detector with high impact parameter resolution, which can detect tracks emerging from a pp interaction located in the range $1.6 < \eta < 4.9$ and $|z| < 10.6$ cm. The geometry of the tracking stations is shown in Fig. 3.7.

Charged particles leaving the VELO are next observed traversing the Tracker Turicensis (TT), which is park of the tracking system situated between the VELO and main tracking stations. The TT uses uses silicon detector technology. Immediately downstream of the TT is the LHCb magnet, followed by the remaining three tracking stations (T1–3). The

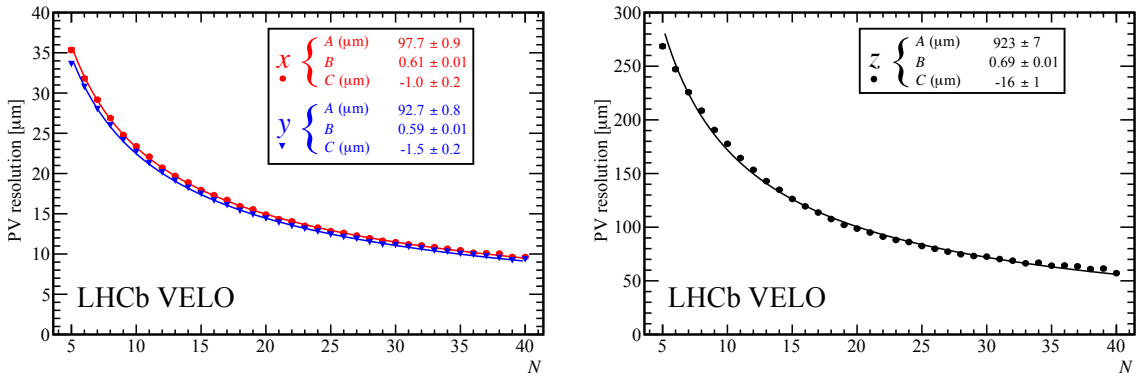


Figure 3.4: PV resolution in the (left red) x, (left blue) y, and (right) z directions of events from 2011 in which there is a single PV as a function of track multiplicity, N [46]. The given values A, B, and C, result from fitting the data points to a resolution function of the form $\sigma_{\text{PV}} = AN^{-B} + C$.

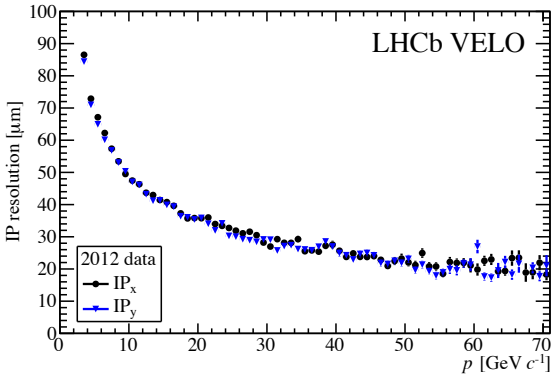


Figure 3.5: Impact parameter resolution as a function of momentum in the (black) x and (blue) y directions as a function of momentum, from data collected in 2012.

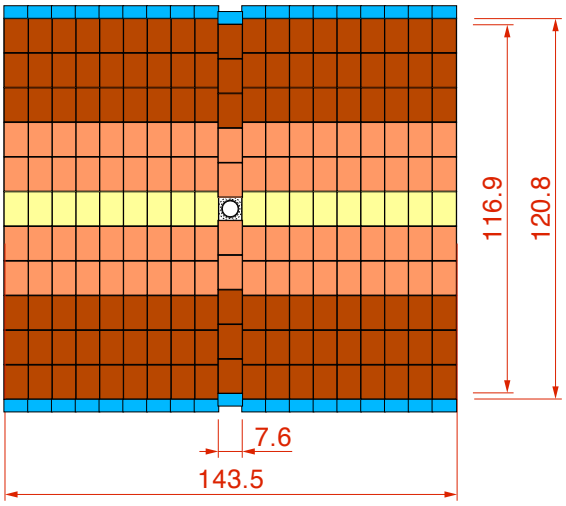


Figure 3.6: Schematic diagram of a single stereo TT layer oriented in the x-direction. Different colours indicate readout density. All units are in cm.

T1–3 tracking stations are each made of two different technologies: the area nearest the beam, aptly named the Inner Tracker (IT), shares the detector technology adopted by the TT; and the Outer Tracker (OT) uses straw drift-tube technology. Each tracking station exhibits an $x - u - v - x$ geometry, where u and v are rotated by -5° and $+5^\circ$ with respect to the y-axis.

Together, the TT and IT are referred to as the Silicon Tracker (ST) which uses silicon strip sensors with a pitch of 200 µm. The TT is the only part of the tracking system that is not used in the trigger, but is instead used to improve the resolution in reconstructing tracks offline. To do this, the TT aids the extrapolation of upstream tracks to hits in T1–3 and the muon station. Spacial resolution of the TT is increased by having a large gap, around 27 cm, between each stereo layer of the detector, whereas in the IT the gap is about 4 cm. To cope with higher occupancy nearest the beam line, the TT has a increased readout density closer to $y = 0$. A schematic diagram of a TT layer is shown in Fig. 3.6.

Figure 3.7 shows a diagram of T1–3. The ITs each occupy a small cross-shaped region closest to the beam in T1–3. The OT is constructed from modules each containing two staggered planes of densely packed straw tubes, each with a diameter of 4.9 mm. In all the VELO, ST and OT give the LHCb detector excellent momentum resolution; $^{\Delta p}/_{p}$ between 0.4 and 0.6 % for particles with momenta of 5 GeV and 100 GeV respectively.

3.2.2 RICH detectors

The primary aim of the LHCb Ring Imaging Cherenkov (RICH) system is to provide PID separation of charged hadrons (π, K, and p), although they are also used to help identify muons and electrons. The ability to correctly assign the correct mass hypotheses is a vital

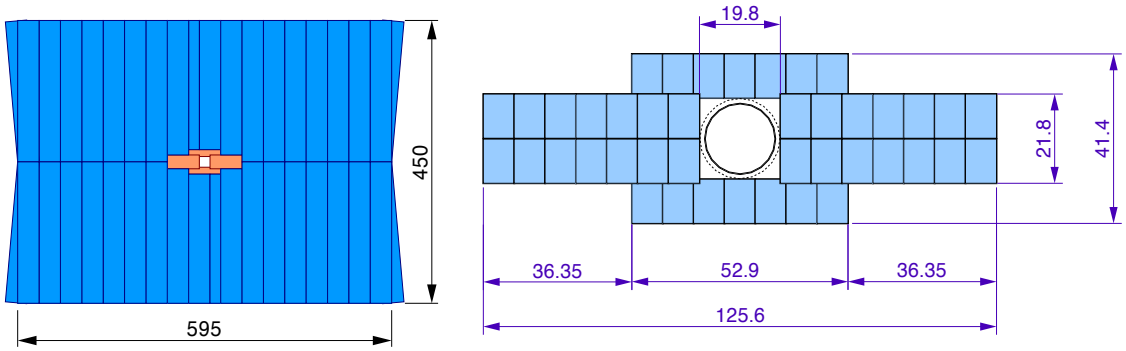


Figure 3.7: Schematic diagram of the (left) tracking stations T1–3, where the outer region is the OT, and the inner cross is the silicon IT. A zoom of an x oriented IT layer is also shown (right). All units are in cm.

part of the LHCb physics program. Since interesting b-hadron decays frequently result in final states with multiple hadrons, misidentifying daughter particles would result in an increase of combinatorial background, and a decrease in signal significance.

The principle behind the operation of the RICH system is that of Cherenkov radiation. When a charged particle travels through a medium, with a refractive index n, at a greater speed than its phase velocity, v_p, it decelerates by emitting a cone of light (Cherenkov radiation) in the forward direction. The opening angle of this light cone, θ_{Ch}, is inversely proportional to, v_p:

$$\cos\theta_{\text{Ch}} = \frac{c}{nv_p}. \tag{3.2}$$

With a measurement of the particle's momentum from the tracking system and only a few possible masses (that of the e, μ, π, K or p) likelihoods are constructed for each track based on the ring of photons which such a particle would emit.

The LHCb RICH system provides PID information for a wide range of momentum with the use of two RICH detectors, containing three different radiators (aerogel and C_4F_{10} in RICH1, and CF_4 in RICH2) between them. Particles passing through these radiators emit a cone of photons, which are focussed using spherical carbon fibre mirrors — which are only 1.5 % X_0 long — on to an array of Hybrid Photomultiplier Detectors (HPDs). Figure 3.8 shows schematic diagrams for RICH1 and RICH2.

Different radiator materials, with different values of n, give the RICH system sensitivity to a range of particle momenta. RICH1 is situated immediately downstream of the VELO and covers the low momentum range, $2 < p < 40\,\text{GeV}$, while RICH2 lies downstream of T3 and covers the range $15 < p < 100\,\text{GeV}$. The variation of θ_{Ch} on momentum for pions, kaons and protons is shown in Fig. 3.9. The variable available to analysts is that of the Delta Log Likelihood (DLL), where DLL_{XY} is the difference between the logarithm of the likelihood of the hypothesis of the particle X compared to the null hypothesis of the Y; where Y is usually the π. In this way, the LHCb detector can achieve excellent pion-kaon

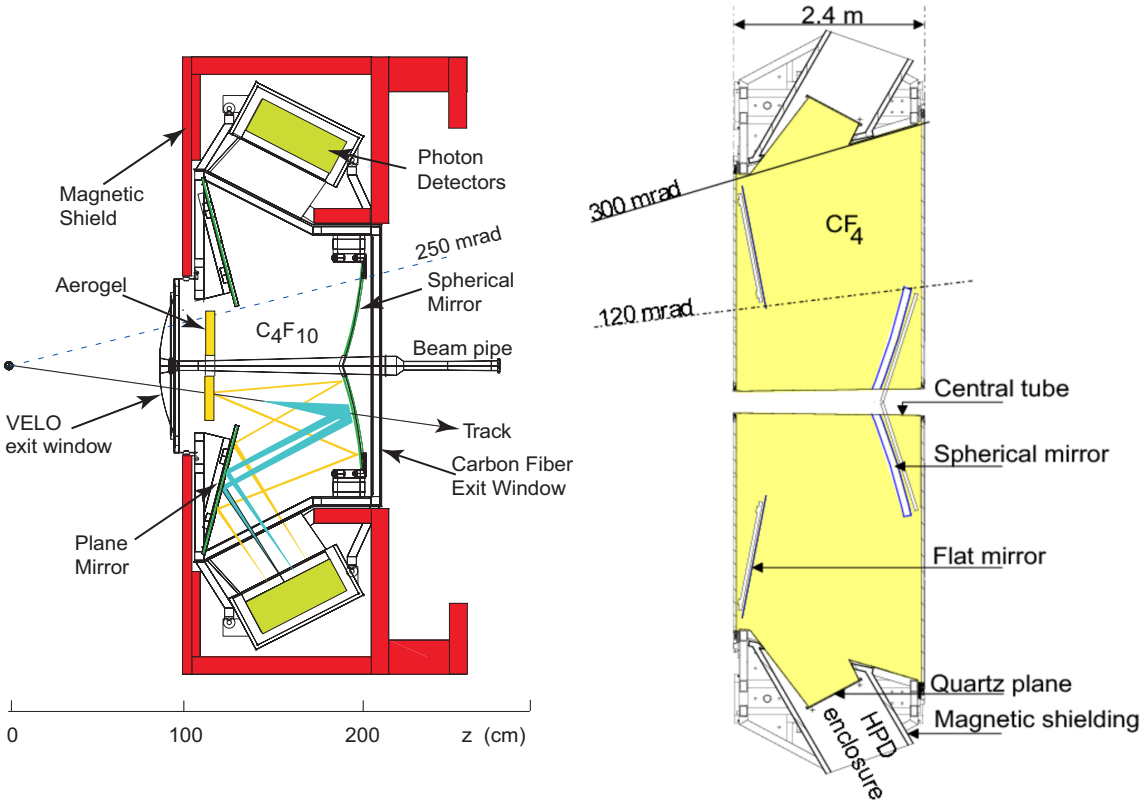


Figure 3.8: Schematic diagrams of (left) RICH1 and (right) RICH2, indicating the radiators, mirrors and HPDs.

separation, for typical kaon produced in a b-hadron decay, which has a momentum of $\sim$20 GeV, the identification rate is near 100 % and the pion misidentification rate is a few percent as shown in Fig. 3.9.

3.2.3 Calorimetry system

The general structure of the calorimetry system is that of an Electromagnetic Calorimeter (ECAL) followed by an Hadron Calorimeter (HCAL), but before both these stations there are additional calorimetry systems designed to provide some PID information in the trigger; this is discussed later. The primary function of the calorimetry system is for triggering high p_{T} hadrons, electrons and photons — this is described in detail in Sec. 3.3.

The Scintillating Pad Detector (SPD) is positioned just downstream of the first muon station, M1, which in turn is immediately downstream of RICH2 and set apart from the other muon stations M2–4. The SPD is made of scintillating tiles which detect charged particles, and therefore can be used to distinguish electrons from photons. Following the SPD is a 15 mm, 2.5 X_0, thick lead plate followed by the Preshower detector (PS) detector, whose design is almost identical to the SPD. The PS exploits the showering caused by the lead plate to distinguish between electromagnetic particles and hadrons, it also helps to

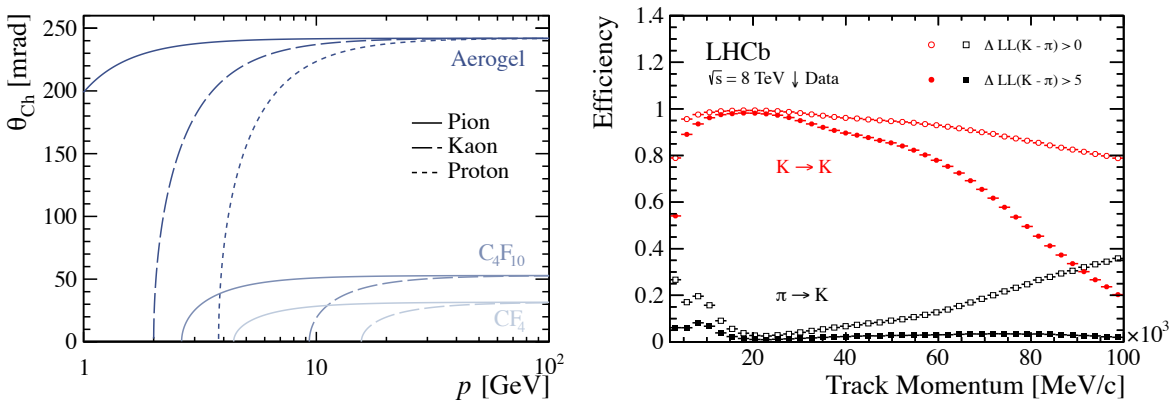


Figure 3.9: Cherenkov angle as a function of momentum for pions, kaons and protons in the three different radiators (left), and the kaon-pion separation performance as a function of momentum, from Ref. [47] (right). The latter plot shows in red circles the efficiency of a kaon being correctly identified as a kaon, and in black squares the rate that pions are misidentified as kaons. Solid and open shapes denote two different cut values at $\mathrm{DLL}_{K\pi} > 0$ and $\mathrm{DLL}_{K\pi} > 5$, respectively. It is observed that for a tighter cut the kaon identification efficiency is lower, but the misidentification of pions is also much more suppressed than the looser cut.

separate photons from π^0 decays.

The ECAL is a shashlik detector, made of alternating layers of lead, reflector, and scintillating material. Light from the scintillators is guided to photomultiplier tubes by wavelength-shifting fibres which penetrate the tiles for the length of each module. This ECAL subdetector is 25 radiation lengths long, ensuring that all electromagnetic energy is deposited before the HCAL. The energy resolution of the ECAL is

$$\frac{\sigma_E}{E} = \frac{10\,\%}{\sqrt{E}} \oplus 1\,\%, \tag{3.3}$$

where E is in GeV.

Figure 3.10 shows the difference in geometry between the ECAL and HCAL. The HCAL has a tiled structure, comprised of alternate layers of scintillating material and iron. The geometry is different to that of the ECAL, since where tiles in the ECAL are oriented towards the incoming particle flux, the HCAL tiles are oriented longitudinally. Each subdetector within the calorimetry system has increased cell density near the beam to cope with higher track multiplicity in this region Due to restrictions in space, the HCAL is only 5.6 nuclear interaction lengths long and has a resolution of

$$\frac{\sigma_E}{E} = \frac{(69 \pm 5)\,\%}{\sqrt{E}} \oplus (9 \pm 2)\,\%, \tag{3.4}$$

where E is in GeV. Light produced in scintillation is guided to photomultiplier tubes by wavelength-shifting fibres.

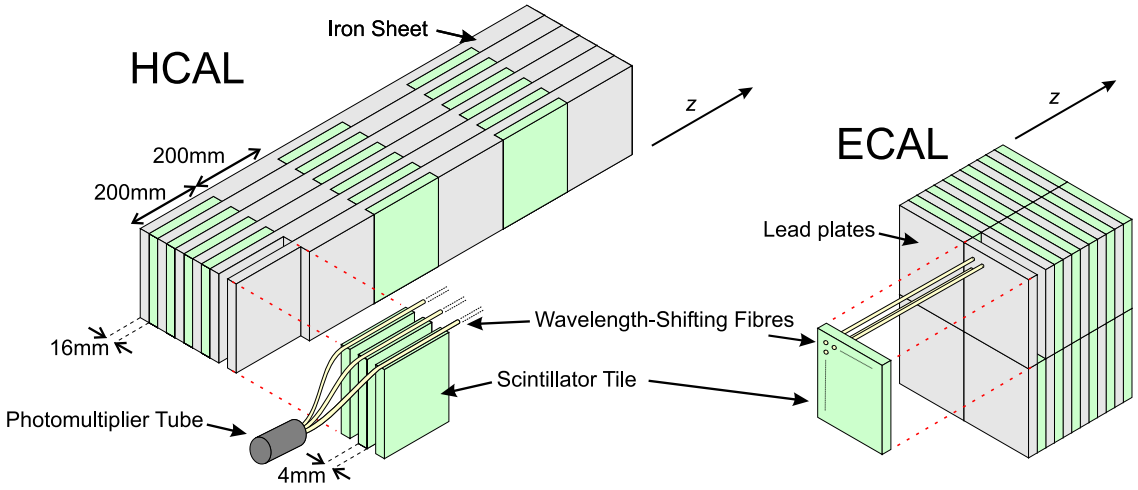


Figure 3.10: Diagrams of the LHCb HCAL and ECAL showing the arrangement of scintillating material, lead plates, and wavelength-shifting fibres.

3.2.4 Muon systems

Muons are the only charged particles that interact weakly enough to pass through the calorimeter system. Muons contribute to the final state of many b-hadron decays, including LHCb's flagship measurements of the branching fraction $B_s^0 \to \mu^+\mu^-$ and the angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$. The muon system is designed to give precise measurements of muon trajectories and triggering of decays with muons in the final state.

The muon system consists of five muon stations, M1–5. The first station, M1, is located directly upstream of the calorimeters; the other stations are all downstream of the calorimetry system and are interleaved with 80 cm thick iron absorbers. The layout of the LHCb muon system is shown in Fig. 3.11.

Each station uses Multi-Wire Proportional Chambers, except for the centre of M1, where the expected muon flux would age this technology too quickly; in this area Gas Electron Multiplier detectors are used. Detector resolution is defined by rectangular logical pads, which have varying spacial resolutions, becoming coarser as the distance from the beam axis increases. These regions are discretised, and labelled R1–4 as shown in Figure 3.12, where the segmentation ratios for R1 : R2 : R3 : R4 are 1 : 2 : 4 : 8, such that the occupancy in each region is approximately equal. The number of detection cells in the bending plane is greater than in the non-bending plane, in order to increase the momentum resolution. The inclusion of the M1 station is primarily to provide improved momentum resolution in the trigger, giving a momentum resolution of approximately 20 %. In total, the absorber material is 20 interaction lengths, such that only highly penetrating muons reach M4–5; in order to pass through all the muon stations a muon must have a p_{T} of at least 6 GeV. Therefore, stations M1–3 have considerably better spacial resolution than

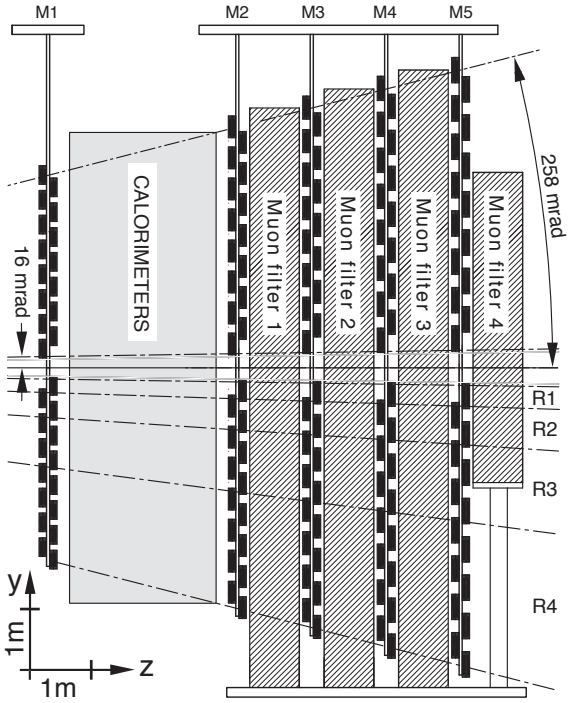


Figure 3.11: Side-on view of the LHCb muon system. The M1 station is located upstream of the calorimeters, and M2–5 are all downstream. Relative scales are the same, but absolute dimensions scale with distance from the pp interaction point, such that the angular coverage is identical for each station. Between stations M2–5 are 80 cm thick iron filters. Regions R1–4 indicate areas of equal flux, the resolution changes accordingly, being courser grained further from the beam line, this is shown clearly in Fig. 3.12.

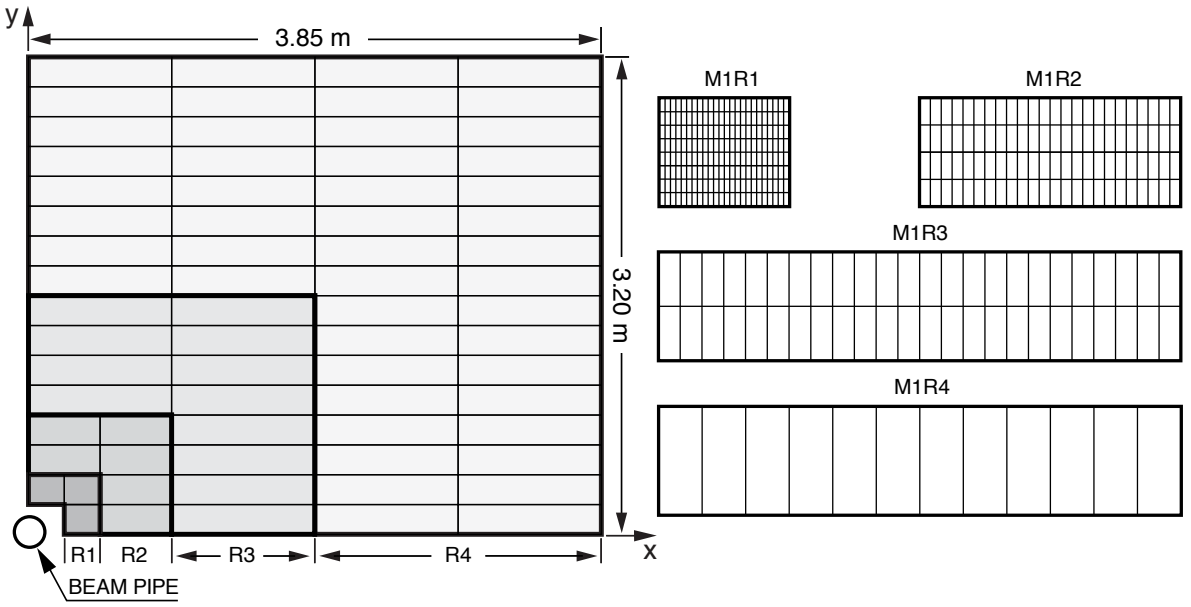


Figure 3.12: Front-on view of a single quadrant of a muon station. The left-hand pane shows the chambers distributed throughout regions R1–4. Each chamber comprises a logical pad structure as shown for M1 in the right-hand diagram. The resolution for the stations in M2–3(M4–5) are double(half) that of M1.

Table 3.1: The `isMuon` flag is an important variable used to identify muons. The criterion depends on the particle's measured momentum. If the `isMuon` condition returns a true then the particle is identified to be a muon.

$p(\mu)$ range (GeV)	`isMuon` condition
$3 < p(\mu) < 6$	M2 and M3
$6 < p(\mu) < 10$	M2 and M3 and (M4 or M5)
$p(\mu) > 10$	M2 and M3 and M4 and M5

stations M4–5, which are primarily used to identify highly penetrating muons.

The muon system provides important PID information. In many LHCb analyses a muon is identified based on hits within the muon system; this criterion is known as `isMuon`. This criterion is boolean, and its exact response is defined by the muon momentum, which is explicitly given in Table 3.1.

3.3 Trigger

With a pp interaction rate of 40 MHz and multiple subdetector stations, there is too much information associated with each event to write everything all to disk. Instead a multistage trigger is employed to filter potentially interesting events based on track momentum and energy. The Level One Trigger (L0) is embedded in the hardware of LHCb and is fully

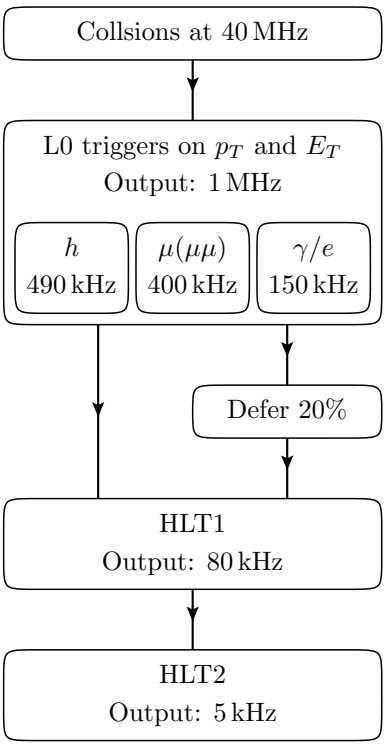


Figure 3.13: The flow of the LHCb trigger system in 2012, where the 20 % of events that are deferred are processed during machine down-time and proton acceleration stages.

synchronous with the bunch crossing rate, it reduces the output 1 MHz. There are five L0 trigger lines; one each for objects identified as photons, electrons, hadrons, muons and pairs of muons (dimuon). Two software triggers follow L0, and are fast enough to perform tracking algorithms and use information from multiple subdetectors to reduce the events written to disk to 5 kHz. The flow of data through the trigger is shown in Fig. 3.13.

The L0 trigger algorithms for photons, hadrons, and electrons are based on calorimeter objects. A single calorimeter cluster is defined as two-by-two calorimeter cells in each the ECAL and HCAL. For each cluster the transverse energy, E_T, is calculated:

$$E_T = \sum_{i=1}^{4} E_i \sin\theta_i, \tag{3.5}$$

where E_i is the energy in cell i and θ_i is the angle between the average interaction point and the cell's centre. Clusters are categorised as follows. A hadron candidate is the largest E_T cluster in the HCAL summed with the E_T of the ECAL cluster in front, if there is one. A candidate photon (electron) is the largest E_T deposit with hits in the PS cells in front and no hits (at least one hit) in the nearest SPD cells. The E_T of each candidate is compared to thresholds, and the event is retained if one or more is exceeded.

L0 trigger lines associated with muons base their acceptance on measurements of p_{T}. Each quadrant of the muon system is read out independently, so muons which cross boundaries

Table 3.2: Thresholds in 2011 and 2012 for L0 trigger lines [48] used in this thesis.

	2011	2012
`L0Muon`	1.48 GeV	1.76 GeV
`L0Dimuon`	$(1.296\,\text{GeV})^2$	$(1.6\,\text{GeV})^2$
`L0Hadron`	3.5 GeV	3.7 GeV

cannot be triggered. Muon candidates with the highest and second highest p_{T} are selected from each quadrant by searching for straight lines through M1-5 in the zy-plane, and in the zx-plane if $p_T > 0.5\,\text{GeV}$. The M1 station is used to increase the p_{T} resolution of muon tracks in the trigger without tracking information: this achieves a resolution of about 25 % that of fully reconstructed tracks. Events are accepted based on candidates from all quadrants with values of $p_T^{\max}$ and $p_T^{\max} \times p_T^{2^{\text{nd}}\max}$ greater than thresholds for the muon and dimuon lines respectively. Thresholds for L0 trigger lines used in this thesis are given in Table 3.2.

Deferred triggering was introduced in 2012. It diverts around 20 % of events that pass the L0 trigger to hard disks for processing when the LHC is not colliding protons; this is schematically shown in Fig. 3.13. The remaining 80 % of events are processed immediately by the software triggers, which are run on around 2000 computing nodes using C++ applications.

Following L0, there are software High Level Triggers (HLTs). The first software trigger, HLT1, performs the full three dimensional VELO track fitting algorithms (but with fewer passes than the offline version). Candidate VELO tracks for triggers which do not require muons are selected based upon the quality of the VELO track and the smallest IP with respect to any of the identified primary vertices. Primary vertices are defined to be points within 300 µm of the mean interaction point in the xy-plane, $\mathrm{PV}_{xy}^{\mathrm{mean}}$, from which at least five tracks originate. The position of $\mathrm{PV}_{xy}^{\mathrm{mean}}$ is measured at the start of each LHC machine fill. For trigger lines requiring muons, each VELO track is extrapolated to a window in the M3 station. The size of this window is narrow in the non-bending direction but wide enough to accommodate a 6 GeV muon in x. If there is a deposit in this window then the VELO track is extrapolated to the cluster and if there are hits consistent with this track in any of the muon stations M2, M4 or M5, the track is tagged as belonging to a muon. The VELO tracks that are selected by IP or the muon system are extrapolated (or interpolated) into the IT and OT. This is known as forward tracking, and provides momentum measurements for all these tracks.

The HLT1 output rate of 80 kHz is sufficiently low to allow the forward tracking algorithm to be performed on all VELO tracks (rather than just those that appear to come from a PV). However, the HLT2 is not fast enough to perform the full off-line track reconstruction, and is limited to VELO tracks with $p > 5\,\text{GeV}$ and $p_T > 0.5\,\text{GeV}$ in 2011; which was

relaxed to $p_T > 0.3\,\text{GeV}$ in 2012 thanks to deferred triggering.

The output of HLT2 is dominated by the *topological* trigger lines, which are designed to identify b-hadrons decaying into charged tracks using vertex and track information consistent with the decay topology of a b-hadron. Vertices formed of two, three and four reconstructed tracks displaced from PVs are triggered based on the response of a Bonsai BDT (BBDT) [49]. A BBDT is a Boosted Decision Tree, which are detailed in Chap. 4, whose input and output distributions are discretised so that a simple look-up table can be used to calculate the response. This approach is not only fast, but 9 % more efficient than using a cuts based selection for a 4-body signal [49].

Events that are selected are flagged as either Trigger On Signal (`TOS`), or Trigger Independent of Signal (`TIS`) — the latter meaning that the event was triggered by a different particle in the event. This allows the analyst the ability to calculate an estimate of the trigger efficiency in data using the `TISTOS` method. One can get an approximate trigger efficiency using

$$\varepsilon_{\text{trig}}^{\texttt{TISTOS}} = \frac{N_{\texttt{TIS\&TOS}}}{N_{\texttt{TIS}}}. \tag{3.6}$$

This can be useful because while simulated events can contain events which were not triggered, this is, obviously, not the case for data. This is not perfect because `TIS` b-hadron candidates are usually fired by the other b-hadron in the event (from $g \to b\bar{b}$), and the kinematics between the two are highly correlated.

3.4 Data, stripping, and simulation

The data collected by the LHCb detector used in this thesis totals $3\,\text{fb}^{-1}$; where $1\,\text{fb}^{-1}$ was collected in the year 2011 with a centre-of-mass energy of 7 TeV, and $2\,\text{fb}^{-1}$ at 8 TeV was collected in 2012. In total, the data collected in 2011 and 2012 is known as Run-1 data.

Even the much reduced HLT2 output rate of 5 kHz is a vast amount of data for an analyst to sift through in a timely manner. To improve the speed to access data, additional selections are applied to the dataset biannually which further categorise each event. This is known as stripping. Stripped datasets are the only ones accessible to analysts, which makes the process of retrieving data of interest fast. Stripping selections in this thesis vary, and will be described when appropriate.

Simulated events are used in LHCb analyses primarily for the purposes of optimising selections and determining efficiencies. Other important uses include the modelling of kinematic distributions and understanding various sources of background. The LHCb collaboration produces simulated events in a pipeline consisting of multiple stages. Proton-proton collisions are generated using Pythia [50] with a configuration specific to LHCb [51], giving details of: the size of the luminous region; the number of interactions per bunch

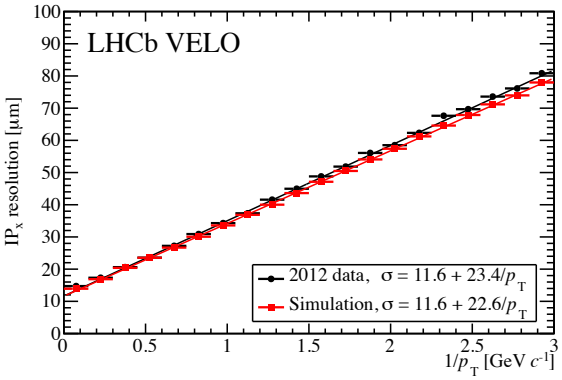


Figure 3.14: Impact parameter resolution as a function of a track's inverse transverse momentum comparing (black) data taken in 2012 which is also shown in Fig. 3.5 and (red) simulation. Excellent agreement is seen, particularly for high momenta tracks, $p_T > 1\,\text{GeV}$ [46].

crossing; and spill-over events from neighbouring proton bunches. Heavy flavour particles are produced by a hadronization process in PYTHIA. Subsequent decays are modelled with EVTGEN [52], which is a package created specifically for modelling the decays of heavy flavour particles. The composition of decays in each sample of simulated events is requested at this level, and EVTGEN forces one of the heavy flavour quarks to decay accordingly, other heavy quarks proceed to decay via a random decay chain. Selection cuts are applied throughout and if the required quark of heavy flavour is produced in the backward direction the z-coordinates are flipped, this is to save time at later stages. At this stage,the simulation is known as *generator level*, which can be useful for determining distributions before the detector interactions have been accounted for. Final state radiation is modelled using PHOTOS [53]. The interactions of each particle with the LHCb detector is modelled using GEANT4 [54] as described in Ref. [55]. Trigger decisions are applied to the simulation, however events which do not pass various triggers are not discarded, and instead are stored. After the simulation is complete, events are treated identically to the collected data; being processed by the same stripping selections.

Considering the importance of simulated events it is important for them to be trusted. To this end each analysis must ensure that the simulation used describes the data to a good degree, sometimes it is necessary for weights to be applied to correct for small differences. This is particularly true for PID variables, which are known to be poorly modelled by the simulation. That being said, the simulated events are seen to reproduce data distributions very accurately. Figure 3.14 shows the IP resolution as a function of the inverse p_T of a track, the performance is roughly linear in this variable. The VELO achieves IP resolutions of less than $25\,\mu\text{m}$ for particles with $p_T > 1\,\text{GeV}$, and the agreement between data and simulation is seen to be excellent [46].

Chapter 4

Multivariate selection techniques

The analyses detailed in Chap. 5, Chap. 6 and Chap. 7 make prodigious use of multivariate techniques to reduce combinatorial backgrounds. Combinatorial backgrounds are formed from random combinations of tracks which appear to form a vertex, pass selection criteria, and satisfy relevant PID assignments. To remove these backgrounds, Multivariate Analysis (MVA) techniques can be employed. A multivariate discriminator exploits correlations between weakly discriminating variables to produce a single, more separating, classifier.

MVA techniques used in HEP tend to be supervised learning algorithms, whereby a selection of events are given input, and an algorithm produces a response based on how best to separate them. Input into the algorithms to separate a background are: a sample of the signal and background candidates that must be separated, and a set of variables to be used to do so. Samples of events are split in two; some are used for training the MVA, and the remaining are used for testing it. The input, or training, variables define an n-dimensional space populated by the input samples. The algorithm then classifies regions in this n-dimensional space as signal- or background-like; such that an arbitrary event placed somewhere in the space would also be classified based on the point it inhabits. The Boosted Decision Tree (BDT) algorithm is used throughout this thesis because it can handle a weighted training sample, including negative weights, and can exploit non-linear correlations between variables [56, 57].

A BDT is composed of a combination of numerous Decision Trees (DTs), each of which is a classifier in its own right — albeit a weak one — being able to distinguish between high density regions of signal and background populations.

Training a DT begins with a single parent node populated by the whole training sample, which inhabits the parameter space defined by the variables x_i, whose true distribution is $f(x_i)$. The sample on the parent node is split by selecting a cut based on maximising some figure-of-merit. Child nodes are then split, and the process is repeated until there is no possible improvement in the separation of signal and background. The definition of

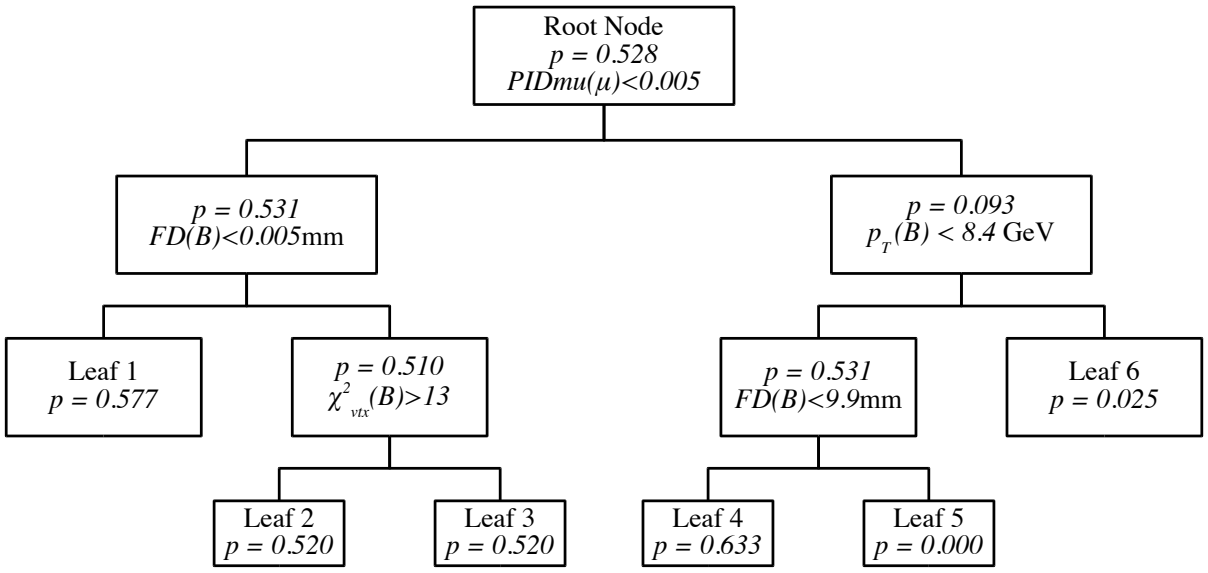


Figure 4.1: An example single DT, taken from the analysis in Chap. 7, showing the root node and decendant nodes. At each node the cut value is shown along with the purity of the node. The variables Flight Distance (FD) and vertex χ^2 are defined in later chapters. Signal nodes have a higher purity, p, and are split to the left, such that candidates landing on leaf 1, 2, or 4(3, 5, or 6) are classified as belonging to signal-like(background-like) events.

improvement is usually related to the purity of a node:

$$p = \left(1 - \varepsilon_{\text{bkg}}\right), \tag{4.1}$$

where ε_{sig} is the signal efficiency on a given leaf. The final child nodes, or leaves, are each associated with signal or background depending on the purity of the sample which populates it. Each leaf therefore maps out an area in n-dimensions, and is classified as a signal or background leaf depending on the purity of the training sample enclosed by that area. Figure 4.1 shows an example of a single DT. The hypothesised category, as output by the DT, $h(x_i)$, will ideally be equal to $f(x_i)$, but in reality there will be events which are misclassified. A figure-of-merit which is often used to determine the cut used at each node is the G_{ini} index, which is defined as

$$\mathrm{G}_{ini} = 2p(1-p) = 2(1-\varepsilon_{\text{sig}})\varepsilon_{\text{sig}} = \frac{2sb}{(s+b)^2}, \tag{4.2}$$

where s and b are the weighted sum of signal and background candidates, respectively, after a given cut.

Decision Trees have the advantages over other machine learning algorithms — such as neural nets — of being easy to interpret and fast to train. They are also able to deal with able to deal with weighted training samples, and are insensitive to variables with very little separation power because the G_{ini} index never identifies a cut on them as being profitable.

However, DTs are sensitive to statistical fluctuations in the training sample. To negate this problem DTs can be *Boosted* using any one of a number of algorithms. The procedure of boosting removes the power that statistical fluctuations has over the final BDT.

A different boosting method is used to train the BDTs in each analyses in this thesis. The algorithms used are outlined in the remainder of this chapter.

4.1 The bagging algorithm

Bootstrap aggregating, or bagging[1], is a method of boosting whereby the effects of statistical fluctuations are negated by making many independent DTs and using them to make a decision based on the average response. Training n DTs, it is possible to define three errors, namely: square of the error of a single estimator

$$\epsilon_t(x_i) = \big(f(x_i) - h_t(x_i)\big)^2 \tag{4.3}$$

where the index t denotes a single DT; the average of individual squared errors

$$\bar{\epsilon}(x_i) = \frac{1}{n}\sum_{t=1}^{n} \epsilon_t(x_i); \tag{4.4}$$

and the squared error of an ensemble of DTs

$$e(x_i) = \big(f(x_i) - \bar{h}(x_i)\big)^2. \tag{4.5}$$

The weighted variance of response of the estimators h_t around a weighted mean is defined as

$$V(x_i) = \frac{1}{n}\sum_{t=1}^{n} \big(h_t(x_i) - \bar{h}(x_i)\big)^2. \tag{4.6}$$

After some basic algebraic manipulation of Eq. 4.6, the relationship

$$e(x_i) = \bar{\epsilon}(x_i) - V(x_i) \tag{4.7}$$

emerges. This means that the error squared of the ensemble of DTs is equal to the average error squared of an individual estimator, minus the weighted variance. Therefore the process of bagging reduces the effect of statistical fluctuations in the training samples [59].

A bagged BDT is trained by randomly selecting events, with replacement, to train a single DT. Hundreds of DTs can then be trained, and the weighted response from all DTs is the result of the classifier, in the range zero to one.

[1] This section is based on Ref. [58].

The bagging algorithm is used to train two BDTs in Chap. 5, one to identify the D_s^+ meson, and the other to identify the ϕ meson, above background combinatorics. Both these BDTs are trained using a large number, $\mathcal{O}(50)$ variables, and therefore the simplicity of bagging is advantageous. Also, because the analysis uses two BDTs, it is better to cut on the product of the BDT responses, which is more natural in the case of the bagging algorithm because the response is between zero and one.

4.2 The AdaBoost algorithm

The Adaptive Boost, AdaBoost, algorithm[2] negates the effect of statistical fluctuations in a data set by increasing the weights of misclassified events. The algorithm begins by training a single DT as described above, where each event has unit weight. In subsequent DTs, the weight for each event, i, is modified for a tree t, to be

$$w_i^t = c_i^t \times w_i^{t-1}, \tag{4.8}$$

where c is based on the classification of the event in the previous tree, it is determined to be

$$c_i^t = \exp\left(\alpha_t \gamma_i^t\right), \qquad \gamma_i^t = \begin{cases} 0 & \text{event } i \text{ classified incorrectly by tree } t-1 \\ 1 & \text{event } i \text{ classified correctly by tree } t-1 \end{cases}. \tag{4.9}$$

Weights are then renormalised such that they sum to unity. The value of α_t is the weight that the DT carries, and is given by

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{4.10}$$

where ϵ is the weighted error rate. This reweighting procedure artificially fluctuates the training sample which is used to train each DT. Multiple DTs are made in this fashion, forming a forest; where the response of the BDT classifier is a combination of responses from all DTs in the forest. The total response of a BDT, T, is

$$T(x_i) = \sum_{t=1}^{n} \alpha_t T_t(x_i) \tag{4.11}$$

where $T_t(x_i)$ is the response of tree t, which returns one if it classifies x_i as being signal-like, and negative one if it is background-like.

The AdaBoost algorithm is fast, and often used in HEP analyses. It is implemented in Chap. 6 to distinguish signal $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$ candidates from

[2] This section is based on Ref. [60].

combinatorial backgrounds.

4.3 The uBoost algorithm

The uniform Boosting, uBoost, algorithm[3] is designed to give a uniform response with respect to the signal efficiency in a set of variables. This is used to create a BDT which is not biased towards some region of parameter space. The procedure of creating a Uniform BDT (uBDT), builds from the weighting technique used in the AdaBoost algorithm, but additional weight is applied to events that lie in a region of parameter space which is not performing with the desired efficiency. Consider a BDT whose signal efficiency in the variables ζ_i is required to be ϵ. The weighting of each event is modified from that used in the AdaBoost algorithm in Eq. 4.8 to

$$w_i^t = u_i^t \times c_i^t \times w_i^{t-1}, \tag{4.12}$$

where u denotes a weighting proportional to the distance away from ϵ in the local region:

$$u_i^t = \exp\left(\beta_t(\bar{\epsilon} - \epsilon_i^t)\right), \tag{4.13}$$

which can be calculated using a nearest neighbours algorithm or by binning. The boosting parameter β is calculated as

$$\beta_t = \frac{1}{2}\ln\left(\frac{1-e_t}{e_t}\right) \tag{4.14}$$

where

$$e_t = \sum_i w_i^{t-1} c_i^t \left|\bar{\epsilon} - \epsilon_i^t\right|. \tag{4.15}$$

This leads to a single BDT whose response is analogous to Eq. 4.16 but with the addition of the target efficiency $\bar{\epsilon}$

$$T(x_i, \bar{\epsilon}) = \sum_{t=1}^{n} \beta_t T_t(x_i, \bar{\epsilon}). \tag{4.16}$$

Therefore, a single of these BDTs is associated with a given target efficiency, where fraction of $T(x_i, \epsilon) > \overline{T}(\bar{\epsilon})$ is $\bar{\epsilon}$. An arbitrary number of BDTs, N, can be concatenated, each with a different target efficiency, and the total response is

$$\mathcal{T}(x_i) = \frac{1}{N}\sum_{e=1}^{n} \Theta\left(T(x_i, \bar{\epsilon}) - \overline{T}(\bar{\epsilon})\right). \tag{4.17}$$

[3] This section is based on Ref. [61].

For $\mathcal{T}(x_i)$ to be a continuous distribution, $N \to \infty$, however in practice $N \sim 100$ is all that is needed for analysis.

Use of the uBoost algorithm is well motivated in Chap. 7, which is a search for a particle of unknown mass and lifetime. In this scenario, the uBoost technique allows the use of a BDT while ensuring that the response is not bias towards some lifetimes or masses.

4.4 Summary

Multivariate selection techniques are a vital part of HEP analyses, and help rare processes to be separated from underlying combinatorial backgrounds. There are a number of algorithms available to the analyst, each should be considered when designing an analysis. However, there are circumstances where the use of a particular algorithm is very important. For example, the BBDTs outlined in Sec. 3.3 is must be fast; and the uBoost algorithm implemented for the analysis in Chap. 7 is required to prevent the selection being bias towards a region of mass and lifetime. Without these various machine learning techniques the field would be restricted to cut-based analyses and the sensitivity of many searches, especially those for very rare decays, would be considerably weakened.

The variables and training samples used to train the BDTs will be discussed in the relevant chapters.

Chapter 5

Search for the decay $B^+ \to D_s^+ \phi$

5.1 Introduction

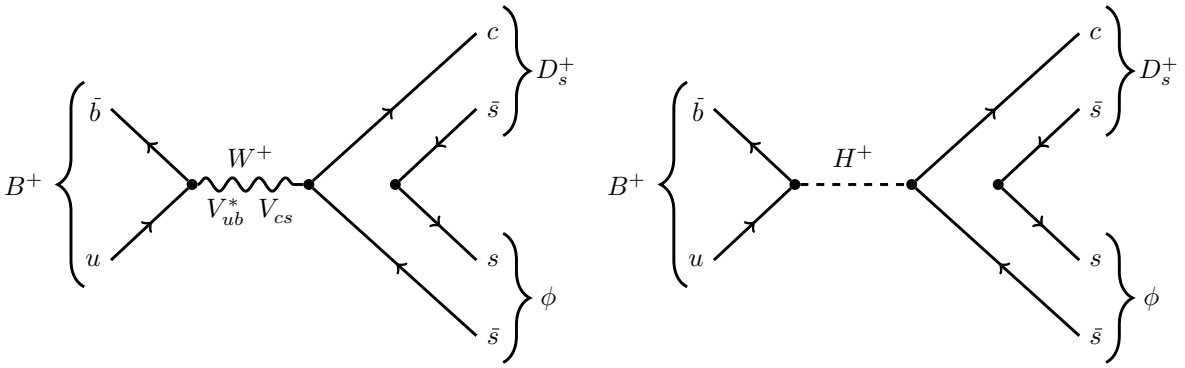


Figure 5.1: A Feynman diagram for the decay $B^+ \to D_s^+ \phi$ being mediated by a (left) W^+ in the SM, and (right) H^+ in SUSY. The $s\bar{s}$ pairs shown here are formed from a gluon that can come from any quark. The arrangement of quarks forming the final state mesons shown is the colour favoured decay

In the SM, the decay $B^+ \to D_s^+ \phi$ proceeds via the annihilation of the constituent b and u quarks of a B^+ meson into a virtual W^+ boson from the CC interaction. This transition is suppressed CKM matrix element V_{ub}[1]. To achieve the final state, the W^+ decays into a $c\bar{s}$ pair and an additional $s\bar{s}$ pair must be created from the QCD field. This is the only diagram that can perpetuate such a decay at tree-level because the initial state quarks are all different to those in the final state. A Feynman diagram of the decay $B^+ \to D_s^+ \phi$ is shown in Fig. 5.1, where the final state mesons can be formed in the way indicated, or the $s\bar{s}$ pair from the QCD field can form the ϕ, although this is colour-suppressed. Also, the gluon that forms the $s\bar{s}$ pair can originate from any of the initial or final state quarks. This analysis was published in Ref. [1].

[1] All mentions of the ϕ meson refer to the $\phi(1020)$.

Annihilation decays of B^+ mesons are rare in the SM due to the magnitude of $|V_{ub}| \sim 4 \times 10^{-3}$. In fact, no fully hadronic decays proceeding via annihilation-type diagrams have yet been observed.

Predictions for the branching fraction $\mathcal{B}\big(B^+ \to D_s^+ \phi\big)$ are calculated using the OPE defined by the effective Hamiltonian [62–65]:

$$\mathcal{H}_{\text{eff}} = -4\frac{G_F}{\sqrt{2}} V_{ub} V_{cs}^* \big[C_1(\Lambda)\mathcal{O}_1 + C_2(\Lambda)\mathcal{O}_2\big] \tag{5.1}$$

where

$$\begin{aligned} \mathcal{O}_1 &= \big(\bar{b}\gamma_\mu P_L u\big)\big(\bar{c}\gamma_\mu P_L s\big) \\ \mathcal{O}_2 &= \big(\bar{b}\gamma_\mu P_L s\big)\big(\bar{c}\gamma_\mu P_L u\big). \end{aligned} \tag{5.2}$$

The Wilson coefficients C_1 and C_2 are defined at the scale $\Lambda = m_b$, and the projection operators are defined as $P_L = \frac{1}{2}(1-\gamma_5)$ and $P_R = \frac{1}{2}(1+\gamma_5)$. The short distance operators $\mathcal{O}_1$ and $\mathcal{O}_2$ both describe the transition $b \to scu$. Therefore, not only is there uncertainty in the branching fraction introduced by V_{ub}, but the number of quarks in the decay make QCD calculations very difficult. Calculating the amplitude of the decay $B^+ \to D_s^+ \phi$ is made particularly complicated because the decay is inherently non-factorizable, since the $s\bar{s}$ pair can come from any of the initial or final state quarks at leading order. There are also inherent uncertainties in the form-factors that describe the hadronization process of the final state quarks. Reference [63] predicts that

$$\mathcal{B}\big(B^+ \to D_s^+ \phi\big)|_{\text{SM}} = 1.88 \times 10^{-6},$$

by naïvely assuming factorizability holds, and by using an improved technique [66], whereby perturbative QCD corrections are applied to the factorisation method, a value of

$$\mathcal{B}\big(B^+ \to D_s^+ \phi\big)|_{\text{SM}} = 0.67 \times 10^{-6},$$

is calculated. The QCD corrections lead to a new branching fraction prediction which differs by a factor of two to the uncorrected result — this is arguably testament to the difficulties of accounting for QCD in such calculations. Other SM predictions tend to lie between $\sim 1 \times 10^{-7}$ and $\sim 7 \times 10^{-7}$ [62–65].

Despite the theoretical uncertainties, the value of in $\mathcal{B}(B^+ \to D_s^+ \phi)$ could be significantly enhanced if the decay to be mediated by additional BSM particles, particularly other charged bosons. For example, a Two Higgs Doublet Model (2HDM) — such as SUSY — the decay $B^+ \to D_s^+ \phi$ would be mediated by a charged Higgs H^+; this is shown in Fig. 5.1. More particles mean more Feynman diagrams that could add to the total amplitude. Reference [63] also makes predictions for the branching fraction of the decay $B^+ \to D_s^+ \phi$

in a 2HDM and a model with R-Parity Violation (RPV):

$$\begin{aligned}\mathcal{B}\big(B^+ \to D_s^+ \phi\big)|_{\text{2HDM}} &= 8.0 \quad \times 10^{-6},\\ \mathcal{B}\big(B^+ \to D_s^+ \phi\big)|_{\text{RPV}} &= 3.06 \times 10^{-4}.\end{aligned}$$

These numbers indicate that, while the exact SM value of $\mathcal{B}\big(B^+ \to D_s^+ \phi\big)$ is not well known, the value for models with additional mediating particles could be enhanced by a factor of over 100.

The CP asymmetry, $\mathcal{A}_{CP}$, of a process is defined in terms of decay rates of B hadrons:

$$\mathcal{A}_{CP} = \frac{\Gamma(\overline{B} \to \overline{f}) - \Gamma(B \to f)}{\Gamma(\overline{B} \to \overline{f}) + \Gamma(B \to f)} \tag{5.3}$$

for some final state f. A positive value of $\mathcal{A}_{CP}$ would indicate a preference of the antimatter process, above the matter process. In the SM $\mathcal{A}_{CP}(B^+ \to D_s^+ \phi) = 0$, because at leading order there is only one phase, in V_{ub}, but interference from BSM physics diagrams could alter this significantly. Predictions from Ref. [63] are:

$$\begin{aligned}\mathcal{A}_{CP}\big(B^+ \to D_s^+ \phi\big)|_{\text{2HDM}} &\leqslant 59\,\%,\\ \mathcal{A}_{CP}\big(B^+ \to D_s^+ \phi\big)|_{\text{RPV}} &\leqslant 14\,\%.\end{aligned} \tag{5.4}$$

So, both measurements of $\mathcal{B}\big(B^+ \to D_s^+ \phi\big)$ and $\mathcal{A}_{CP}\big(B^+ \to D_s^+ \phi\big)$ could lead to evidence for NP.

5.1.1 Other annihilation-type hadronic decays

The annihilation of the B^+ meson can perpetuate numerous decays resulting in fully hadronic states, including a charmed meson. The decay $B^+ \to D^+ K^{*0}$ proceeds in the same way as $B^+ \to D_s^+ \phi$, but the former needs a $d\bar{d}$ pair to be created from the QCD vacuum, rather than an $s\bar{s}$ pair. Similarly, the decay $B^+ \to D_s^+ \overline{K}^{*0}$ is identical to the $B^+ \to D_s^+ \phi$ excepting that instead of $W^+ \to c\bar{s}$ the W^+ decays into an $c\bar{d}$ pair. The decays $B^+ \to D^+ \overline{K}^{*0}$ and $B^+ \to D_s^+ K^{*0}$ are non-trivial diagrams in the SM, and heavily suppressed, but have similar final states. The same final states can also come from the annihilation of the constituent quarks of the B_c^+ meson. While the following chapter only discusses the search for the decay $B^+ \to D_s^+ \phi$, these other interesting decay modes are searched for in Ref. [1].

5.2 Data selection

The data sample used in this analysis amounts to $1\,\text{fb}^{-1}$ of pp collisions at 7 TeV collected by the LHCb detector in 2011. Events are selected by the trigger at hardware level if they fulfil the requirements of the L0 hadron trigger, or any track in the event fulfils any L0 trigger line requirement (`TIS`). Further trigger requirements are applied at the HLT2 level, where events are required to pass at least one of the hadronic topological triggers (see Sec. 3.3 for more details) with `TOS`.

The D_s^+ meson is only reconstructed from the Cabibbo-favoured decay $D_s^+ \to K^+K^-\pi^+$, which has a branching fraction of $(5.39 \pm 0.21) \times 10^{-2}$. Furthermore, the mass of the reconstructed particle must fall within 25 MeV of $m_{D_s^+}^{\text{PDG}} = 1968.30 \pm 0.11\,\text{MeV}$; where the superscript PDG indicates the nominal mass of the indicated particle from Ref. [21]. The D_s^+ meson decays weakly, and therefore has a non-zero lifetime, $\tau_{D_s^+} = (5.00 \pm 0.07) \times 10^{-13}\,\text{s}$, and thus an extremely narrow width, so the value of 25 MeV is primarily to account for detector resolution effects. In order for the candidate to have the correct decay topology, it is also required that the D_s^+ vertex lies downstream of the B^+ decay vertex.

Candidate ϕ mesons are reconstructed from the decay mode $\phi \to K^+K^-$, where $\mathcal{B}(\phi \to K^+K^-) = (0.498 \pm 0.005)$ and are accepted if the invariant K^+K^- mass, $m_{K^+K^-}$, is within 40 MeV of $m_\phi^{\text{PDG}} = (1019.461 \pm 0.019)\,\text{MeV}$ [21]. The ϕ meson decays strongly, and therefore has an appreciable width, $\Gamma_\phi = (4.266 \pm 0.031)\,\text{MeV}$, but the detector resolution is better for the ϕ than the D_s^+ because it has zero lifetime. Therefore, a mass window of 40 MeV is extremely wide; but further mass constraints of the ϕ are applied in the fit used to obtain the signal yield: a signal region is defined by a window that extends only 20 MeV from m_ϕ^{PDG}, and a sideband region which contains the rest.

All tracks forming the candidate mesons must fulfil requirements on the transverse momentum, $p_{\text{T}} > 100\,\text{MeV}$, and tracks from the $D_s^+(\phi)$, $p > 1(2)\,\text{GeV}$. Geometrical constraints are also placed on the tracks. One important geometrical variable is the IP, which is defined as the perpendicular distance between the vertex in question and the line of flight of a particle. The χ^2 per degree of freedom of the track fit, $\chi^2_{\text{trk}}/\text{DOF}$, must be less than four. The variable χ^2_{IP} is defined as the increase in the χ^2 of the vertex fit (χ^2_{vtx}) when the signal track is combined with the PV; the $\min(\chi^2_{\text{IP}})$ is the minimum χ^2_{IP} with respect to all PVs — this selection requires $\min(\chi^2_{\text{IP}}) > 4$. Loose PID requirements are also placed on all tracks, and further PID constraints are applied in the BDT, which is detailed later. All the χ^2 variables, with the exception of the χ^2_{trk}, are not strictly χ^2 values and do not follow χ^2 distributions. Therefore, for these values the number of degrees of freedom are irrelevant and need not be accounted for in the selection. The number of degrees of freedom is the same for all selected events though, so would simply be an additional scaling.

The B^+ vertex fit is performed by constraining the mass of the D_s^+ candidate to its known mass [21], and requiring a χ^2_{vtx} per degree of freedom of less than ten. The angle between the momentum vector of the B^+ candidate and the vector formed by the PV and decay vertex of the B^+ is known as the direction angle, θ_{dir}. Were the resolution of the LHCb detector to be perfect, a real decay would have $\cos\theta_{\text{dir}} = 1$, but here $\cos\theta_{\text{dir}}$ must be greater than 0.999. Prompt background from the PV is suppressed by requiring that the lifetime of the B^+, τ_{B^+}, is greater than 0.2 ps. A cut on the Distance of Closest Approach (DOCA) is applied to the D_s^+ candidate; where the DOCA is defined as the maximum distance of closest approach between all pairs of daughter particles. Selection criteria are summarised in Table 5.1.

Candidate	Selection criterion			
B^+	$\sum p_T^{\text{tracks}}$	$>$	5	GeV
	χ^2_{vtx}/DOF	$<$	10	
	χ^2_{IP}	$<$	25	
	τ	$>$	0.2	ps
	$\cos\theta_{\text{dir}}$	$>$	0.999	
D_s^+	$\sum p_T^{\text{tracks}}$	$>$	1.8	GeV
	χ^2_{vtx}/DOF	$<$	10	
	DOCA	$<$	0.5	mm
Tracks from D_s^+	p_{T}	$>$	100	MeV
	p	$>$	1	GeV
	χ^2_{trk}/DOF	$<$	4	
	$\min(\chi^2_{\text{IP}})$	$>$	4	
	χ^2_{FD}	$>$	36	
	$\text{DLL}_{K\pi}(K)$	$>$	-10	
	$\text{DLL}_{K\pi}(\pi)$	$<$	20	
ϕ	$\sum p_{\text{T}}{}^{\text{tracks}}$	$>$	1	GeV
	χ^2_{vtx}/DOF	$<$	16	
Tracks from ϕ	p_{T}	$>$	100	MeV
	p	$>$	2	GeV
	χ^2_{trk}/DOF	$<$	4	
	$\min(\chi^2_{\text{IP}})$	$>$	4	
	χ^2_{FD}	$>$	16	
	$\text{DLL}_{K\pi}(K)$	$>$	-2	

Table 5.1: Selection criteria applied to the $B^+ \to D_s^+ \phi$ candidates during the LHCb stripping phase.

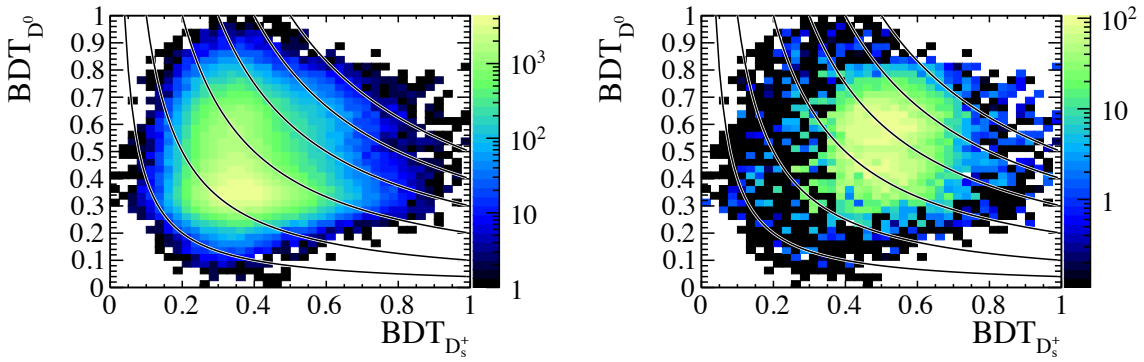


Figure 5.2: An example distribution of the response for a BDT_{D^0} against a $\mathrm{BDT}_{D_s^+}$ for a sample of (left) $B^+ \to D_s^+ \overline{D}^0$ candidates where the background dominates and occupies the lower left of the plots, and (right) the same candidates after sWeighting has been applied where the signal peak is observed further towards the upper right. The lines overlaid on the plot are show the boundaries for the cuts of $\mathrm{BDT}_{D_s^+} \times \mathrm{BDT}_{D^0} > 0.01$, 0.04, 0.10, 0.20, 0.30, 0.40, and 0.50. It is seen that using the product of the BDT discriminants is more effective for rare decays where a tighter cut will be needed; in this region candidates are selected if their response is particularly positive, at the expense of the other meson.

5.2.1 Suppression of combinatorial background

A pair of BDTs are employed to separate $D_s^+ \to K^+K^-\pi^+$ and $\phi \to K^+K^-$ candidates from combinatorial background; referred to as $\mathrm{BDT}_{D_s^+}$ and BDT_ϕ, respectively. The BDTs are designed to identify the decays $D_s^+ \to K^+K^-\pi^+$ and $\phi \to K^+K^-$, with topologies consistent with coming from a parent B-meson. The methodology used to train each BDT is the same. Both are trained using the bagging algorithm, as outlined in Chap. 4, with the StatPatternRecognition package [67], using the same set of input variables. This technique of using a BDT to identify each meson is also used in Ref. [68], which measures branching fraction ratios of various $B \to DD$ decays. The bagging boosting technique used gives a response between zero and one. Therefore, it is natural to cut on the product of the two BDT responses, $\mathrm{BDT}_{D_s^+} \times \mathrm{BDT}_\phi > X$, as opposed to $\mathrm{BDT}_{D_s^+} > X_1$ and $\mathrm{BDT}_\phi > X_2$. Cutting on the product of the BDTs improves the performance of the selection, because a very strong D_s^+ candidate (for example) will be selected at the expense of a slightly weaker ϕ selection, this is particularly true for tighter cuts. Figure 5.2 shows the effect of cutting on $\mathrm{BDT}_{D_s^+} \times \mathrm{BDT}_{D^0}$ in the normalisation channel $B^+ \to D_s^+ \overline{D}^0$.

The $D_s^+ \to K^+K^-\pi^+$ BDT was trained using $D_s^+ \to K^+K^-\pi^+$ decays from data taken from the high statistics channel $\overline{B}_s^0 \to D_s^+\pi^-$. The signal sample of D_s^+ decays came from the $D_s^+ \to K^+K^-\pi^+$ candidates that fell within 3σ of the known D_s^+ mass, and the sample was background-subtracted using the sWeighting technique [69] on the B^+ mass spectrum. Background data was taken from candidates falling in the upper mass sideband of the B^+, and either sideband of the D_s^+. The upper mass sideband is guaranteed to be

Particle		Variable
D_s^+, ϕ	Kinematic variables	p, p_{T}
	Geometric variables	χ^2_{vtx}, χ^2_{IP}, χ^2_{FD}
Tracks	Kinematic variables	p, p_{T}
	Geometric variables	$\min(\chi^2_{\mathrm{IP}})$
	Track variables	4 variables characterizing the track quality
	PID variables	16 variables containing PID information, such as `isMuon` and DLL variables from the RICH detectors

Table 5.2: List of training variables used in the D_s^+ and ϕ BDTs. Each BDT uses five variables associated with the parent particle and 23 variables from each daughter track.

pure combinatorics. Decays of higher mass states in which daughter particles are missed, misidentified, or both, may be in this sample, but they constitute real combinatorial background. Only real decays of B_s^0 mesons can contaminate this sample, specifically decay chains that contain D^+ and Λ_c^+ mesons, these are removed using the efficient vetoes as described in Sec. 5.2.2. Similarly, for the ϕ BDT, the signal sample was sWeighted and the background comes from the ϕ mass sidebands; but the sample is taken from the high statistics $B_s^0 \to J/\psi\,\phi$ mode.

In total, there are five kinematic and geometric training variables for the parent (D_s^+ or ϕ) meson. For the daughter tracks there are a total of 23 variables, including kinematic, geometric and PID variables. Since the BDT was trained using data, it is possible to use PID variables that are poorly described in simulation. The addition of the PID variables is seen to improve the performance of the discriminator, but they are not the most important variables, meaning that the background sample taken from the upper mass sideband is a real sample of random tracks, as opposed to real decays. A summary of all training variables is given in Table 5.2.

The cut for the BDT was optimised using the metric $S/\sqrt{S+B}$, In this case, the number of signal events, S, was estimated from the yield from the decay $B_s^0 \to D_s^- \pi^+$, according to:

$$S = \frac{\mathcal{B}\big(B^+ \to D_s^+ \phi\big)}{\mathcal{B}\big(\overline{B}{}_s^0 \to D_s^+ \pi^-\big)} \frac{\varepsilon_{\mathrm{gen}}\big(B^+ \to D_s^+ \phi\big)}{\varepsilon_{\mathrm{gen}}\big(\overline{B}{}_s^0 \to D_s^+ \pi^-\big)} \frac{f_d}{f_s} N\big(\overline{B}{}_s^0 \to D_s^+ \pi^-\big), \tag{5.5}$$

where f_s/f_d quantifies the fraction of B_s^0 mesons produced relative to B^0 mesons. The generator-level efficiency, $\varepsilon_{\mathrm{gen}}$, is the efficiency introduced by the acceptance region of the LHCb detector, and the necessity that all daughter particles must travel through the detector. Background yield for a given cut is estimated as:

$$B = c \cdot N_{\mathrm{c}}\big(B_s^0 \to D_s^- \pi^+\big) \cdot N_{\mathrm{c}}\big(B_s^0 \to J/\psi\,\phi\big), \tag{5.6}$$

where N_{c} indicates the yield of the combinatoric background for the indicated decay, and c is

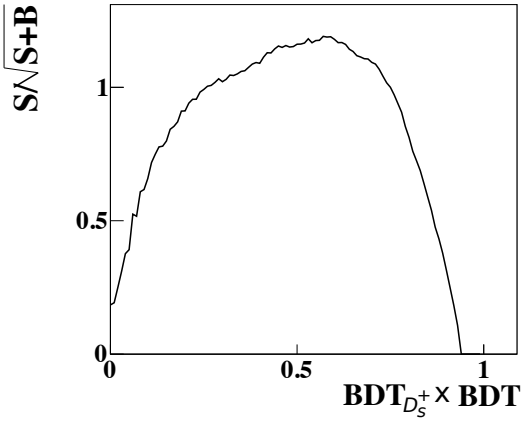


Figure 5.3: Value of the figure of merit $S/\sqrt{S+B}$ is shown as a function of the BDT response, $\mathrm{BDT}_{D_s^+} \times \mathrm{BDT}_{\phi}$. The maximum value of the figure of merit is 0.57, which is chosen as the final BDT cut.

a constant scaled such that $N_{\mathrm{c}}\big(B_s^0 \to D_s^- \pi^+\big) \cdot N_{\mathrm{c}}\big(B_s^0 \to J/\psi\,\phi\big) = N_{\mathrm{c}}\big(B^+ \to D_s^+ \phi\big)$ with no BDT cut. The optimisation procedure results in the optimal cut as $\mathrm{BDT}_{D_s^+} \times \mathrm{BDT}_{\phi} > 0.57$, as is shown in Fig. 5.3.

5.2.2 Vetoes of D^+ and Λ_c^+ decays

There are few backgrounds from real particles that could contaminate the final selection after the BDT selection. The decay topology of $D_s^+ \to K^+K^-\pi^+$ is very similar to the other weak decays $D^+ \to K^-\pi^+\pi^+$ and $\Lambda_c^+ \to pK^-\pi^+$, which both have relatively large branching fractions [21]:

$$\mathcal{B}(D^+ \to K^-\pi^+\pi^+) = (9.13 \pm 0.19) \times 10^{-2},$$

and

$$\mathcal{B}(\Lambda_c^+ \to pK^-\pi^+) = (5.0 \pm 1.3) \times 10^{-2}.$$

If the daughter $\pi^+(p)$ from the $D^+(\Lambda_c^+)$ decay is misidentified as a K^+, the decay $D_s^+ \to K^+K^-\pi^+$ can be mimicked. Henceforth, the notation h_i will be used to denote a particle assigned the mass of an h which has been identified as an i — for example, a particle identified as a kaon but under the pion mass hypothesis would be π_K. Simple generator level simulations of phasespace, as shown in Fig. 5.4, illustrate that the mass combination of $K_\pi^+K^-\pi^+$ and $K_p^+K^-\pi^+$ from a real D^+ or Λ_c^+, respectively, can fall within 25 MeV of $m_{D_s^+}^{\mathrm{PDG}}$.

The cross-feed from the D^+ and Λ_c^+ is suppressed by vetoes, whereby tight PID constraints

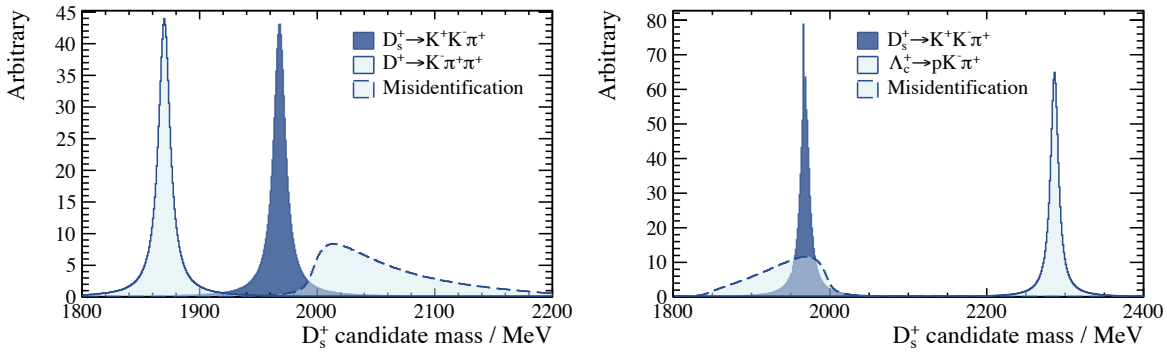


Figure 5.4: Simple phasespace simulations at generator level of the decay $D_s^+ \to K^+K^-\pi^+$, with (left) $D^+ \to K^-\pi^+\pi^+$, and (right) $\Lambda_c^+ \to pK^-\pi^+$, The distributions of the D^+ and Λ_c^+ decays, where one particle has been misidentified as a K^+ are also shown. Distributions after the misidentification are shown with a dotted outline, and sit under the D_s^+ mass peak. Shapes are normalised to the same area.

are applied if the $D_s^+ \to K^+K^-\pi^+$ candidate could have come from either a D^+ or Λ_c^+. Firstly, if the invariant mass of the K^+K^- pair lies within 10 MeV of the nominal ϕ mass then it is highly likely that it is a real D_s^+ decaying via $D_s^+ \to \phi\pi^+$, which has a branching fraction of $(4.5 \pm 0.4) \times 10^{-2}$ and therefore the $K^+K^-\pi^+$ combination is immediately accepted as a D_s^+ candidate. Secondly, if the invariant mass of the $p_K K^-\pi^+(K^-\pi_K^+\pi^+)$ object falls within 25 MeV of the known $D^+(\Lambda_c^+)$ mass the ambiguous particle is subject to harsh PID requirements, such that: $\mathrm{DLL}_{K\pi} > 10(\mathrm{DLL}_{Kp} > 0)$. These vetoes are highly, $\sim$95 %, efficient.

Invariant mass distributions of selected candidate D_s^+ and ϕ mesons after the selection are shown in Fig. 5.5.

5.3 Calculation of the branching fraction

The branching fraction of the decay $B^+ \to D_s^+\phi$ is determined with respect to the normalisation channel $B^+ \to D_s^+\overline{D}^0$, where $D^0 \to K^-\pi^+$, using

$$\mathcal{B}\big(B^+ \to D_s^+\phi\big) = \frac{N\big(B^+ \to D_s^+\phi\big)}{N\big(B^+ \to D_s^+\overline{D}^0\big)} \cdot \frac{\varepsilon\big(B^+ \to D_s^+\overline{D}^0\big)}{\varepsilon\big(B^+ \to D_s^+\phi\big)} \cdot \frac{\mathcal{B}\big(\overline{D}^0 \to K^-\pi^+\big)}{\mathcal{B}\big(\phi \to K^+K^-\big)} \cdot \mathcal{B}\big(B^+ \to D_s^+\overline{D}^0\big), \tag{5.7}$$

where N denotes a yield and ε denotes an efficiency. Using $B^+ \to D_s^+\overline{D}^0$ as the normalisation channel is advantageous because the decay topology is similar to $B^+ \to D_s^+\phi$, and the 5 % uncertainty from $\mathcal{B}(D_s^+ \to K^+K^-\pi^+)$ cancels. Reference [68] uses an almost identical selection, and finds that $N(B^+ \to D_s^+\overline{D}^0) = 5182.0 \pm 73.9$, which is taken from the fit which is shown in Fig. 5.6.

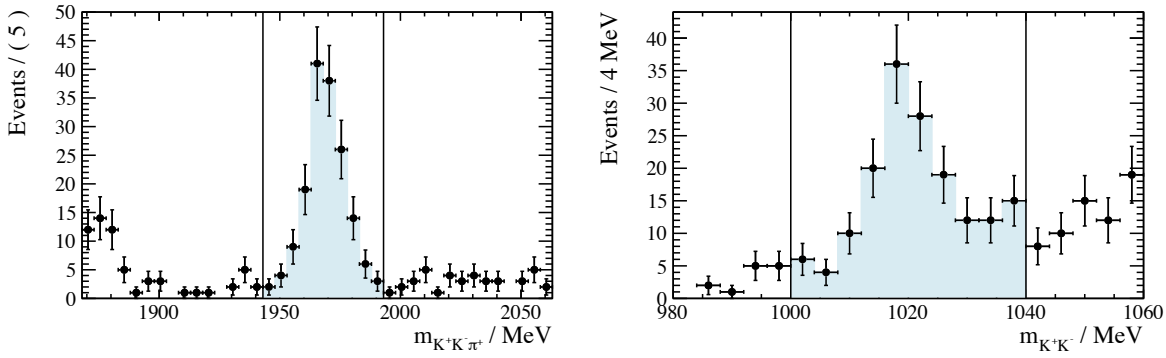


Figure 5.5: Invariant mass distributions of the candidate (left) $D_s^+ \to K^+K^-\pi^+$, and (right) $\phi \to K^+K^-$ candidates. The $K^+K^-\pi^+$ spectrum shows a range of masses, where the vertical black lines indicate the boundaries of the mass cut $|m_{K^+K^-\pi^+} - m_{D_s^+}^{\rm PDG}| < 25\,\text{MeV}$, where the shaded candidates are those that are accepted. On the far left of this distribution, a mass peak from $D^+ \to K^+K^-\pi^+$ is also visible. The $\phi \to K^+K^-$ spectrum is shown in the range $|m_{K^+K^-} - m_{\phi}^{\rm PDG}| < 40\,\text{MeV}$, and the signal region is indicated by the vertical black lines and shaded data.

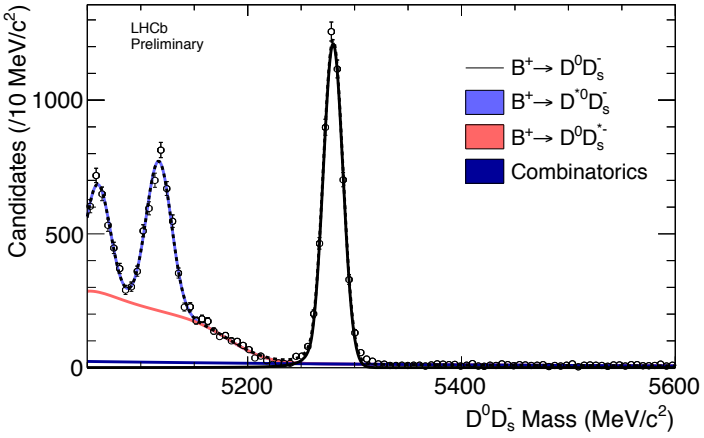


Figure 5.6: Invariant mass distribution for $B^+ \to D_s^+ \overline{D}^0$ candidates in $1\,\text{fb}^{-1}$ of LHCb data, from Ref. [68].

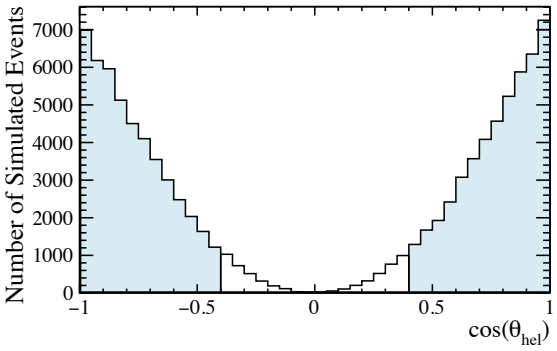


Figure 5.7: Distribution of $\cos\theta_{\rm hel}$, where $\theta_{\rm hel}$ is the angle between the B^+ and K^+ in the rest frame of the ϕ from simulated events. The shaded region indicates $|\cos\theta_{\rm hel}| > 0.4$, which defines the signal region. For the decay $B^+ \to D_s^+\phi$, where the B^+ and D_s^+ mesons are both pseudoscalars, and the ϕ is a vector, the ϕ is forced into the $j = 0$ state. Therefore, the angular distribution of $\theta_{\rm hel}$ is proportional to $\cos^2\theta_{\rm hel}$.

5.3.1 Mass fits

Both the B^+ and D_s^+ mesons have quantum numbers $J^P = 0^-$, and the ϕ is 1^- . Therefore the decay $B^+ \to D_s^+\phi$ is a transition of a pseudoscalar to a pseudoscalar and a vector meson. In order for angular momentum to be conserved, the vector particle must be produced in the $j = 0$ state, where the spin is orthogonal to the particle's momentum. Therefore, the vector ϕ must be longitudinally polarised in the final state, and its daughter kaons have an angular distribution proportional to $\cos^2\theta_{\rm hel}$, as shown in Fig. 5.7. This proves to be an excellent variable for separating signal and background, because most backgrounds are flat in $\cos\theta_{\rm hel}$. It transpired that a cut of $|\cos\theta_{\rm hel}| > 0.4$ was 93 % signal-efficient.

Fit regions are further split according to the invariant mass of the ϕ candidate. A signal region is defined for $\phi \to K^+K^-$ candidates with a mass within 20 MeV of the nominal ϕ mass, and a sideband region is defined for candidates with a mass in the range $20 < \left|m_\phi^{\rm PDG} - m_{K^+K^-}\right| < 40\,{\rm MeV}$.

The four fit regions — defined by $\theta_{\rm hel}$ and $m_{K^+K^-}$ — have a signal region, **A**, containing most of the signal, and a purely background region, **D**. Region **B** has a signal-like helicity angle, but is in the ϕ sideband region, while region **C** is the opposite. A summary of cuts defining the four fit regions is shown in Table 5.3. By simultaneously fitting $B^+ \to D_s^+\phi$ candidates in all four regions allows regions **B**, **C**, and **D** to help constrain the background in the signal region.

The signal yield of the decay $B^+ \to D_s^+\phi$ is determined with an unbinned maximum likelihood fit performed simultaneously to the invariant mass spectrum of the candidate B^+ mesons in the four regions defined above. In each region of the fit, there are several components: the signal $B^+ \to D_s^+\phi$; combinatorial background; and backgrounds from specific decays. The final state particles in this analysis of the decay $B^+ \to D_s^+\phi$ are

		$\lvert m_{K^+K^-} - m_\phi^{\rm PDG}\rvert$ (MeV)	
		$\in [0, 20]$	$\in [20, 40]$
$\lvert\cos\theta_{\rm hel}\rvert$	> 0.4	**A**	**B**
	< 0.4	**C**	**D**

Table 5.3: Definitions of fit regions used to search for the decay $B^+ \to D_s^+ \phi$. Approximately 89 % was expected to be in region **A**.

$K^+K^-\pi^+K^+K^-$; given the mass cuts on the D_s^+ and ϕ candidates, there are no sources of background which peak at the B^+ mass. However, there are backgrounds from genuine B-hadron decays in which a particle — or multiple particles — are not reconstructed, and therefore the invariant mass of the $B^+ \to D_s^+ \phi$ candidate falls below the mass of the B^+.

After the selection requirements the most significant backgrounds above 5100 MeV of these are the decays[2]: $B^+ \to D_s^{*+}\phi$, $\overline{B}{}_s^0 \to D_s^+ K^{*0} K^-$, and $\overline{B}{}_s^0 \to D_s^{*+} K^{*0} K^-$. The decay $B^+ \to D_s^{*+}\phi$ proceeds via

$$
\begin{aligned}
B^+ \to \quad & D_s^{*+} && \phi \\
& \hookrightarrow D_s^+\gamma && \hookrightarrow \boldsymbol{K^+K^-} \\
& \quad \hookrightarrow \boldsymbol{K^+K^-\pi^+}
\end{aligned}
$$

where an emboldened particle indicates a reconstructed track. There is no measured branching fraction for the decay $B^+ \to D_s^{*+}\phi$, however there is a prediction from Ref. [62] of $(3.09\,^{+1.06}_{-1.10}) \times 10^{-6}$, compared to the prediction of $\mathcal{B}\big(B^+ \to D_s^+\phi\big) = (0.13\,^{+0.06}_{-0.05}) \times 10^{-6}$ from the same reference, it is therefore likely to form a considerable background in region **A**, just below the B^+ mass after the γ is not reconstructed. The other background decay chains are:

$$
\begin{aligned}
\overline{B}{}_s^0 \to \quad & D_s^+ && K^{*0} && \boldsymbol{K^-} \\
& \hookrightarrow \boldsymbol{K^+K^-\pi^+} && \hookrightarrow \boldsymbol{K^+}\pi^- \\
\\
\overline{B}{}_s^0 \to \quad & D_s^{*+} && K^{*0} && \boldsymbol{K^-} \\
& \hookrightarrow D_s^+\gamma && \hookrightarrow \boldsymbol{K^+}\pi^- \\
& \quad \hookrightarrow \boldsymbol{K^+K^-\pi^+}
\end{aligned}
$$

where, once again, only the particles in bold are reconstructed in the vertex fit. The decay $\overline{B}{}_s^0 \to D_s^+ K^{*0} K^-$ has never been observed, but given that the branching fraction of the decay $B^0 \to D^- \overline{K}{}^{*0} K^+$ is $(8.8 \pm 1.9) \times 10^{-4}$ [21], it should be in the $B^+ \to D_s^+\phi$ selection. There is no contribution in the mass range of interest from the $B^0 \to D^- \overline{K}{}^{*0} K^+$ mode, because of the mass difference between the B_s^0 and B^0 mesons. Daughter particles from the unobserved decay $\overline{B}{}_s^0 \to D_s^{*+} K^{*0} K^-$ are also expected to be present in the sample. All these specific backgrounds are irreducible and must be accounted for in the fit.

[2] For these decays, the K^{*0} refers to the $K^*(892)^0$ meson.

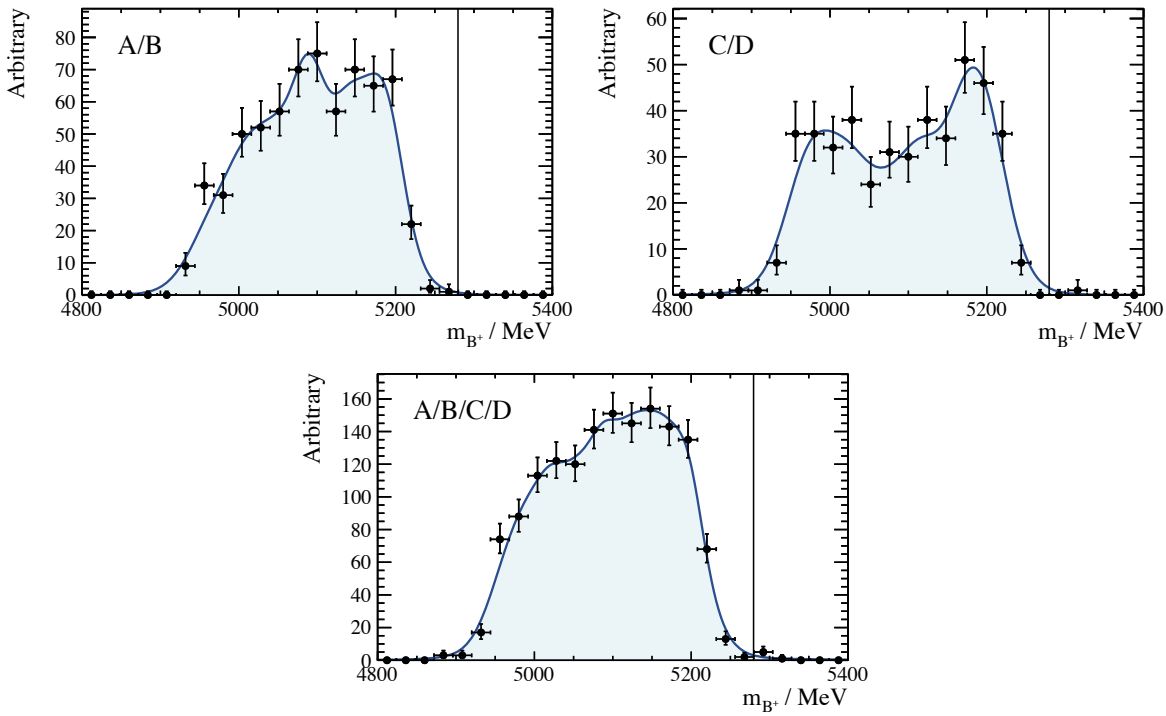


Figure 5.8: Distributions of the invariant mass distribution of simulated events of $B^+ \to D_s^{*+}\phi$ in the indicated fit regions and the resulting kernelised function. The shape of these background distributions varies depending on the helicity cut, but the rising edge on the right hand side is the same in each case. In the fit to the data, the distribution made from all events. The vertical line indicates the B^+ mass.

5.3.2 Modelling components in the mass fit

The signal shape of $B^+ \to D_s^+ \phi$ is described by a Gaussian function, with a mean μ and standard deviation σ. Combinatorial background is modelled using a decaying exponential function.

The shape of B^+ candidates originating from the $B^+ \to D_s^{*+}\phi$ background is taken from simulated events. The ϕ from the decay $B^+ \to D_s^{*+}\phi$ does not need to be longitudinally polarised because the D_s^{*+} is a vector meson ($J^P = 1^-$). Therefore the background from the decay $B^+ \to D_s^{*+}\phi$ contributes in all fit regions. Since the shapes of these reconstructed candidates are non-trivial, a kernel density estimation technique [70] is used to describe the shape. Figure 5.8 shows the kernelised distribution for the whole set of simulated events, as well as each individual helicity region. It was assumed that D_s^{*+} was unpolarised when generating events, some difference is observed in the shape of the mass distribution for this background caused by the helicity angle of the ϕ. Although there are differences, the rising edge near the B^+ mass is very similar in each region.

Backgrounds from the decay $\overline{B}{}_s^0 \to D_s^+ K^{*0} K^-$ result in highly non-trivial invariant mass shapes when the π^+ from the K^{*0} decay is missed. Once again, kernel density estimation

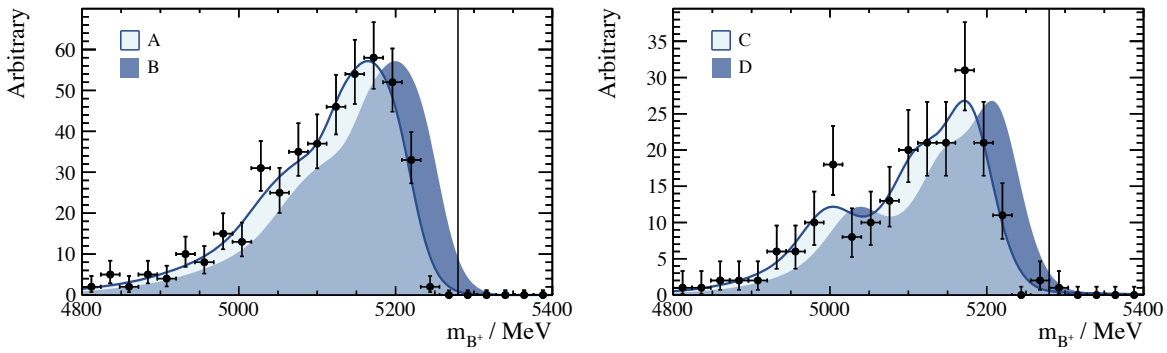


Figure 5.9: Distributions of the invariant mass distribution of simulated events of $\overline{B}{}_s^0 \to D_s^+ K^{*0} K^-$ in the (left) signal helicity regions **A** and **B**, and (right) background helicity regions **C** and **D**. The pale blue shapes at lower mass are from the signal region in $m_{K^+K^-}$, and the dark shapes are shifted up in mass by 35 MeV to model the ϕ mass sidebands.

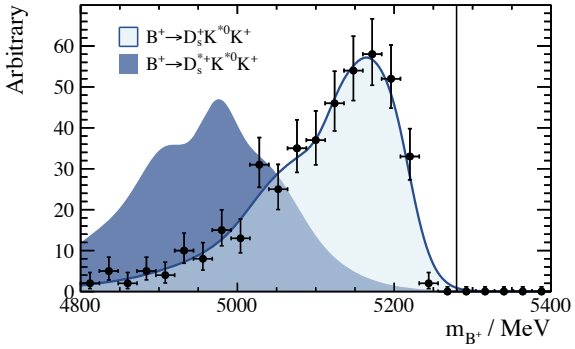


Figure 5.10: Distribution of the $\overline{B}{}_s^0 \to D_s^{*+} K^{*0} K^-$ background shape, in regions **A** and **B**, relative to that of the $\overline{B}{}_s^0 \to D_s^+ K^{*0} K^-$ from region **A**. The $\overline{B}{}_s^0 \to D_s^{*+} K^{*0} K^-$ background in **C** and **D** are the same as those shown, but shifted up in mass by 5 MeV. The shape of the $\overline{B}{}_s^0 \to D_s^{*+} K^{*0} K^-$ background is made by shifting the displayed data down in mass and smearing to account for the additional lost photon from the D_s^{*+}. This resulting shape is then kernelised. The vertical line indicates the B^+ mass.

techniques [70] are used to get an understanding of the background shape. Due to the low statistics, simulated $\overline{B}{}_s^0 \to D_s^+ K^{*0} K^-$ events falling in regions **A** and **B** are used to make the background distributions, and then shifted up by 35 MeV for regions **C** and **D**. The value of 35 MeV is an amount that was observed from data, the upwards shift is because of the shape of the S-wave under the ϕ mass peak. These shapes are shown in Fig. 5.9.

An estimate of the shape of the background distribution of $\overline{B}{}_s^0 \to D_s^{*+} K^{*0} K^-$ is estimated by taking the simulated $\overline{B}{}_s^0 \to D_s^+ K^{*0} K^-$ events, shifting them down in mass and smearing to account for the additional missing γ from the D_s^{*+} decay. The amount of shifting and smearing is the same as that seen between $B^+ \to D_s^+ \phi$ and $B^+ \to D_s^{*+} \phi$. There is 5 MeV increase in the mean of the background in the ϕ sideband regions. This shape is shown in Fig. 5.10.

5.3.3 Constraining the mass fit

The mass fit constitutes a simultaneous fit to four regions, in each of which there are four background distributions as well as a signal component (a total of at least 32 free parameters). Clearly this is very challenging, especially considering that there are low statistics and complex background models. It is therefore necessary to use relationships — derived from simulation or data — to fix as many parameters as possible. The following section will outline how the shape of each distribution is derived and how parameters are constrained.

The Gaussian function modelling the signal is determined to have a value of $\sigma = 11\,\text{MeV}$ from simulation, which is then scaled up by 20 % to account for differences in resolution between simulation and data. The mean, μ, is fixed to be 5283 MeV, which is the mean mass observed in $B^+ \to D^0\pi^+$ (and also observed in the B_s^0 mode). It is also determined from simulation that the total signal is distributed between the regions: **A** 89 %, **B** 4 %, **C** 7 %, and in **D** there is negligible expected signal contribution. Therefore, **A** and **D** will be referred to as the signal and background regions, respectively.

It was expected from simulation that ~ 7 events from the decay $B^+ \to D_s^{*+}\phi$ contribute to the background of the four fit regions, spread over $\sim 300\,\text{MeV}$. At this level, the difference in the true distributions and those shown in Fig. 5.8 — especially considering rising shape is the same for all polarisations of the ϕ — leads to a negligible difference in yield. For this reason, the kernel produced with a longitudinally polarised ϕ in $B^+ \to D_s^{*+}\phi$ is used in all regions, this makes sense in **A** and **B**, and in the other regions there is so little contribution that it makes no difference. Just as was done with the signal component, the ratios between yields of each fit region was fixed using simulation. Approximately 95 % of the contribution from the decay $B^+ \to D_s^{*+}\phi$ is expected to be in the signal region.

Yields from the decays $\overline{B}_s^0 \to D_s^+ K^{*0} K^-$ and $\overline{B}_s^0 \to D_s^{*+} K^{*0} K^-$ can, clearly, not be estimated. Especially since the yield from the decay $\overline{B}_s^0 \to D_s^+ K^{*0} K^-$ is highly sensitive to the width of the $a_1(1260)$ — because it decays strongly via $a_1(1260)^+ \to K^{*0}K^+$ — which is poorly known [21]. However, the ratios of yields in each fit regions for the decay $\overline{B}_s^0 \to D_s^+ K^{*0} K^-$ can be determined using simulated events. The ratios between the yields from regions **A**/**B** and **C**/**D** was found to be 0.5 ± 0.24, and between the regions **A**/**C** and **B**/**D** was determined to be $1.50 \pm .034$. These values are used as Gaussian constraints in the fit.

The ratio of branching fractions

$$\frac{\mathcal{B}\big(\overline{B}^0 \to D^+ K^{*0} K^-\big)}{\mathcal{B}\big(\overline{B}^0 \to D^{*+} K^{*0} K^-\big)} \sim 1.5, \tag{5.8}$$

and it is reasonable to expect the same to be true for the branching fraction ratio for the

Fit component	Parameter	Value	
$B^+ \to D_s^+ \phi$	yield **A**	6.00 ± 2.70	
	B/**A**	0.044	f
	C/**A**	0.075	f
	D/**A**	0.003 %	f
	μ	5283 MeV	f
	σ	13.2 MeV	f
$B^+ \to D_s^{*+} \phi$	yield **A**	8.67 ± 7.36	
	B/**A**	0.044	f
	C/**A**	0.00 ± 0.12	
	D/**C**	0.044	f
$\overline{B}_s^0 \to D_s^+ K^{*0} K^-$	yield **A**	4.94 ± 1.29	
	A/**B**, **C**/**D**	0.50 ± 0.24	s
	A/**C**, **B**/**D**	1.50 ± 0.34	s
$\overline{B}_s^0 \to D_s^{*+} K^{*0} K^-$	$\frac{\text{yield } \mathbf{R}}{\text{yield}(\overline{B}_s^0 \to D_s^+ K^{*0} K^-)\ \mathbf{R}}$	1.5	f
Combinatorial	yield **A**	24.0 ± 6.7	
	yield **B**	16.5 ± 6.0	
	A/**C**	1.5	f
	B/**D**	1.5	f
	exponent	$-(1.8 \pm 0.2) \times 10^{-3}$	d

Table 5.4: Fit parameters used in in the fit to determine the yield of the decay $B^+ \to D_s^+ \phi$. A label of f, means that the value is fixed in the fit; and labels of s and d mean constrained using simulated events and data over a wider mass range, respectively. The use of **R** for the constraints of $\overline{B}_s^0 \to D_s^{*+} K^{*0} K^-$ indicate that it applies in each fit region separately.

$\overline{B}_s^0 \to D_s^+ K^{*0} K^-$ and $\overline{B}_s^0 \to D_s^{*+} K^{*0} K^-$ modes. Therefore, yields of the $\overline{B}_s^0 \to D_s^{*+} K^{*0} K^-$ background is set to a factor of 1.5 less than the yield of $\overline{B}_s^0 \to D_s^+ K^{*0} K^-$ in every fit region.

The last remaining background component to be constrained is that of the combinatorial background, which is modelled with a decaying exponential function. Since the distribution of $\cos\theta_{\text{hel}}$ is flat for combinations of random tracks, the yields between the regions **A**/**C** and **B**/**D** are fixed to 1.5. The value of the slope is Gaussian constrained to a fit across a wider range of mass in data.

A summary of all constraints in the fit model, and fit yields, are given in Table 5.4.

5.3.4 Efficiency calculations

The calculation of the branching fraction $\mathcal{B}(B^+ \to D_s^+ \phi)$ requires the total efficiencies for the signal and normalisation channels. These are calculated as the product of all efficiencies

Source of efficiency	$B^+ \to D_s^+ \phi$	$B^+ \to D_s^+ \overline{D}^0$
Geometry of the LHCb detector	14.62 ± 0.05	12.75 ± 0.05
Reconstruction and stripping	1.53 ± 0.04	1.98 ± 0.04
Trigger	95.8 ± 0.3	94.4 ± 0.3
Preselection	86.0 ± 0.9	75.0 ± 0.6
BDT	51.4 ± 0.2	99.2 ± 0.2
D^+ and Λ_c^+ vetoes	95.0	95.0 ± 0.2
Total	0.091 ± 0.003	0.166 ± 0.003

Table 5.5: Efficiencies, in %, for the signal decay $B^+ \to D_s^+ \phi$ and the normalisation channel $B^+ \to D_s^+ \overline{D}^0$. The veto efficiency of $B^+ \to D_s^+ \phi$ is assumed to be the same as for $B^+ \to D_s^+ \overline{D}^0$. All efficiencies were calculated using simulated events, with the exception of the BDT, which is calculated using a data driven method, as described in text.

that contribute to the final selection. For the majority of stages in the selection process, the efficiency can be calculated using simulated events of $B^+ \to D_s^+ \phi$ and $B^+ \to D_s^+ \overline{D}^0$. However, the efficiency for the BDT cut cannot be estimated with simulation because the PID variables, upon which the BDT output depends, are poorly described by simulation.

To obtain the efficiency of BDT_X, for $X \in \{D_s^+, \phi\}$ a data driven method is used. First, the validation sample is binned in three dimensions: $p_\mathrm{T}(X)$, $\chi^2_\mathrm{FD}(X)$, and BDT_X. The bin widths are chosen such that each two dimensional bin in $(p_\mathrm{T}, \chi^2_\mathrm{FD})$ have approximately equal statistics. Here, χ^2_FD is the flight distance of X in units of χ^2. The variables p_T and χ^2_FD are used because they are two of the most discriminating variables in the BDT, and well described by simulation. The large statistics of the validation sample mean that in each bin defined by p_T and χ^2_FD there is a BDT distribution. Then, each individual simulated event is assigned an efficiency based on the BDT distribution in the bin defined by the p_T and χ^2_FD of the X. These individual efficiencies can then be amalgamated into an overall efficiency. This method is also used in Ref. [71].

A summary of all efficiencies can be found in Table 5.5.

Figure 5.11 shows the results of the fit of the signal and background functions to the $B^+ \to D_s^+ \phi$ candidates from data. The total signal yield gives a value of $N\big(B^+ \to D_s^+ \phi\big) = 6.7\,^{+4.5}_{-2.6}$, with a statistical significance of $3.6\,\sigma$, as calculated using Wilks Theorem [72]. This is confirmed with an ensemble of 10^5 toy datasets.

The branching fraction of the decay $B^+ \to D_s^+ \phi$ was determined to be

$$\mathcal{B}\big(B^+ \to D_s^+ \phi\big) = \big(1.87\,^{+1.25}_{-0.73}(\mathrm{stat}) \pm 0.19(\mathrm{syst}) \pm 0.32(\mathrm{norm})\big) \times 10^{-6}, \qquad (5.9)$$

where stat, syst, and norm refer to uncertainties introduced by statistics, systematics, and the normalisation channel, respectively. This assumes the branching fraction values of $\mathcal{B}\big(B^+ \to D_s^+ \overline{D}^0\big) = \big(1.00 \pm 0.17\big) \times 10^{-2}$, $\mathcal{B}\big(\overline{D}^0 \to K^- \pi^+\big) = \big(3.88 \pm 0.05\big) \times 10^{-2}$, and

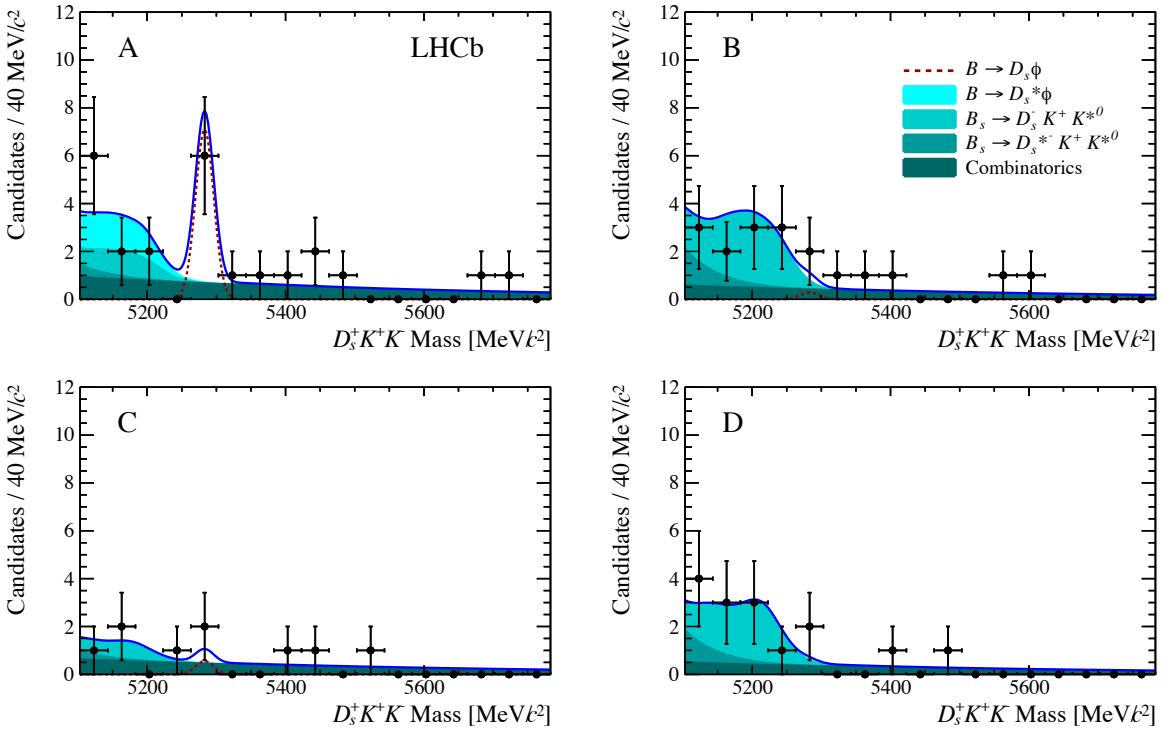


Figure 5.11: Fits to the four analysis regions, as given in Table 5.3, in the search for the decay $B^+ \to D_s^+ \phi$. The region **A** is contains the majority of the signal candidates, while region **D** contains none.

$\mathcal{B}(\phi \to K^+K^-) = (48.9 \pm 0.5) \times 10^{-2}$ [21]. The uncertainty of the total branching fraction of the normalisation channel introduced a systematic uncertainty of 17 %. Other systematic uncertainties included in the above result are discussed in Sec. 5.4.

5.4 Systematic uncertainties

Sources of systematic uncertainty appear at all levels of the selection and modelling of the decay. The following section lists each contribution and defends the choice of the value of the uncertainty assigned. It should be noted that it is only the relative efficiency between the signal and normalisation channels that contributes to the systematics, rather than their absolute values.

The selection of candidate $B^+ \to D_s^+ \phi$ decays is the source of some small uncertainty. It is known that the geometry of the detector and the reconstruction process is well described in simulation. As described above, the efficiency of the BDT is evaluated using data, so there is a negligible difference in the efficiency ratio. In total, a 1 % uncertainty is assigned for the selection. It is important to check the validity of the choice of the optimised BDT cut value. This was investigated by optimising the BDT using the newly measured value of $\mathcal{B}(B^+ \to D_s^+ \phi)$, resulting in a looser cut than used in the analysis. It is seen that using

recalculated BDT cut value has negligible effect on the resulting branching fraction.

There are some differences between the performance of the trigger in data and in the simulation. The HLT lines are observed to be described well in simulation, as should be expected, since they are software level triggers. Also, it is observed that the L0 hardware trigger is reliable above $p_{\mathrm{T}} > 4\,\mathrm{GeV}$, but slightly mismodelled at low p_{T}. Using only the events above 4 GeV, it is observed that the trigger efficiency alters by about 4 %. This is the value of the systematic uncertainty that is applied.

Both the signal and normalisation channels have a $D_s^+ \to K^+K^-\pi^+$, and therefore efficiencies from the D^+ and Λ_c^+ vetoes cancel to a large extent. Given that these cuts are very efficient, the difference is negligible, and the efficiency ratio is assumed to be unity for the calculation. The only differences between the two modes are due the decay kinematics, since the ϕ is lighter than the D^0. There is a difference of 1.5 % between $\varepsilon_{\mathrm{veto}}$ for the decays $\overline{B}{}_s^0 \to D_s^+\pi^-$ and $B^+ \to D_s^+\overline{D}{}^0$, and since the mass of the ϕ is nearer the D^0 mass than the pion mass, a systematic uncertainty of 1 % is assigned.

There are other systematic uncertainties that affected the selection. The BDT cut is assigned an uncertainty of 3 %, which is due to the sizes of the validation samples used to calculate the efficiencies. Mass windows around the D_s^+ and ϕ lead to a 3 % systematic uncertainty. Also, the low statistics of the simulation samples used to deduce efficiencies led to a 3 % uncertainty.

The mass fits introduced systematic uncertainties from each component. If the parameters describing the signal shape are allowed to float the fit results in a yield which is 5% higher than the nominal fit. This is assigned as a systematic uncertainty.

The total uncertainty from the background shape is 5 %; and is estimated by making changes to the background model. By removing either the $B^+ \to D_s^{*+}\phi$ or $\overline{B}{}_s^0 \to D_s^{*+}K^{*0}K^-$ components, the yield changes by only 1 %. Changing the constraints on $\mathbf{A}/\mathbf{B}$ and $\mathbf{C}/\mathbf{D}$ for $\overline{B}{}_s^0 \to D_s^+K^{*0}K^-$ by a factor of 2 results in a 1 % change in signal yield. The combinatorial background is estimated by allowing the slope to float free, this led to an approximate 3.5 % systematic uncertainty.

Contributions from all sources of systematic uncertainties are summarised in Table 5.6. The dominant systematic uncertainty — discounting the uncertainties on the branching fraction of the normalisation channel — is from the mass fits, which is unsurprising considering the complexity of the fit, and the treatment of the backgrounds. Regardless of the constraints that are, or are not, included in the fit: the lowest significance obtained is still greater than $3\,\sigma$. Therefore the significance is quoted as greater than $3\,\sigma$.

Source of systematic	Uncertainty (%)
Selection	1
Trigger	4
BDT	3
D^+ and Λ_c^+ vetoes	1
Mass windows	3
Simulation statistics	3
Mass shape	5
Background shapes	5
Total	10

Table 5.6: Summary of the systematic uncertainties contributing to the branching fraction of the decay $B^+ \to D_s^+ \phi$.

5.5 Direct *CP* asymmetry

The *CP* asymmetry is defined in Eq. 5.4, but this must be modified to account for background in the sample:

$$\mathcal{A}_{CP}\big(B^+ \to D_s^+ \phi\big) = \frac{N(B^- \to D_s^- \phi) - N(B^+ \to D_s^+ \phi)}{N(B^- \to D_s^- \phi) + N(B^+ \to D_s^+ \phi) - N_{\text{bkg}}}. \tag{5.10}$$

Here, the yields, N, refer to the decay with the indicated charge, and N_{bkg} denotes the amount of background contaminating the signal.

Values of N are determined by defining a signal region, which extends $2\,\sigma$ either side of the mean B^+ mass in region **A**, where $\mu = 5283\,\text{MeV}$ and $\sigma = 13.2\,\text{MeV}$. In this region, the number of B^+ and B^- candidates are $N(B^+ \to D_s^+ \phi) = 3$ and $N(B^- \to D_s^- \phi) = 3$. The background prediction is determined by integrating across this region, which results in $N_{\text{bkg}} = 0.75$. From these values, $\mathcal{A}_{CP}^{\text{raw}}\big(B^+ \to D_s^+ \phi\big) = \big(0.0 \pm 0.41\big)$, where the coverage was obtained using the Feldman-Cousins method [73]; which gives a $1\,\sigma$ interval $\in [-0.41, 0.41]$. Figure 5.12 shows the Feldman-Cousins intervals for the *CP* asymmetry when observing six signal events with a background expectation of 0.75.

In order to find the final value of $\mathcal{A}_{CP}$, the *CP* asymmetries caused by production ($\mathcal{A}_{CP}^{\text{prod}}$), detection ($\mathcal{A}_{CP}^{\text{det}}$), and selection ($\mathcal{A}_{CP}^{\text{sel}}$) must be corrected for.

The B production asymmetry has been estimated by LHCb for the decays $B^+ \to J/\psi\,\pi^+$ [74] and $B^+ \to D^0 K^+$ [71] to be $-(0.3 \pm 0.9)\,\%$ and $-(0.8 \pm 0.7)\,\%$, respectively. For the decay $B^+ \to D_s^+ \phi$, the only source of detection asymmetry is the pion from the D_s^+ decay. This has been shown to be a very small effect, but the edges of the detector preferentially selects one charge or the other depending on the magnet polarity, and because the dataset collected in 2011 is approximately 30% larger in one polarity than the other, there could conceivably be an effect. Therefore, from these values an estimate of the total $\mathcal{A}_{CP}^{\text{prod}} + \mathcal{A}_{CP}^{\text{det}} = -(1 \pm 1)\,\%$

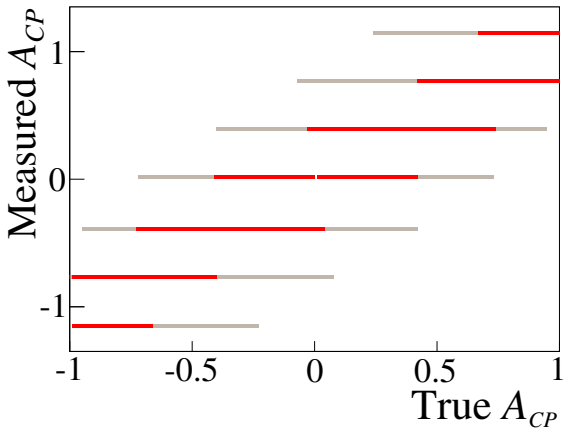


Figure 5.12: Feldman-Cousins intervals for the *CP* asymmetry, when six signal events are observed, and the background expectation is 0.75. The red stripes indicate $1\,\sigma$ intervals, and the grey are $2\,\sigma$.

is used. Although there are no selection requirements that preferentially select one charge or the other, including input variables to the BDT, the selection did show a slight bias. The selection asymmetry was given a conservative value of $(2 \pm 3)\,\%$.

Corrections to the $\mathcal{A}_{CP}^{\text{raw}}$ total to $(1 \pm 3)\,\%$, where the error here is more than a factor of ten less than the statistical uncertainty. Accounting for these corrections results in

$$\begin{aligned} \mathcal{A}_{CP}\big(B^+ \to D_s^+ \phi\big) &= \mathcal{A}_{CP}^{\text{raw}} - \big(\mathcal{A}_{CP}^{\text{prod}} + \mathcal{A}_{CP}^{\text{det}} + \mathcal{A}_{CP}^{\text{sel}}\big) \\ &= -\big(0.01 \pm 0.41(\text{stat}) \pm 0.03(\text{syst})\big), \end{aligned} \tag{5.11}$$

which is consistent with no observable *CPV*, as expected in the SM.

5.6 Summary

The analysis outlined in the preceding chapter shows first evidence for the decay $B^+ \to D_s^+ \phi$, with greater than $3\,\sigma$ significance. While the measured branching fraction

$$\mathcal{B}\big(B^+ \to D_s^+ \phi\big) = \big(1.87\,^{+1.25}_{-0.73}(\text{stat}) \pm 0.19(\text{syst}) \pm 0.32(\text{norm})\big) \times 10^{-6}.$$

is somewhat higher than SM predictions, they are not incompatible when considering large theoretical uncertainties. Another measurement was the *CP* asymmetry, which could deviate significantly from zero were NP to be present at leading order. The value measured was

$$\mathcal{A}_{CP}\big(B^+ \to D_s^+ \phi\big) = -\big(0.01 \pm 0.41(\text{stat}) \pm 0.03(\text{syst})\big),$$

which is consistent with the SM expectation of zero *CPV*.

Chapter 6

Search for the decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$

6.1 Introduction

The decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$ both are $b \to s\mu^+\mu^-$ FCNC transitions, which are forbidden at tree-level in the SM[1]. Therefore, these processes are sensitive to virtual NP particles contributing to the decay amplitude in loops. The analysis described in the following chapter makes a measurement of the branching fractions of both decays and the differential branching fraction of $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ in bins of q^2, where q^2 is the invariant mass of the dimuon system squared. This analysis was published in Ref. [2].

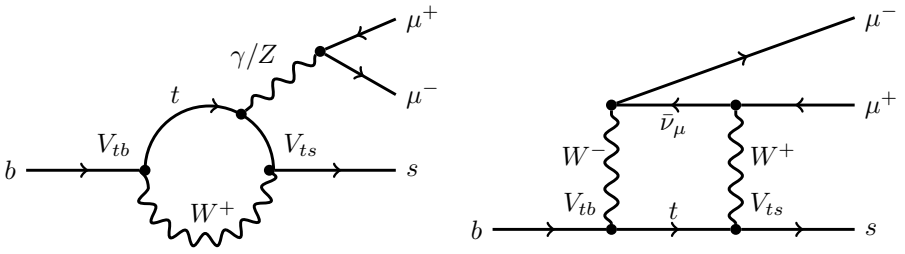


Figure 6.1: Schematic Feynman diagrams for the operators $\mathcal{O}_7$, $\mathcal{O}_9$, and $\mathcal{O}_{10}$ which are most sensitive to the $b \to s\mu^+\mu^-$ FCNC. The propagators are the (left) photonic and Z penguin diagram, and the (right) W^+-mediated box diagram. Operator $\mathcal{O}_7$ describes the photonic penguin diagram; while $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are the vector and axial-vector parts of both the Z and W^+ diagrams.

[1] All mentions of the ϕ refer implicitly to the $\phi(1020)$ meson.

Transitions of the FCNC $b \to s\ell^+\ell^-$ can be expressed with the effective Hamiltonian

$$\mathcal{H}_{\text{eff}} = -4\frac{G_F}{\sqrt{2}}V_{ts}^*V_{tb}\frac{e^2}{16\pi^2}\sum_{i=1}^{10}\big[C_i(\Lambda)\mathcal{O}_i(\Lambda) + C_i'(\Lambda)\mathcal{O}_i'(\Lambda)\big]. \tag{6.1}$$

Operators which are particularly sensitive to NP contributions in $b \to s\mu^+\mu^-$ transitions are

$$\begin{aligned}
\mathcal{O}_7 &= \frac{m_b}{e}\big(\bar{s}\sigma_{\mu\nu}P_R b\big)F^{\mu\nu} & \mathcal{O}_7' &= \frac{m_b}{e}\big(\bar{s}\sigma_{\mu\nu}P_L b\big)F^{\mu\nu} \\
\mathcal{O}_9 &= \big(\bar{s}\gamma_\mu P_L b\big)\big(\bar{\ell}\gamma^\mu\ell\big) & \mathcal{O}_9' &= \big(\bar{s}\gamma_\mu P_R b\big)\big(\bar{\ell}\gamma^\mu\ell\big) \\
\mathcal{O}_{10} &= \big(\bar{s}\gamma_\mu P_L b\big)\big(\bar{\ell}\gamma^\mu\gamma_5\ell\big) & \mathcal{O}_{10}' &= \big(\bar{s}\gamma_\mu P_R b\big)\big(\bar{\ell}\gamma^\mu\gamma_5\ell\big).
\end{aligned} \tag{6.2}$$

The operator $\mathcal{O}_7$ describes the transition of $b \to s\gamma$; in the SM this is via a photonic penguin diagram. Operators $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are the vector and axial-vector components of the four point $b \to s\mu^+\mu^-$ interaction; in the SM these operators are made up of the Z mediated penguin, and the W^+ box diagrams. Figure 6.1 shows schematic Feynman diagrams for $\mathcal{O}_7$, $\mathcal{O}_9$, and $\mathcal{O}_{10}$. Primed operators are the suppressed helicity (usually right-handed), whose contributions are vanishingly small in the SM. The operators $\mathcal{O}_{1-6}$ encapsulate long distance contributions, such as $c\bar{c}$ loops, and $\mathcal{O}_8$ is the gluonic penguin operator.

Both decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$ proceed via a $b \to s\mu^+\mu^-$ FCNC with a spectator u quark. The diagrams describing these decays mediated by penguin loops are shown in Fig. 6.2. It should be noted that the B^+ and K^+ in both these decays have the same charge assignment. In this way, both these decays are the same as the FCNC decay $B^0 \to K^{*0}\mu^+\mu^-$, but differ in the $q\bar{q}$ pairs that are popped from the QCD field.

There is significant interest in the decay $B^0 \to K^{*0}\mu^+\mu^-$ because it has access to a range of angular observables, many of which are sensitive to new physics. The angular observable P_5' is a parameter formed of combinations of K^{*0} decay amplitudes, and designed to have reduced theoretical uncertainties: being nearly free of form-factor dependence [75]. A measurement from LHCb indicates that P_5' is $3.7\,\sigma$ above predictions in the region $4.0 < q^2 < 8.0\,\text{GeV}^2$ [76], indicating that C_9 is low with respect to the SM [77]. This either points to a new vector contribution or vastly misunderstood QCD effects. Analyses of other decays with similar topologies and simpler final states are also interesting. For example, a measurement from LHCb of the differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$ [78] shows some deviation from the SM prediction.

In the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$, the structure of the $K^+\pi^+\pi^-$ system's mass distribution results from the decay of a variety of strange resonances. Contributions of resonances to the $K^+\pi^+\pi^-$ system has been previously studied by the Belle collaboration in the

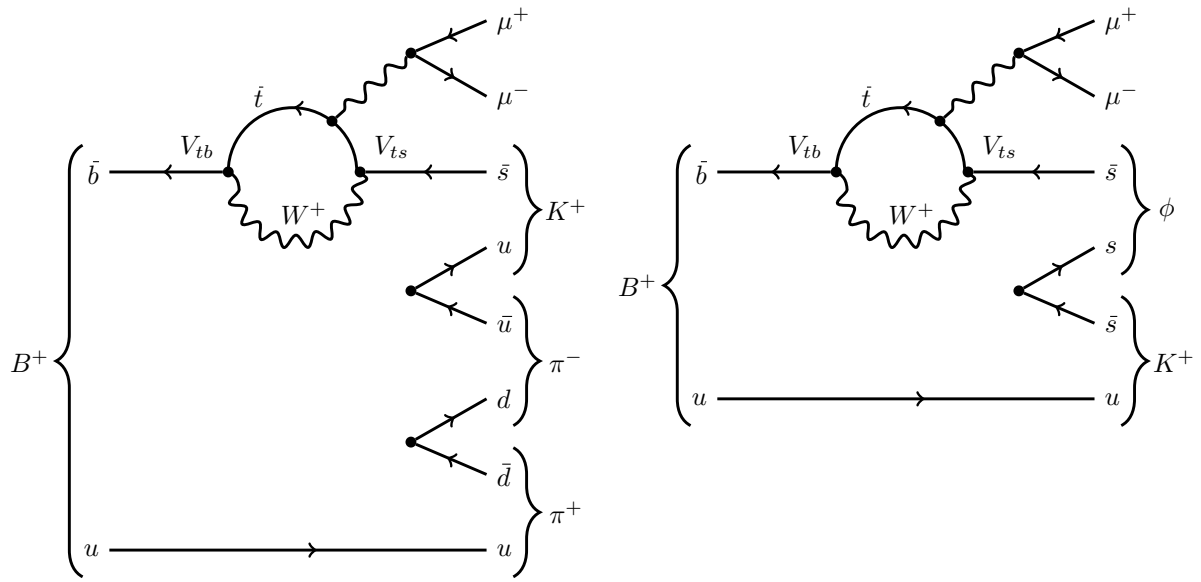


Figure 6.2: Feynman diagrams illustrating how the decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$ can proceed in the SM via operators $\mathcal{O}_7$, $\mathcal{O}_9$ and $\mathcal{O}_{10}$. The operators $\mathcal{O}_{1-6}$ also contribute a small amount.

tree-level decay $B^+ \to J/\psi\, K^+\pi^+\pi^-$, where $J/\psi \to \mu^+\mu^-$ [79]. This study indicated that the dominant contribution to the $K^+\pi^+\pi^-$ system should be expected to be from decaying $K_1(1270)^+$ mesons. The total branching fraction of the decay $K_1(1270)^+ \to K^+\pi^+\pi^-$ is $\mathcal{B}\big(K_1(1270)^+ \to K^+\pi^+\pi^-\big) = \big(35.7 \pm 3.7\big)\,\%$ [21]. This value includes the non-resonant mode, and the decays via variaous resonances including $K_1(1270)^+ \to K^+\rho(\to \pi^+\pi^-)$ and $K_1(1270)^+ \to \pi^+K^{*0}(\to K^+\pi^-)$. The $K_1(1270)^+$ meson, together with the $K_1(1400)^+$, are mass eigenstates resulting from the mixing of the P-wave axial vector states K_{1A} and K_{1B} according to:

$$\begin{pmatrix} |K_1(1270)^+\rangle \\ |K_1(1400)^+\rangle \end{pmatrix} = \begin{pmatrix} \sin\theta_{K_1} & \cos\theta_{K_1} \\ \cos\theta_{K_1} & -\sin\theta_{K_1} \end{pmatrix} \begin{pmatrix} |K^+_{1A}\rangle \\ |K^+_{1B}\rangle \end{pmatrix}. \tag{6.3}$$

Here, θ_{K_1} is the mixing angle and has been measured to have central values of both -34° and -57° [80–85]. However, more recent measurements favour a value of $-(34\pm13)^\circ$ [82–85], and the most recent rule out the solution at -57° entirely [84, 85] (but with a different sign convention).

Due to the unknown composition of the $m_{K^+\pi^+\pi^-}$ spectrum, an inclusive prediction of the branching fraction $\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)$ does not exist. However, the branching fraction of the rare decay $B^+ \to K_1(1270)^+\mu^+\mu^-$ is predicted to be [86]

$$\mathcal{B}\big(B^+ \to K_1(1270)^+\mu^+\mu^-\big) = \big(2.3\,^{+1.3\,+0.0}_{-1.0\,-0.2}\big) \times 10^{-6}, \tag{6.4}$$

where uncertainties arise from form-factor calculations and the mixing angle, respectively.

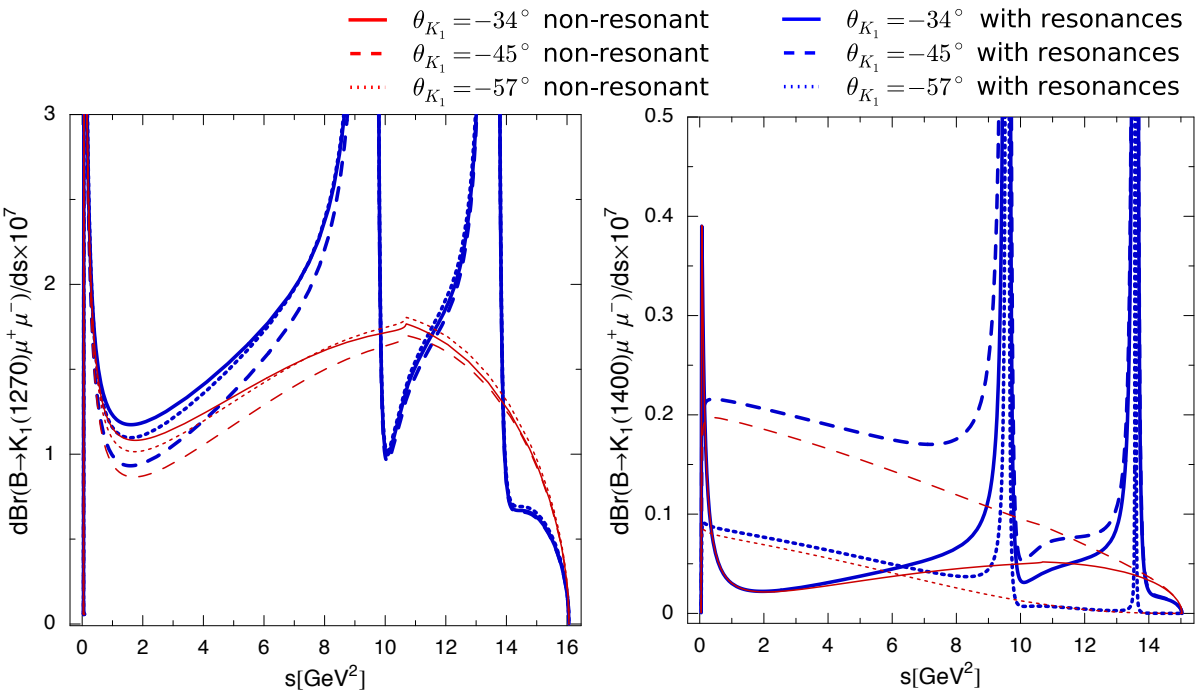


Figure 6.3: The dimuon invariant mass distributions for the differential decay rates, as taken from Ref. [86]. $\mathrm{d}\mathcal{B}(B^+ \to K_1^+\mu^+\mu^-)/\mathrm{d}q^2$, (where $s = q^2$), for the $K_1(1270)^+$ and $K_1(1400)^+$. Central values of input form factors are used. The thick blue lines and thin red curves indicate the differential branching fractions with and without corrections from resonances, respectively. Solid, dotted and dashed lines correspond to values of the mixing angle $\theta_{K_1} = -34°, -45°, -57°$ respectively.

Just as with theoretical predictions of the branching fraction $B^+ \to D_s^+\phi$, QCD makes predictions very difficult, giving fractional errors of 50 %.

Figure 6.3 shows the theoretical q^2 distribution for the decay $B^+ \to K_1^+\mu^+\mu^-$, for both the $K_1(1270)^+$ and $K_1(1400)^+$ and varying θ_{K_1}. The $b \to s\mu^+\mu^-$ can be mediated by a virtual photon which, for some values of θ_{K_1}, can be transversely polarised. However, for some values of θ_{K_1}, the $\mu^+\mu^-$ pair is fully longitudinally polarised and the decay via a photon is forbidden.

There are no predictions for the branching fraction of the decay $B^+ \to \phi K^+\mu^+\mu^-$, but one would expect it to be smaller than that of the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$. This is because it requires an $s\bar{s}$ to be created from the QCD field rather than a $u\bar{u}$ and $d\bar{d}$ pair.

It is possible to use some other decays to get an estimate using previously measured branching fractions. The decays $B^+ \to K^+\pi^+\pi^-$ and $B^+ \to \phi K^+$ proceed, at leading order, in the same way as $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$; being propagated

by gluonic penguins. These fully hadronic branching fractions are measured to be

$$\begin{aligned}\mathcal{B}(B^+ \to K^+\pi^+\pi^-) &= (5.10 \pm 0.29) \times 10^{-5} \\ \mathcal{B}(B^+ \to \phi K^+) &= (8.8\,^{+0.7}_{-0.6}) \times 10^{-6}.\end{aligned} \tag{6.5}$$

One would expect that the ratio of branching fractions to be approximately the same

$$\frac{\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)}{\mathcal{B}(B^+ \to \phi K^+\mu^+\mu^-)} \simeq \frac{\mathcal{B}(B^+ \to K^+\pi^+\pi^-)}{\mathcal{B}(B^+ \to \phi K^+)} = 5.80\,^{+0.6}_{-0.5}. \tag{6.6}$$

Absent from this analysis are the searches for the decays $B^+ \to K^+K^-\pi^+\mu^+\mu^-$ and $B^+ \to \pi^+\pi^-\pi^-\mu^+\mu^-$. These were not included because they are suppressed by a factor of $|V_{td}/V_{ts}|^2 \simeq 23$ with respect to the decays that are studied. The more interesting of the two decays is $B^+ \to \pi^+\pi^-\pi^-\mu^+\mu^-$ because this has the higher branching fraction, and so the easiest access to the ratio $|V_{td}/V_{ts}|$ using $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$; however, it suffers from large backgrounds because of the number of pions in the final state.

6.2 Selection

Signal candidates for the decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$ must first pass the L0 muon trigger. Subsequent software trigger stages required that at least one final-state muon has $p_{\rm T} > 1.0\,\text{GeV}$ and at least one hadron has $p_{\rm T} > 1.6\,\text{GeV}$, both of which have an IP larger than $100\,\mu\text{m}$ with respect to any V in the event. The response of the topological BBDT, which was described in Sec. 3.3, in HLT2 must be consistent with a decaying B meson with muons in the final state.

Candidate B^+ hadrons are then formed from combinations of three hadrons and a pair of opposite sign muons. Fully reconstructed candidates must form a good quality vertex, with a χ^2 of the vertex fit < 6. This secondary vertex must be well displaced from any V, having a flight distance inconsistent with zero, $\chi^2_{\rm FD} > 121$. Each track must satisfy $\chi^2_{\rm IP} > 16$, where the $\chi^2_{\rm IP}$ of a track is defined as the change in $\chi^2_{\rm IP}$ when calculated with and without the track in question. The muons must both satisfy the `isMuon` criteria and have $\text{DLL}_{\mu\pi} > 0$, while PID criteria for hadrons are applied later. Each hadron must have $p_{\rm T} > 500\,\text{MeV}$ and the total invariant mass of the $K^+\pi^+\pi^-$ system is required to be in the range $750 < m_{K^+\pi^+\pi^-} < 2400\,\text{MeV}$ in order to reduced the rate of accepted events at stripping level. For the ϕK^+ system, in additional constraint is that the $\phi \to K^+K^-$ object must have an invariant mass within $12\,\text{MeV}$ of $m_\phi^{\rm PDG}$.

There are other decay channels which are used throughout this analysis, specifically: $B^+ \to J/\psi\, K^+\pi^+\pi^-$, $B^+ \to \psi(2S)K^+$, and $B^+ \to J/\psi\,\phi K^+$. In all these decays

Candidate	Selection criterion			
B^+	χ^2_{vtx}	$<$	6.0	
	χ^2_{IP}	$<$	16.0	
	χ^2_{FD}	$>$	121.0	
$K^+\pi^+\pi^-$	Vtx χ^2	$<$	12.0	
	$m_{K^+\pi^+\pi^-}$	$\in$	$[750, 2400]$	MeV
	χ^2_{IP}	$<$	4.0	
	χ^2_{FD}	$>$	25.0	
$\mu^+\mu^-$	χ^2_{vtx}	$<$	12.0	
	χ^2_{FD}	$>$	81.0	
tracks	χ^2_{IP}	$>$	16.0	
	track χ^2/ndf	$<$	2.5	
K^+, π^+	p_T	$>$	500	MeV
μ^+	`isMuon`			

Table 6.1: Selection criteria applied to candidates in the LHCb stripping phase. The definitions of χ^2 values can be found in the text in this section or in Sec. 5.2.

$\psi(2S) \to J/\psi\,\pi^+\pi^-$ and $J/\psi \to \mu^+\mu^-$. These are each selected by requiring that both J/ψ and $\psi(2S)$ candidates have an invariant mass within 50 MeV of their known masses [21].

6.2.1 Background contributions

The tree level decays $B^+ \to J/\psi\,K^+\pi^+\pi^-$ and $B^+ \to J/\psi\,\phi K^+$, where $J/\psi \to \mu^+\mu^-$, have large branching fractions:

$$\mathcal{B}\big(B^+ \to J/\psi\,K^+\pi^+\pi^-\big) \cdot \mathcal{B}\big(J/\psi \to \mu^+\mu^-\big) = (4.8 \pm 0.8) \times 10^{-5} \tag{6.7}$$

$$\mathcal{B}\big(B^+ \to J/\psi\,\phi K^+\big) \cdot \mathcal{B}\big(J/\psi \to \mu^+\mu^-\big) = \big(3.1 \pm 1.1\big) \times 10^{-6}, \tag{6.8}$$

and the same final state particles as the signal modes. They therefore constitute peaking backgrounds that lie under the signal peak. The same is true for the large contributions from $B^+ \to \psi(2S)K^+\pi^+\pi^-$ and $B^+ \to \psi(2S)\phi K^+$ decays, where $\psi(2S) \to \mu^+\mu^-$. These charmonium decays are large irreducible backgrounds that must be removed with vetoes around the J/ψ and $\psi(2S)$ masses. The vetoes used remove events where the invariant dimuon mass falls in either region $2946 < m_{\mu^+\mu^-} < 3176\,\text{MeV}$ or $3586 < m_{\mu^+\mu^-} < 3766\,\text{MeV}$.

Figure 6.4 shows the boundaries defined by these vetoes on data. Mis-reconstructed decays to charmonium contribute to the upper mass sideband. To remove these, the veto windows are extended up by 40 MeV in the region $5330 < m_{K^+\pi^+\pi^-\mu^+\mu^-} < 5450\,\text{MeV}$. Radiative tails from the decays $J/\psi \to \mu^+\mu^-\gamma$ and $\psi(2S) \to \mu^+\mu^-\gamma$ are suppressed by extending the

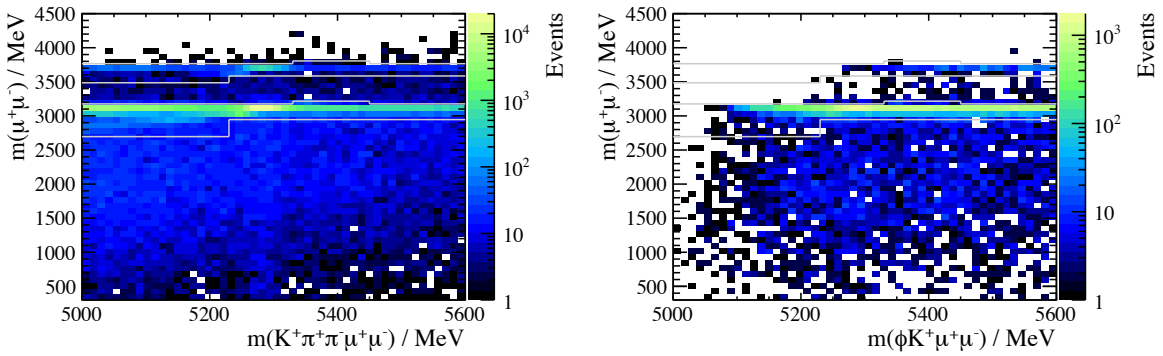


Figure 6.4: The variation of the invariant mass of the dimuon candidate with the mass of the B^+ candidate for (left) $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$, and (right) $B^+ \to \phi K^+\mu^+\mu^-$. The grey lines indicate the boundaries of the charmonium vetoes.

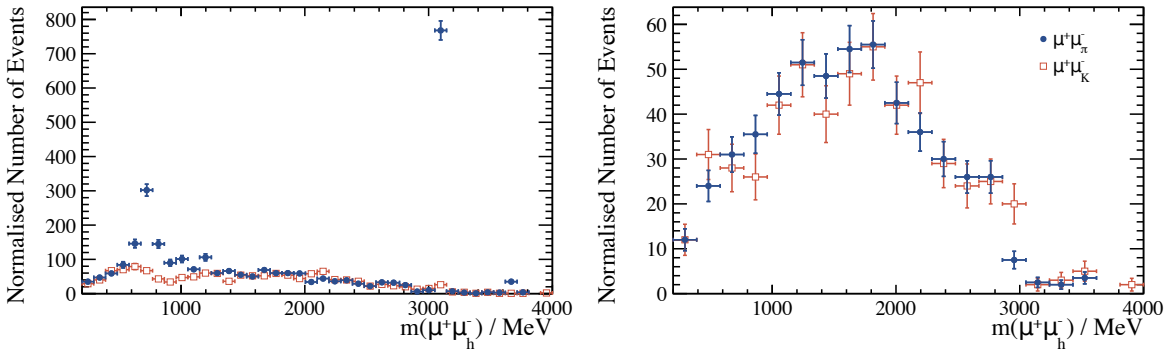


Figure 6.5: Invariant mass of the combination of a muon and a reconstructed hadron of the opposite charge under the muon mass hypothesis (left) before, and (right) after the veto in the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$. Both pion charges are included in the histogram with solid markers. There is a feature at low mass in the $m(\mu^+\mu^-_\pi)$ spectrum which is removed by the vetoes, and originates from the background decay $B^+ \to J/\psi\,\rho(770)^0 K^+$.

vetoes down by 250 MeV and 100 MeV respectively in the region $m_{K^+\pi^+\pi^-\mu^+\mu^-} < 5230\,\text{MeV}$.

Considering the large branching fractions of the charmonium decays given above, and the probability of misidentifying a pion as a muon is $\mathcal{O}(1\,\%)$ [87] (somewhat less for a kaon) there could be significant contamination from mis-identified candidates. This background was removed by calculating the invariant mass of each $\mu^+\pi^-$ and μ^+K^- combination, where the hadron was assigned the muon mass. If the mass of this object fell within 50 MeV of $m_{J/\psi}$ or $m_{\psi(2S)}$, then the candidate was vetoed. Figure 6.5 shows the effect on these vetoes, and demonstrates that a large part of the background that is removed by these vetoes is from the decay $B^+ \to J/\psi\,\rho(770)^0 K^+$.

Other background contributions can come from doubly misidentified fully hadronic decays, such as the decay $B^0 \to \overline{D}{}^0\pi^+\pi^+\pi^-$ followed by $\overline{D}{}^0 \to K^+\pi^-$, which has a total branching fraction of $(2.21 \pm 0.85) \times 10^{-4}$ [21]. This decay has a final state of $K^+\pi^+\pi^-\pi^+\pi^-$. If a pair of opposite sign pions are misidentified as muons, the final state of $K^+\pi^+\pi^-\mu^+\mu^-$

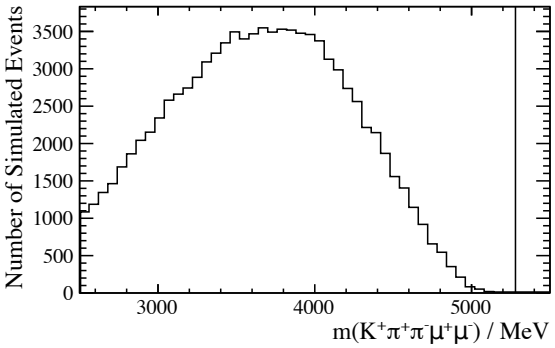


Figure 6.6: Background from the cascade decay $B^+ \to D^-\pi^+\mu^+\nu_\mu$, where $D^- \to K^+\pi^-\mu^-\bar{\nu}_\mu$ does extend as high in mass as the nominal B^+ mass, due to lost energy from neutrinos. The vertical black line indicates the nominal mass of the B^+ meson.

is mimicked, and therefore should be considered as a background. The same can be said for the decay $B^0 \to \overline{D}^0\pi^+$ and $\overline{D}^0 \to K^-\pi^+\pi^+\pi^-$, which has a total branching fraction of $(3.88 \pm 0.15) \times 10^{-4}$.

In order to ascertain if these charmed decays are potential backgrounds, invariant mass distributions are inspected for evidence of D^0 mesons. The only distribution in which there is evidence of contamination is in the $K^+\pi^+_\mu$ distribution[2]. To remove the small possible contribution from $B^0 \to \overline{D}^0\pi^+\pi^+\pi^-$ events which fall within 30 MeV of the nominal D^0 mass in the $K^+\pi^-_\mu$ spectrum are vetoed.

Semileptonic cascades, where a b decays via $b \to c\mu^-\bar{\nu}_\mu$ and subsequently $c \to s\mu^+\nu_\mu$, can have branching fractions as high as $\mathcal{O}(10^{-4})$. For example, the decay $B^+ \to D^-\pi^+\mu^+\nu_\mu$ followed by $D^- \to K^+\pi^-\mu^-\bar{\nu}_\mu$ has a total branching fraction of $(1.6 \pm 0.3) \times 10^{-4}$ [14]. Selection requirements on the χ^2_{vtx} suppress these decays significantly, and the energy lost to the neutrinos means that this background sits considerably below the known B^+ mass, as shown in Fig. 6.6.

6.2.2 Multivariate selection

Combinatorial background is suppressed using a BDT trained using the AdaBoost algorithm [60], which is described in detail in Chap. 4.2, and is implemented using the Toolkit for Multivariate Analysis (TMVA) [88]. The signal-proxy is taken from the background subtracted sample of $B^+ \to J/\psi K^+\pi^+\pi^-$ candidates. Background-like events are selected from the upper mass sideband of the signal selection in the range $5530 < m_{K^+\pi^+\pi^-\mu^+\mu^-} < 5780$ MeV. These candidates are not used for the determination of the signal yield, and at this high mass are comprised solely of combinatorial background.

[2] Where, as defined previously, the notation h_i is a particle under the mass hypothesis of h which was reconstructed as an i.

Particle	Variables				
B^+	p_{T}	χ^2_{IP}	χ^2_{FD}	χ^2_{vtx}	θ_{dir}
Tracks	p_{T}	χ^2_{IP}			

Table 6.2: Input variables used to train the BDT to distinguish between signal $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ decays and combinatorial backgrounds.

This BDT was trained using selection of geometric and kinematic variables, the exact variables are listed in Table 6.2.

Optimisation of particle identification criteria and multivariate classifier

The determination of the optimum cut value on the BDT is required in conjunction with optimizing PID criteria on each hadron. For the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$, a requirement that $\mathrm{DLL}_{K\pi}(K^+) - \mathrm{DLL}_{K\pi}(\pi^+) > 10$ is made, to ensure that of the two hadrons with the same sign, the one identified as a kaon had more of a kaon like signature than the other. This also reduces the number of multiple candidates per event. Before this cut, there is, on average, 1.32 candidates per event within $3\,\sigma$ of $m^{\mathrm{PDG}}_{B^+}$, which is reduced to 1.05 candidates per event. Then, optimisation was made in the three dimensions of BDT, $\mathrm{DLL}_{K\pi}(K^+)$ and $\mathrm{DLL}_{K\pi}(\pi^\pm)$ with a grid search, by maximising the figure of merit $S/\sqrt{S+B}$, where S and B are the expected signal and background yields respectively. The value of S was determined by scaling the weighted sum of selected $B^+ \to J/\psi\, K^+\pi^+\pi^-$ events, $N\big(B^+ \to J/\psi\, K^+\pi^+\pi^-\big)$, according to

$$S\big(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-\big) = \frac{\mathcal{B}\big(B^+ \to K_1(1270)^+\mu^+\mu^-\big)\mathcal{B}\big(K_1(1270)^+ \to K^+\pi^+\pi^-\big)}{\mathcal{B}\big(B^+ \to J/\psi\, K^+\pi^+\pi^-\big)\mathcal{B}\big(J/\psi \to \mu^+\mu^-\big)} \cdot N\big(B^+ \to J/\psi\, K^+\pi^+\pi^-\big). \quad (6.9)$$

Branching fraction values used in this calculation are taken to be $\mathcal{B}\big(K_1(1270)^+ \to K^+\pi^+\pi^-\big) = (35.7 \pm 3.7) \times 10^{-2}$, $\mathcal{B}\big(J/\psi \to \mu^+\mu^-\big) = (5.93 \pm 0.06) \times 10^{-2}$, and $\mathcal{B}\big(B^+ \to J/\psi\, K^+\pi^+\pi^-\big) = (8.1 \pm 1.3) \times 10^{-4}$ [21]. Estimation of the branching fraction of the signal decay $B^+ \to K_1(1270)^+\mu^+\mu^-$ is made assuming the ratio of the branching fractions for the known decays $B \to X_s\mu^+\mu^-$ to $B \to X_s\gamma$ is the same for $X_s \in \{K_1(1270)^+, K^*(892)^0\}$. Thus,

$$\begin{aligned}\mathcal{B}\big(B^+ \to K_1(1270)^+\mu^+\mu^-\big) &= \mathcal{B}\big(B^+ \to K_1(1270)^+\gamma\big) \cdot \frac{\mathcal{B}\big(B^+ \to K^*(892)^0\mu^+\mu^-\big)}{\mathcal{B}\big(B^+ \to K^*(892)^0\gamma\big)} \\ &= \big(1.05 \pm 0.34\big) \times 10^{-6}. \end{aligned} \quad (6.10)$$

The value of the estimated background yield, B, is determined from interpolating $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ candidates from the mass sidebands into the regions around the mass of the B^+ meson. Candidates falling in the low and high mass sidebands are fit to a decaying exponential, and value of B taken to be the integral of this fitted distribution within $3\,\sigma$ of the known B^+ mass [21]. Sideband regions are defined by B^+ candidates with masses between 5000 MeV and 5750 MeV, but more than 120 MeV from the nominal B^+ mass.

The optimisation procedure for the decay $B^+ \to \phi K^+\mu^+\mu^-$ employs a similar strategy as $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$, but only in two dimensions (BDT and $\mathrm{DLL}_{K\pi}(K^\pm)$). The calculation of B was made in the same way as described above, and the value of S was determined in a similar way, by scaling the *sWeighted* sum of candidates that passed given cuts. So,

$$S\big(B^+ \to \phi K^+\mu^+\mu^-\big) = \frac{\mathcal{B}\big(B^+ \to \phi K^+\mu^+\mu^-\big)}{\mathcal{B}\big(B^+ \to J\!/\!\psi\,\phi K^+\big)\mathcal{B}\big(J\!/\!\psi \to \mu^+\mu^-\big)} \cdot N\big(B^+ \to J\!/\!\psi\,\phi K^+\big). \quad (6.11)$$

where $N\big(B^+ \to J\!/\!\psi\,\phi K^+\big)$ is the weighted number of selected $B^+ \to J\!/\!\psi\,\phi K^+$ events. The expected background yield is determined using an exponential fit across the signal region, as is done for the $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ channel. The prediction for the branching fraction of the signal channel is taken to be

$$\begin{aligned}\mathcal{B}\big(B^+ \to \phi K^+\mu^+\mu^-\big) &= \mathcal{B}\big(B^+ \to \phi K^+\gamma\big) \cdot \frac{\mathcal{B}\big(B^+ \to K^*(892)^0\mu^+\mu^-\big)}{\mathcal{B}\big(B^+ \to K^*(892)^0\gamma\big)} \\ &= \big(0.66 \pm 0.12\big) \times 10^{-7}. \end{aligned} \quad (6.12)$$

The maximum value of the figure of merit was found, and corresponding cut values used for the analysis. Exact requirements for the PID variables were that $\mathrm{DLL}_{K\pi}(K^+) > 3.5$, $\mathrm{DLL}_{K\pi}(\pi^\pm) < 14.5$ and BDT > 0.025. After the full selection multiple candidates are removed at random such that there is, at most, one $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ candidate per event. For the analysis of the decay $B^+ \to \phi K^+\mu^+\mu^-$, the optimisation procedure yields cut values of $\mathrm{DLL}_{K\pi}(K) > -3$ and BDT > 0.05. The value of the $\mathrm{DLL}_{K\pi}$ criteria is looser for $B^+ \to \phi K^+\mu^+\mu^-$ than for $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ because the implicit PID requirements when selecting a ϕ candidate.

6.3 Efficiency calculations

Although the normalisation channels have the same final state particles, the efficiency does not completely cancel due to kinematic differences. Therefore the relative efficiency between signal and normalisation decays must be calculated. The efficiency for each decay

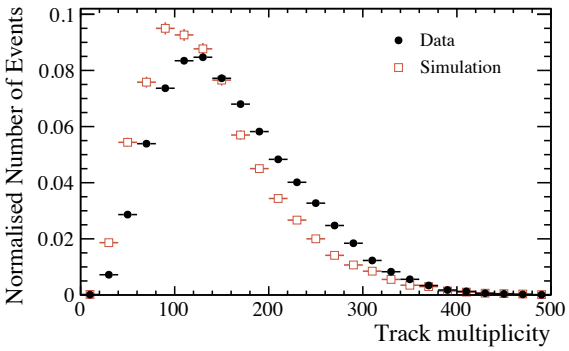


Figure 6.7: Distributions of track multiplicity for (black circles) data and (red squares) simulation. Simulated events are known to mis-model the track multiplicity, having a lower average number of tracks per event, this is caused by the modelling of hadronization in PYTHIA.

is computed using simulated events, under the assumption that the simulation accurately describes data, where there are discrepancies the simulations must be corrected. There are some variables which are known to be poorly described in simulation, particularly track multiplicity and the χ^2_{vtx} of the B^+ candidate. The effects of these discrepancies are minimised by reweighting simulated events using the distributions of $B^+ \to J/\psi\, K^+\pi^+\pi^-$. Track multiplicity is known to be poorly described in simulation, generally there are fewer tracks in simulation than observed in data, these discrepancies are caused by the modelling of hadronization in PYTHIA and are illustrated in Fig. 6.7. Aside from the direct differences, the low track multiplicity of in simulated events is a contributing factor of PID variables being badly described in simulations.

To ensure that efficiencies from PID cuts are determined accurately from simulation, the variables must be corrected. This is done using data driven methods, using highly pure samples of pions, kaons and muons (coming from the decays $D^{*+} \to D^0(\to K^+\pi^-)\pi^+$ and $J/\psi \to \mu^+\mu^-$). For each hadron track in the simulated B^+ candidate, a new PID variable is resampled from PID distributions of the pure track samples as a function of the track's pseudorapidity, momentum, and track multiplicity. Figure 6.8 shows the effect of this resampling technique, it is observed that the simulated PID distributions that have been resampled matches data distributions much better than the raw distributions; the differences that remain are accounted for in the systematic uncertainties. Muon PID distributions are sufficiently well modelled in simulation that correcting them is unnecessary.

Tracking efficiency varies depending upon the regions of the detector through which the particle passes, and the modelling of the detector in the simulation behaves differently to in actuality. To correct for this, each candidate is weighted based on the relative tracking efficiency between data and simulation, this is dependent upon p and η. The same is true for the response of the `isMuon` variable. Figure 6.9 shows how the tracking efficiency and `isMuon` criteria are corrected for throughout the detector volume. After all reweighting all

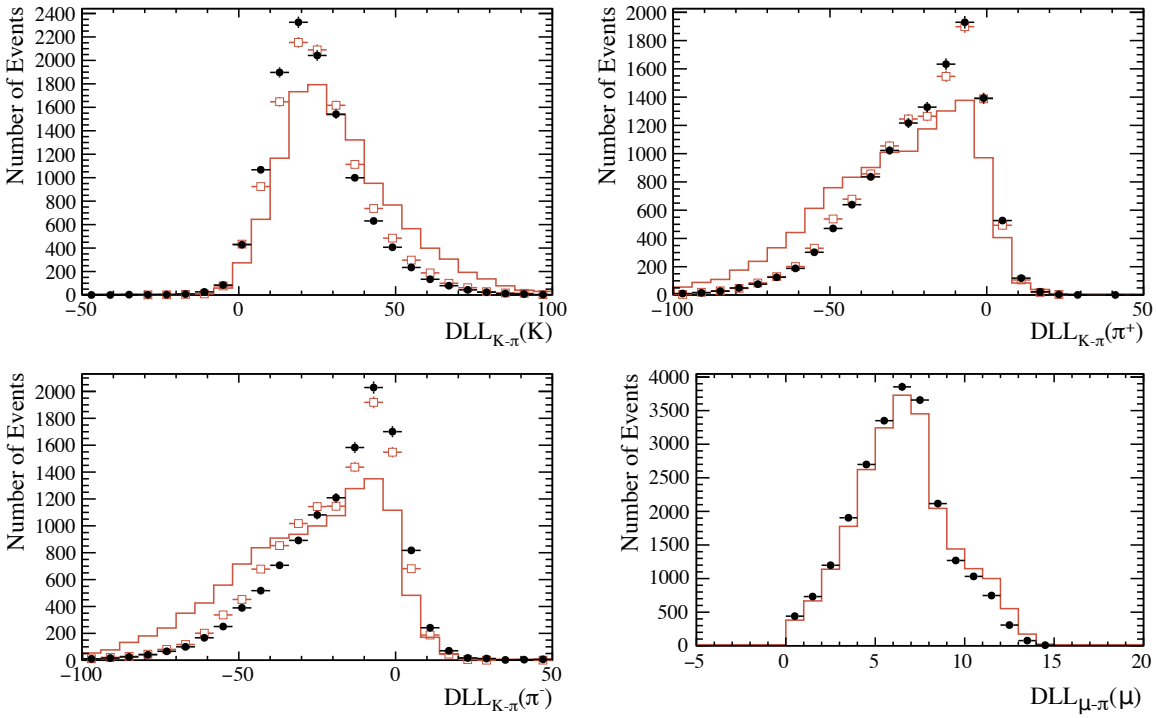


Figure 6.8: The effect of PID resampling for simulated tracks using pure samples of particles for kaons and pions. There is marked improvement in the similarity of the (black circles) $B^+ \to J/\psi K^+\pi^+\pi^-$ data and simulated events before (red line) resampling and after (red squares). Muon candidates need not be resampled since muon PID is well described in simulation.

the aforementioned variables in simulated events, the BDT distributions are seen to be in agreement, this is shown in Fig. 6.10.

Once the simulation has been corrected, the total efficiency, ε_{tot} was calculated for each normalisation and signal mode using simulated events. The value for ε_{tot} is calculated to be $\varepsilon_{\text{gen}} \times \varepsilon_{\text{reco\&sel}} \times \varepsilon_{\text{trig}}$, where: ε_{gen} is the generator selection efficiency; $\varepsilon_{\text{reco\&sel}}$ is the reconstruction and selection efficiency; $\varepsilon_{\text{trig}}$ trigger efficiency. The generator efficiency defines the probability that a B^+ decays into daughter particles which all pass through the LHCb detector acceptance, and is approximately 15 % for each signal and normalisation channel.

Since efficiency calculations require reliable simulated samples of decays, accurate physics models must, or should, be used. This raises a dilemma deciding how to calculate the efficiency of the signal decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$, because no physics models exist for them. For the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$, an appropriate choice was the physics model for $B^+ \to K_1(1270)^+\mu^+\mu^-$ from Ref. [86] and $\theta_{K_1} = 34°$, because this was assumed to be a dominant contribution. As there is no available physics model for the decay $B^+ \to \phi K^+\mu^+\mu^-$, simulated events are produced using a phasespace model. These models introduce systematic uncertainties, which are discussed in Sec. 6.4.1 and Sec. 6.5.1. The efficiencies for signal decays in each q^2 bin are shown in Fig. 6.11. Reliable

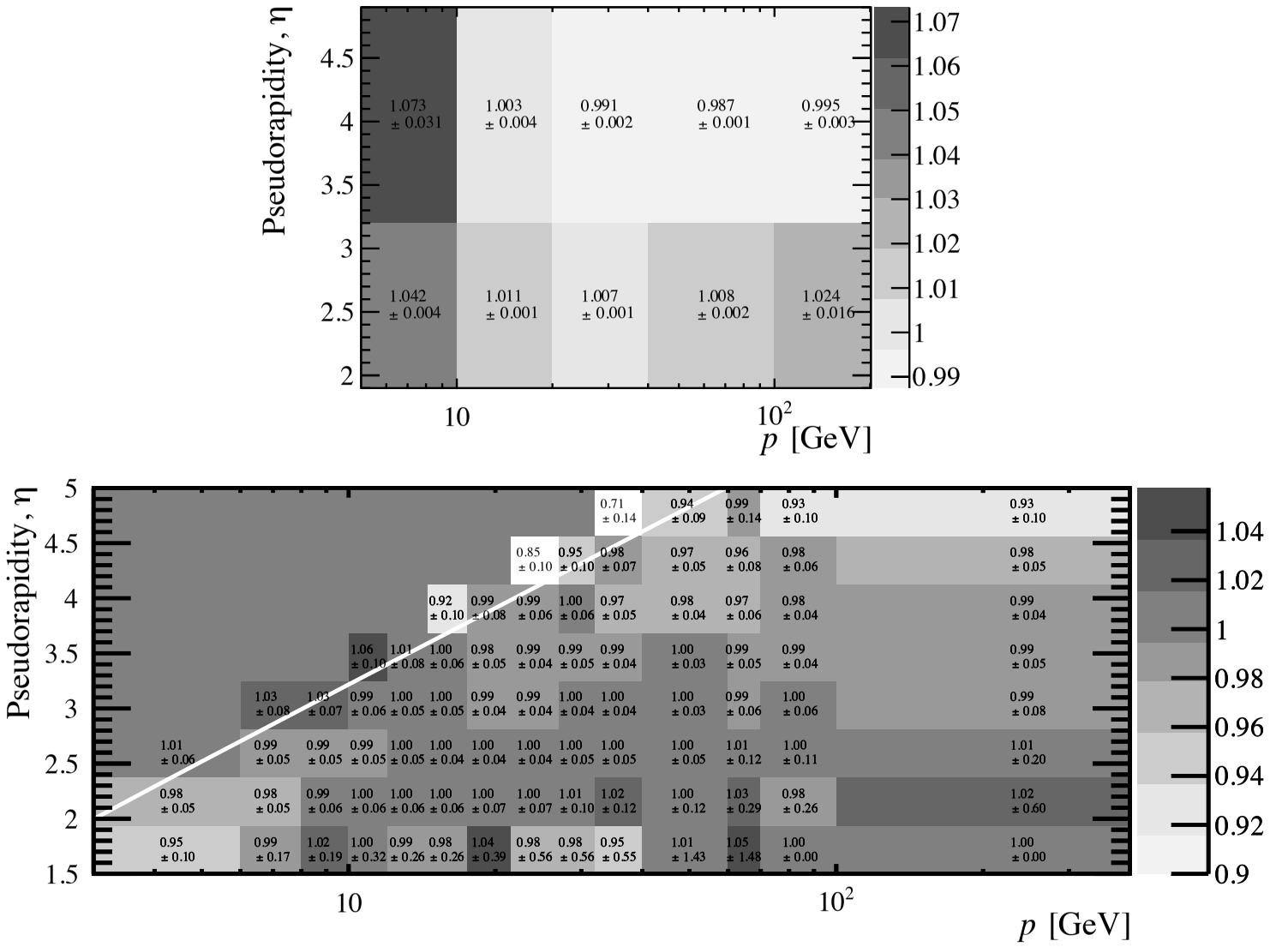


Figure 6.9: The ratio of efficiencies for (top) tracking, and (bottom) `isMuon` between data and simulation. Simulated tracks in simulation are reweighted according to their momentum and psudorapidity. For the lower plot, the calibration sample does not contain muons with $p_{\mathrm{T}} < 800\,\mathrm{MeV}$, this geometric threshold is indicated by the white line.

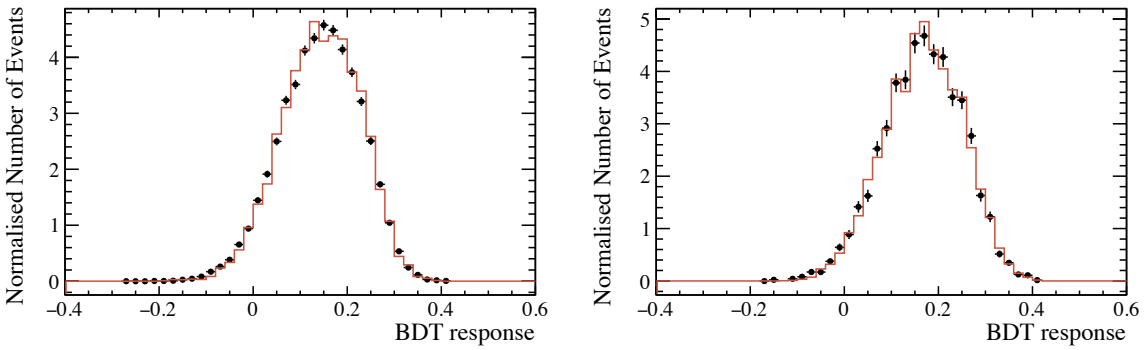


Figure 6.10: Distributions of the BDT for data (black points) and simulation (red line) for the decay (left) $B^+ \to J/\psi K_1(1270)^+$ (right) $B^+ \to \psi(2S)K^+$.

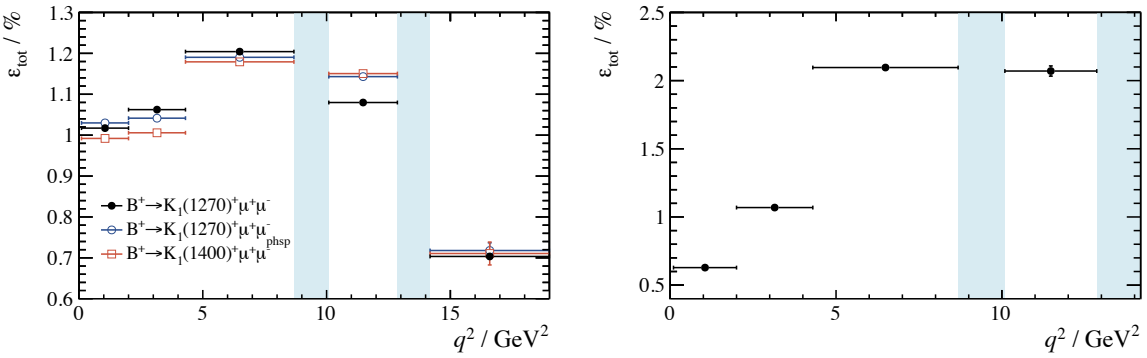


Figure 6.11: The efficiency of the signal decay (left) $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$, showing a range of decay models; and (right) $B^+ \to \phi K^+\mu^+\mu^-$ in bins of q^2. For the $B^+ \to \phi K^+\mu^+\mu^-$ analysis, there are no events in the $14.18 < q^2 < 19.00\,\text{GeV}^2$ bin which is present in the $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ analysis. Light blue shaded regions indicate the vetoed J/ψ and $\psi(2S)$ regions.

physics models exist for the normalisation channels, which have efficiencies measured to be $\varepsilon(B^+ \to \psi(2S)K^+) = (0.41 \pm 0.01)\,\%$ and $\varepsilon(B^+ \to J/\psi\,\phi K^+) = (2.11 \pm 0.01)\%$ the inefficiency of the former is due to the soft pions from the $\psi(2S)$ decay failing p_{T} requirements in the stripping.

6.4 Differential branching fraction of the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$

Given the statistics available for this channel, the differential branching fraction, $\text{d}\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)/\text{d}q^2$ was calculated in five bins of q^2 with respect to the normalisation channel $B^+ \to \psi(2S)K^+$. This normalisation channel is chosen because it has the same final state particle if $\psi(2S) \to J/\psi\,\pi^+\pi^-$ and $J/\psi \to \mu^+\mu^-$. This has a total branching fraction of $\mathcal{B}_{\text{tot}}\left(B^+ \to \psi(2S)K^+\right) = (1.264 \pm 0.0052) \times 10^{-5}$ [21], which has a relative uncertainty of 4 %, the alternative normalisation channel is $B^+ \to J/\psi\,K^+\pi^+\pi^-$ but this

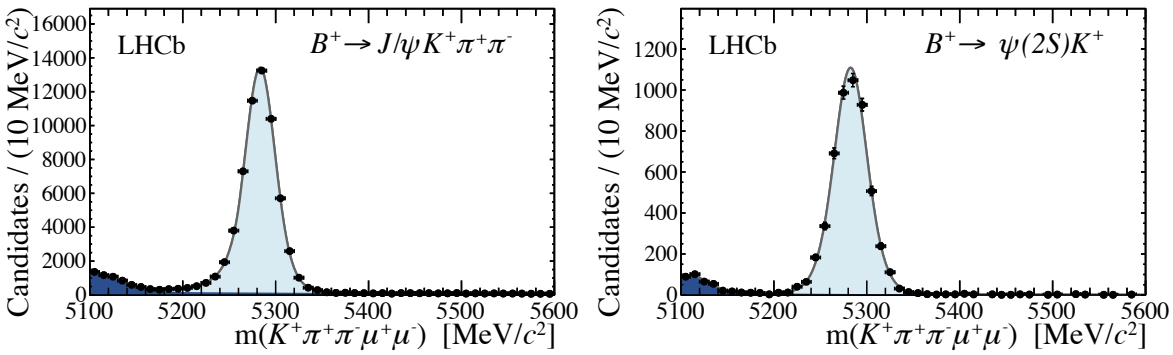


Figure 6.12: Distributions of the invariant mass of the (left) cross-check channel $B^+ \to J/\psi K^+\pi^+\pi^-$ and (right) normalisation mode $B^+ \to \psi(2S)K^+$. Projections from the fit are overlaid, where the light blue is the yield of the indicated decay, and the dark blue is the background component.

has a relative uncertainty of 16 %.

The differential branching fraction for a bin of width Δq^2 is calculated

$$\frac{d}{\mathrm{d}q^2}\mathcal{B}\big(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-\big) = \frac{1}{\Delta q^2} \cdot \frac{N\big(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-\big)}{N\big(B^+ \to \psi(2S)K^+\big)} \cdot \frac{\varepsilon\big(B^+ \to \psi(2S)K^+\big)}{\varepsilon\big(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-\big)} \cdot \mathcal{B}_{\mathrm{tot}}\big(B^+ \to \psi(2S)K^+\big) \tag{6.13}$$

where N and ε denote yields and efficiencies, respectively.

The yield of the normalisation channel is extracted from an unbinned extended maximum likelihood fit to the invariant mass distribution of $B^+ \to \psi(2S)K^+$ candidates. The signal shape is modelled using the sum of two Gaussian functions, each having a power-law tail at low-mass, where the tail parameters are shared between both Gaussians. Two Gaussian functions are due to different resolution effects, the wider one is due to events where tracks have been through the OT or undergone multiple scattering. Background is modelled using the sum of an exponential to model combinatorial background, and a Gaussian at low mass to model partially reconstructed candidates. The control channel $B^+ \to J/\psi K^+\pi^+\pi^-$ is fit to the same model. These fits are shown in Fig. 6.12 and they yield $N\big(B^+ \to \psi(2S)K^+\big) = 5128 \pm 67$ and $N\big(B^+ \to J/\psi K^+\pi^+\pi^-\big) = 59\,335 \pm 343$.

Yields of the signal decay in regions of q^2 were extracted from fits using a distribution that was fixed to as large an extent as possible. The lower limit of the fit was taken to be 5150 MeV, because partially reconstructed background becomes significant below that. The combinatorial background shape is an exponential distribution, where the exponent and yield was allowed to float in the fit. In q^2 regions where charmonium vetoes are effective, the background distribution must account for the areas in which charmonium

vetoes are extended, this is done by incorporating discrete steps in the combinatorial background; the size of which are determined from data. A fit from the crosscheck channel fixes the signal model used in the lower statistics signal channel. Figure 6.13 shows the invariant mass distribution for all signal candidates for the whole q^2 region and for each bin. The total signal yield is measured to be $N(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) = 367\,^{+24}_{-23}$.

Differential branching fractions were calculated according to Eq. 6.13, where yields were extracted from fits in Fig. 6.13. Results are shown graphically in Fig. 6.14 and are given numerically — along with signal yields — in Table 6.3. These results also quote a number for the $1.0 < q^2 < 6.0\,\mathrm{GeV}^2$ region, which is an area of lower theoretical uncertainty for dimuon FCNC decays, as it is away from the photon pole at low mass, and away from the J/ψ at high mass. Uncertainty from the normalisation channel is fully correlated between all q^2 bins. The sum over all q^2 bins results in an integrated branching fraction of

$$\mathcal{B}\big(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-\big) = \big(3.43\,^{+0.23}_{-0.21}(\mathrm{stat}) \pm 0.15(\mathrm{syst}) \pm 0.14(\mathrm{norm})\big) \times 10^{-7};$$

where quoted uncertainties are statistical, systematic and due to the errors on the branching fraction of the normalisation channel. The fraction of signal events removed by the charmonium vetoes is determined from simulated $B^+ \to K_1(1270)^+\mu^+\mu^-$ events to be $(21.3 \pm 1.5)\,\%$. Finally, the adjusted integrated branching fraction is

$$\mathcal{B}\big(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-\big) = \big(4.36\,^{+0.29}_{-0.27}(\mathrm{stat}) \pm 0.21(\mathrm{syst}) \pm 0.18(\mathrm{norm})\big) \times 10^{-7}.$$

The statistical significance of this result is in excess of $20\,\sigma$ according to Wilks' theorem [72].

Since the uncertainty due to the normalisation channel branching fraction is significant, the ratio of branching fractions is also quoted

$$\frac{\mathcal{B}\big(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-\big)}{\mathcal{B}\big(B^+ \to \psi(2S)K^+\big)} = \big(6.95\,^{+0.46}_{-0.43}(\mathrm{stat}) \pm 0.34(\mathrm{syst})\big) \times 10^{-4}.$$

Figure 6.15 shows the background subtracted invariant mass distribution's of the $K^+\pi^+\pi^-$ system in the case of both decays $B^+ \to J/\psi\, K^+\pi^+\pi^-$ and $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$. Numerous broad resonances are visible in the distributions. The $K^+\pi^+\pi^-$ system from the resonant $B^+ \to J/\psi\, K^+\pi^+\pi^-$ decay shows significant, though not dominant, contributions from the $K_1(1270)^+$ state. There is also a contribution visible as $m_{K^+\pi^+\pi^-} \simeq 1750\,\mathrm{MeV}$, similar to resonances seen by Belle [79]. However, there are significant differences between the invariant mass distributions of the $K^+\pi^+\pi^-$ system in the decays $B^+ \to J/\psi\, K^+\pi^+\pi^-$ and $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$. These differences can be explained by differences in the allowed spin states due to the pseudoscalar J/ψ.

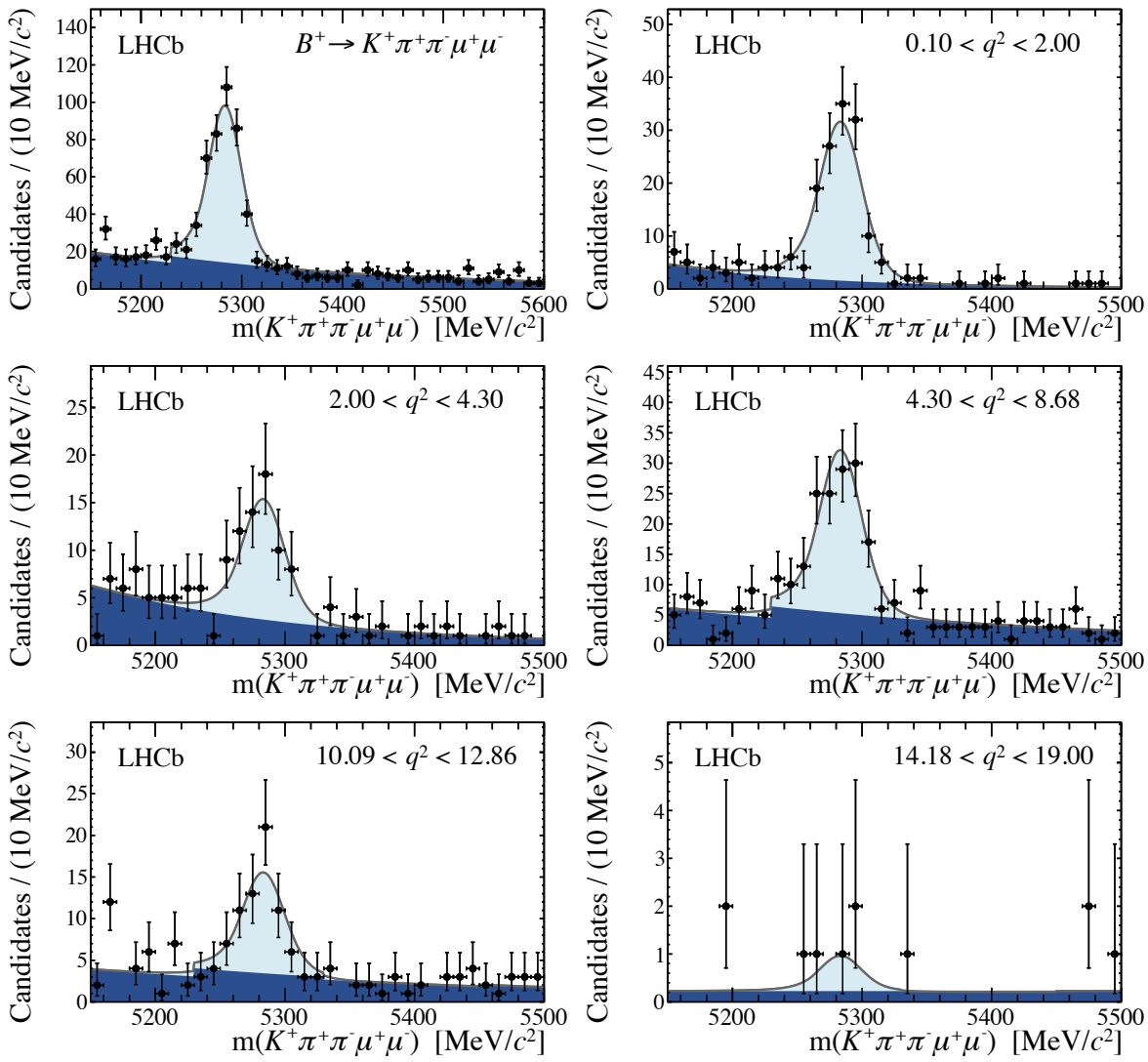


Figure 6.13: Invariant mass distributions of $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ candidates in bins of q^2 with projections of fits overlaid, the upper left plot is a separate fit to the full q^2 range. The signal signal component (light blue) is modelled by the sum of two Gaussian distributions, and the background component (dark blue) is an exponential function. In the three q^2 bins $4.30 < q^2 < 8.68\,\text{GeV}^2$, $10.09 < q^2 < 12.86\,\text{GeV}^2$, and $14.18 < q^2 < 19.00\,\text{GeV}^2$ scaling factors in the background components are used to account the removal of the radiative tails of charmonium vetoes.

q^2 bin (GeV^2)	$N_{\rm sig}$	$\frac{\mathrm{d}\mathcal{B}}{\mathrm{d}q^2}$ $\left(\times 10^{-8}\,\mathrm{GeV}^{-2}\right)$
[0.10, 2.00]	$134.1\,^{+12.9}_{-12.3}$	$7.01\,^{+0.69}_{-0.65} \pm 0.47$
[2.00, 4.30]	$56.5\,^{+\ 9.7}_{-\ 9.1}$	$2.34\,^{+0.41}_{-0.38} \pm 0.15$
[4.30, 8.68]	$119.9\,^{+14.6}_{-13.7}$	$2.30\,^{+0.28}_{-0.26} \pm 0.20$
[10.09, 12.86]	$54.0\,^{+10.1}_{-\ 9.4}$	$1.83\,^{+0.34}_{-0.32} \pm 0.17$
[14.18, 19.00]	$3.3\,^{+\ 2.8}_{-\ 2.1}$	$0.10\,^{+0.08}_{-0.06} \pm 0.01$
[1.00, 6.00]	$144.8\,^{+14.9}_{-14.3}$	$2.75\,^{+0.29}_{-0.28} \pm 0.16$

Table 6.3: Signal yields and resulting differential branching fractions for the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ in bins of q^2.

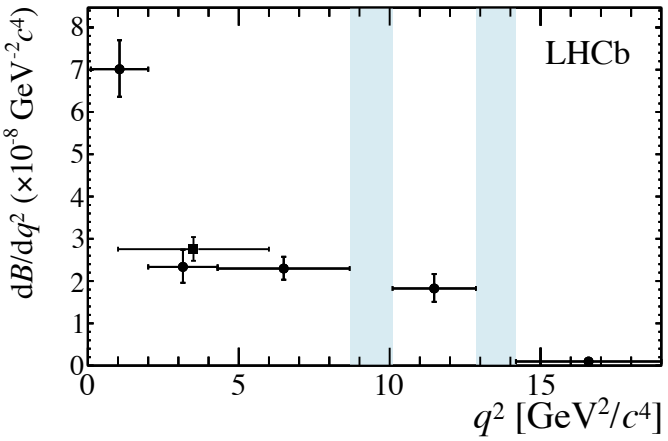


Figure 6.14: Differential branching fraction $\frac{d}{\mathrm{d}q^2}\mathcal{B}\left(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-\right)$ as given in Table 6.3, where the errors shown include both statistical and systematic uncertainties. Shaded areas indicate the vetoed J/ψ and $\psi(2S)$ resonances. The result for the $1.0 < q^2 < 6.0\,\mathrm{GeV}^2$ region which has reduced theoretical uncertainties is also shown.

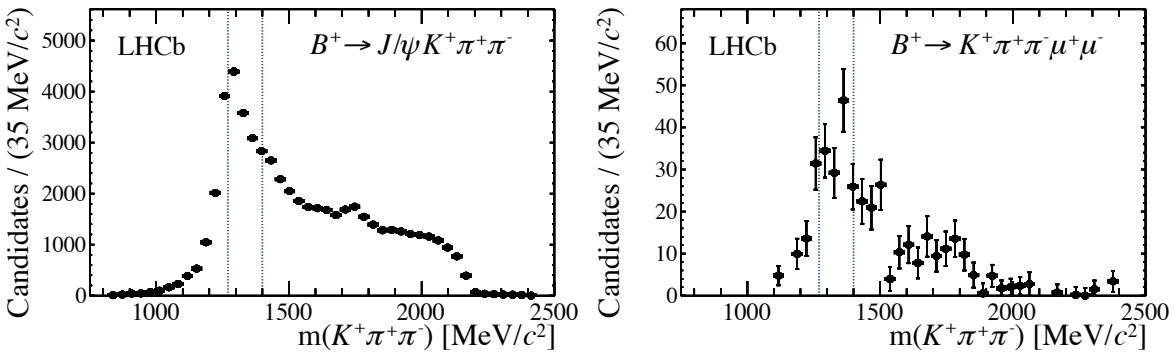


Figure 6.15: Invariant mass distributions of the $K^+\pi^+\pi^-$ system, from the (left) control channel $B^+ \to J/\psi\, K^+\pi^+\pi^-$, and the (right) signal decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ which have been background subtracted using the *sPlot* [69] technique. The vertical lines indicate the central masses of the $K_1(1270)^+$ and $K_1(1400)^+$ resonances.

6.4.1 Systematic uncertainties

Systematic uncertainties introduced to the measurement of the differential branching fraction of $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ can be categorised into three main categories: corrections applied to simulated events for the efficiency calculation; models used to describe the signal and normalisation channel shapes; and modelling of the signal decay.

Section 6.3 describes the process of correcting simulated events such that they best describe data. This procedure introduces systematic uncertainties which must be accounted for. This is done by recalculating the relative efficiency for each q^2 bin: without reweighting track multiplicity and the χ^2 of the fit to the B^+ vertex; without accounting for differences in tracking and `isMuon` efficiencies; and without performing PID resampling. The resulting systematic uncertainty from the selection was assigned a value of 1 %. Each source of uncertainty is treated separately, and the total systematic uncertainty is taken to be the difference in the branching fraction with the newly calculated relative efficiency.

In order to assess the degree of uncertainty introduced by the distributions used to model the signal and background components. The systematic uncertainty is taken to be the difference between the nominal branching fraction, and the branching fraction recalculated using different signal or background components in the fit. A single Gaussian with a power-law tail is used as an alternative signal model and — in a different fit — a linear function is used to describe the background. The modelling of the mass distributions led to a systematic uncertainty of 2 %.

The way the signal decay is modelled in simulation were the source of numerous systematic uncertainties. Efficiencies were calculated using simulated data using the model described in Ref. [86], but this is imperfect for a number of reasons. This model does not match data in q^2 or $m_{K^+\pi^+\pi^-}$; it also assumes that $\theta_{K_1} = -34^\circ$, and was used to determine the fraction of signal removed by charmonium vetoes.

To assess the systematic introduced by the difference between the q^2 distribution in data and simulation a different decay model was used. Rather than the efficiency for the signal channel being determined from simulated $B^+ \to K_1(1270)^+\mu^+\mu^-$ generated with the model in Ref. [86], it was decayed using a phasespace model. The differences in the invariant mass distribution of $K^+\pi^+\pi^-$ cannot be assessed using simulation because of the difficulties associated with theoretically describing $K^+\pi^+\pi^-$ resonances. Instead, the signal efficiencies are calculated after being reweighted to match the sWeighted $m(K^+\pi^+\pi^-)$ distribution from signal, as shown in Fig. 6.15.

The value of the mixing angle θ_{K_1} has been a source of some theoretical contention, where $\theta_{K_1} = -34^\circ$ or -57° [80–85]. But, more recently measurements favour a value of approximately $-(34 \pm 13)^\circ$ [82–85]. To estimate the systematic uncertainty associated with this value, the simulated data was reweighted on an event-by-event basis such that

Source	q^2 bin [GeV^2] [0.10, 2.00]	[2.00, 4.30]	[4.30, 8.68]	[10.09, 12.86]	[14.19, 19.00]
$\mathcal{B}(B^+ \to \psi(2S)K^+)$	0.288	0.096	0.094	0.075	0.004
Corrections to simulation	0.218	0.082	0.074	0.048	0.002
$K^+\pi^+\pi^-$ composition	0.175	0.009	0.010	0.061	0.004
Value of θ_{K_1}	0.110	0.036	0.062	0.003	0.001
Modelling q^2	0.087	0.047	0.027	0.094	0.002
Mass model: background	0.154	0.038	0.052	0.093	0.003
Mass model: signal	0.138	0.052	0.052	0.039	0.002
Relative efficiencies	0.014	0.005	0.003	0.003	0.000
Peaking backgrounds	0.000	0.000	0.126	0.000	0.000
Total	0.473	0.154	0.201	0.175	0.007

Table 6.4: Systematic uncertainties on the differential branching fraction of $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ in bins of q^2 [$\times 10^{-8}\,\text{GeV}^{-2}$].

the value of θ_{K_1} was varied by $1\,\sigma$.

The signal efficiency was recalculated after varying the θ_{K_1} by $1\,\sigma$ in simulated $B^+ \to K_1(1270)^+\mu^+\mu^-$ events. Branching fractions were then evaluated using these new efficiencies, and the maximum difference from the nominally calculated value is taken as the systematic uncertainty. The same procedure was used to calculate systematics with respect to $\theta_{K_1} = -57^\circ$, and the values are found to be consistent with the theoretically favoured value of $-(34 \pm 13)^\circ$; leading to an assigned uncertainty of $\sim$1.5 %.

Charmonium vetoes removes a percentage of signal candidates, that must be accounted for in order to calculate the total integrated branching fraction. This fraction is subject to a systematic uncertainty due to the mismodelling of the q^2 distribution in the simulated signal events. Form factors in Ref. [86] gave the central value of 21.3 %. Varying these form factors at generator level by a random number forms a Gaussian distribution, whose standard deviation is the uncertainty on the form factor. This is done multiple times and leads to an uncertainty of 1.5 %.

6.5 Branching fraction of the decay $B^+ \to \phi K^+\mu^+\mu^-$

The branching fraction of the signal decay $B^+ \to \phi K^+\mu^+\mu^-$ is determined relative to the normalisation channel $B^+ \to J/\psi\,\phi K^+$:

$$\mathcal{B}\big(B^+ \to \phi K^+\mu^+\mu^-\big) = \frac{N'\big(B^+ \to \phi K^+\mu^+\mu^-\big)}{N\big(B^+ \to J/\psi\,\phi K^+\big)} \cdot \mathcal{B}\big(B^+ \to J/\psi\,\phi K^+\big) \cdot \mathcal{B}\big(J/\psi \to \mu^+\mu^-\big), \tag{6.14}$$

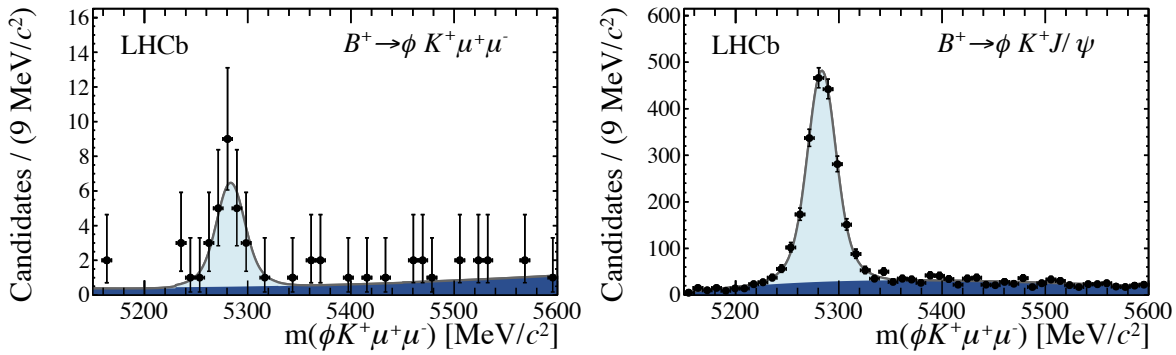


Figure 6.16: Invariant mass distributions for the (left) signal decay $B^+ \to \phi K^+\mu^+\mu^-$, and (right) normalisation mode $B^+ \to J/\psi\,\phi K^+$. The signal component (light blue) is modelled by the sum of two Gaussian distributions, each with a power-law tail on the low-mass side; and the background component is a second order Chebychev polynomial. Variables describing the signal shape in the fit to $B^+ \to \phi K^+\mu^+\mu^-$ are fixed by the fit to $B^+ \to J/\psi\,\phi K^+$.

here, $N'\big(B^+ \to \phi K^+\mu^+\mu^-\big)$ denotes the number of signal events extracted from an weighted unbinned maximum likelihood fit of $B^+ \to \phi K^+\mu^+\mu^-$ candidates. Each event is weighted by the relative efficiency, so that

$$N'\big(B^+ \to \phi K^+\mu^+\mu^-\big) = \sum_i \frac{\varepsilon\big(B^+ \to J/\psi\,\phi K^+\big)}{\varepsilon^{(i)}_{q^2}\big(B^+ \to \phi K^+\mu^+\mu^-\big)}, \tag{6.15}$$

where the denominator is the efficiency of the signal decay for candidate i, in bins of q^2. The variation of efficiency in q^2 for the signal decay is shown in Fig. 6.11. This is done because the efficiency of the decay $B^+ \to \phi K^+\mu^+\mu^-$ was shown to vary significantly over the full q^2 range, the weights are determined in bins of q^2. The branching fraction measurements used were $\mathcal{B}\big(B^+ \to J/\psi\,\phi K^+\big) = \big(5.2 \pm 1.7\big) \times 10^{-5}$ [21], and $\mathcal{B}\big(J/\psi \to \mu^+\mu^-\big) = \big(5.93 \pm 0.06\big) \times 10^{-2}$ [21].

Yields for both the signal and normalisation channels are extracted from unbinned maximum likelihood fits of the invariant mass of the B^+ candidates, these fits are shown in Fig. 6.16. The signal component for the normalisation channel is the sum of two Gaussian functions, with a power-law tail on the low mass side; the same function is used for fitting the weighted signal distribution where all parameters are fixed from a fit to the high statistics normalisation mode. Combinatorial background is modelled as a second order Chebychev polynomial. These fits give the values $N'\big(B^+ \to \phi K^+\mu^+\mu^-\big) = 25.2\,^{+6.0}_{-5.3}$ and $N\big(B^+ \to J/\psi\,\phi K^+\big) = 1908 \pm 63$. The statistical significance of this is $6.6\,\sigma$ according to Wilks' theorem [72].

The above values lead to a measured branching fraction of

$$\mathcal{B}\big(B^+ \to \phi K^+\mu^+\mu^-\big) = \big(0.81\,^{+0.18}_{-0.16}(\text{stat}) \pm 0.03(\text{syst}) \pm 0.27(\text{norm})\big) \times 10^{-7}. \tag{6.16}$$

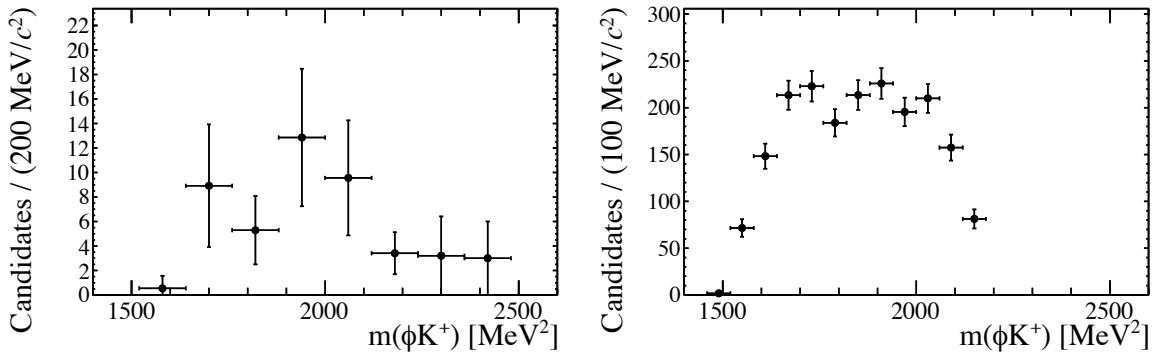


Figure 6.17: Distributions of the sWeighted invariant mass ϕK^+ object from the decay (left) $B^+ \to \phi K^+\mu^+\mu^-$, and (right) $B^+ \to J/\psi\,\phi K^+$.

However, the charmonium vetoes remove $\left(2\,^{+10}_{-\ 2}\right)\%$ of signal events, as calculated using simulated events. This systematic uncertainty in $B^+ \to \phi K^+\mu^+\mu^-$ is larger than in $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ because the q^2 distribution in the case of the former drops off and ends in the J/ψ region, and reweighting the q^2 distribution causes significant differences in the amount vetoed. Taking this into account results in a value of

$$\mathcal{B}\big(B^+ \to \phi K^+\mu^+\mu^-\big) = \big(0.82\,^{+0.19}_{-0.17}(\text{stat})\,^{+0.10}_{-0.04}(\text{syst}) \pm 0.27(\text{norm})\big) \times 10^{-7}. \tag{6.17}$$

The ratio of branching fractions of the signal and normalisation channels is

$$\frac{\mathcal{B}\big(B^+ \to \phi K^+\mu^+\mu^-\big)}{\mathcal{B}\big(B^+ \to J/\psi\,\phi K^+\big)} = \big(1.58\,^{+0.36}_{-0.32}(\text{stat})\,^{+0.19}_{-0.07}(\text{syst})\big) \times 10^{-3}, \tag{6.18}$$

which is quoted because there are large relative uncertainties associated with the branching fraction of the normalisation channel ($\sim$1 %).

Just as in $K^+\pi^+\pi^-$, it is expected that the ϕK^+ system will be composed of numerous strange resonances. However, low statistics and some phasespace limitations make it impossible to draw and meaningful conclusions. The background subtracted ϕK^+ distributions for the signal and normalisation channels are shown in Fig. 6.17. As in the case of the invariant mass distributions of the $K^+\pi^+\pi^-$ system in the decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to J/\psi\,K^+\pi^+\pi^-$, there are differences between the ϕK^+ mass in $B^+ \to \phi K^+\mu^+\mu^-$ and $B^+ \to J/\psi\,\phi K^+$; and is explained by the spin state of the J/ψ.

6.5.1 Systematic uncertainties

Systematic uncertainties contributing to the measurement of $\mathcal{B}(B^+ \to \phi K^+\mu^+\mu^-)$ are determined in very similar ways as for the measurement of $\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)$, as detailed in Sec. 6.4.1. The uncertainties introduced by correcting the simulated events by weighting and resampling are calculated in the same way as for the decay

Source	$[\times 10^{-8}]$
$\mathcal{B}(B^+ \to J/\psi\,\phi K^+)$	2.688
Corrections to simulation	0.145
Relative efficiency	0.065
Mass model: background	0.245
Mass model: signal	0.128
q^2 model	0.118

Table 6.5: Systematic uncertainties on the branching fraction of $B^+ \to \phi K^+\mu^+\mu^-$.

$B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$. The systematic uncertainty arising from the signal mass model was evaluated by substituting the nominal double Gaussian with a power-law tail, for a single Gaussian with a tail and recalculating the signal yield. Similarly, the systematic from the background model is evaluated by repeating the fit with a first order, rather than a second order, polynomial. These changes in the fit model lead to to a total systematic uncertainty of $\sim$3 %. The uncertainty on the q^2 distribution in the simulation was estimated by reweighting the from the nominal phasespace simulation to the distribution of the $B^+ \to K_1(1270)^+\mu^+\mu^-$ mode. Therefore, a photon pole at low q^2 is introduced. The total systematic uncertainty evaluates to $\sim$1.5 %.

The fraction of events that are removed by the charmonium vetoes are determined from simulation. But, as discussed, the q^2 distribution that is described by the $B^+ \to \phi K^+\mu^+\mu^-$ simulation was inaccurate since there is no theoretical prediction available. To circumnavigate this problem, generator level simulated events is used to obtain the central value. Vetoed fractions are then calculated by generating events whose q^2 distribution is taken from the $B^+ \to K_1(1270)^+\mu^+\mu^-$ decay, where the mass of the $K_1(1270)^+$ system is replaced with an estimated mass of the ϕK^+ system. The value of $m_{\phi K^+} = 1960\,\text{MeV}$, which is taken fro the sWeighted $m_{K^+\pi^+\pi^-}$ distribution from data. In this way, it was determined that $(2\,^{+10}_{-\ 2})$ % of signal events were removed by the charmonium vetoes.

All other systematic uncertainties are evaluated in the same way as in the analysis of the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$, and are summarised in Table 6.5.

6.6 Summary

The two FCNC decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$ were both observed for the first time. Their branching fractions were measured to be:

$$\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) = \left(4.36\,^{+0.29}_{-0.27}(\text{stat}) \pm 0.21(\text{syst}) \pm 0.18(\text{norm})\right) \times 10^{-7},$$
$$\mathcal{B}(B^+ \to \phi K^+\mu^+\mu^-) = \left(0.82\,^{+0.19}_{-0.17}(\text{stat})\,^{+0.10}_{-0.04}(\text{syst}) \pm 0.27(\text{norm})\right) \times 10^{-7}.$$

The ratio of these branching fractions is $\sim$5.3, which is consistent with the ratio of branching fractions for the decays $B^+ \to K^+\pi^+\pi^-$ and $B^+ \to \phi K^+$. Additionally, the differential decay rate of the decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ was calculated with respect to q^2; these results are given in Table 6.3 and shown in Fig. 6.14.

Also shown were the resonances of the $K^+\pi^+\pi^-$ and ϕK^+ systems for the signal decays, and for the same final states where the muons come from a J/ψ.

Chapter 7

Search for the decay of a dark sector particle $\chi \to \mu^+\mu^-$ in $B^0 \to K^{*0}\mu^+\mu^-$

7.1 Introduction

This chapter describes a procedure for searching for a dark boson, χ, of unknown mass and lifetime[1] as published in Ref. [3]. However, some of the results in this thesis diverge from the LHCb analysis and are presented separately. A frequentist method is applied to the dimuon distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates, to search for an excess of events above the SM background, consistent with a χ decaying into a pair of muons. Lifetime information is added by splitting candidates into two bins of decay time: those which are prompt, and the χ vertex is the same as the K^{*0} vertex; and those which are displaced.

Chapter 2 elucidates as to how the SM fails to explain the numerous experimental observations of DM. Little is known about dark matter, except that it interacts gravitationally, and does not interact with electromagnetic radiation to any significant extent. A possible extension to the SM is to introduce a dark sector, which can contain a rich variety of distinct particles operating through forces that are hitherto unknown. Dark sector particles would be gauge singlet states with respect to the SM, and only be able to communicate with known particles via weakly interacting messenger particles through one of four *portals*: the vector, axion, Higgs, and neutrino portals [89]. Interaction terms for messengers in each of these portals are given in Table 7.1. Theories involving dark sectors that are weakly coupled to the SM are tremendously attractive because it is relatively easy to construct a complex theory accommodating various unexplained phenomena, while having little impact on the SM observables. These theories are therefore difficult to rule out.

The Higgs portal has a scalar messenger particle which can mix with the SM Higgs. There

[1] Throughout this chapter, the symbol χ shall denote a general dark boson and all references to a K^{*0} will be implicitly referencing the $K^*(892)^0$; unless explicitly stated otherwise.

Portal	Particles	Operator(s)		
Vector	Dark photons		$-\frac{\epsilon}{2\cos\theta_W}F_{\mu\nu}F'^{\mu\nu}$	
Axion	Pseudoscalars	$\frac{\chi}{f_\chi}F_{\mu\nu}\widetilde{F}^{\mu\nu}$	$\frac{\chi}{f_\chi}G^a_{\mu\nu}\widetilde{G}^{\mu\nu}_a$	$\frac{\partial_\mu\chi}{f_\chi}\overline{\psi}\gamma^\mu\gamma^5\psi$
Higgs	Dark scalars		$(\mu\chi + \lambda\chi^2)H^\dagger H$	
Neutrino	Sterile neutrinos		$Y_N\ell H\chi$	

Table 7.1: A summary of portals through which a new dark boson could operate, as given in Ref. [89]. Terms are defined as: $F_{\mu\nu}$ is the field strength tensor of the photon; $F'_{\mu\nu}$ is the dark photon field; ϵ characterises mixing between the SM and the dark photon; f_χ is scale at which Peccei-Quinn global $U(1)$ symmetry is spontaneously broken; $G_{\mu\nu}$ is the gluon field strength tensor; a dark scalar field has coupling strengths μ and λ to the Higgs field; and the sterile neutrino couples to a H with a strength Y_N.

are a number of models which incorporate a scalar messenger particle that interacts with the Higgs. One class of models incorporate an *inflaton*, which is the quanta of the hypothesised inflaton field responsible for the inflationary period of the Universe, beginning around $t = 10^{-36}$ s. Figure 7.1 shows the allowed parameter space of the mixing angle between the DM and Higgs, θ, as a function of mass. It is possible that inflatons are light, in the range $270 < m_\chi < 10^4$ MeV [90], and might therefore be accessible in the decay $B^0 \to K^{*0}\chi$. These models also help to solve other problems, such as the BAU [91, 92].

Chapter 2 introduces the idea of PQ symmetry breaking leading to an axion which resolves the strong *CP* problem. Unlike other dark boson portals, the axion portal introduces a term in the Lagrangian which couples messenger axions to fermions directly. In order for the axion portal to couple to a dark sector containing TeV-scale DM, the messenger particles are predicted to have a mass in the range $360 < m_\chi < 800$ MeV and a decay constant in the range $1 \lesssim f_\chi \lesssim 3$ TeV [96]. Figure 7.2 shows a Feynman diagram of how the decay $B^0 \to K^{*0}\chi$ might proceed; it shows an FCNC where the χ results from a coupling to a t quark. Therefore, searching for evidence of a $B^0 \to K^{*0}\chi$ where $\chi \to \mu^+\mu^-$ is particularly sensitive to portals which couple strongly to mass, as is the case for axions.

Were the dark sector to be gauged under a different group to the SM, the $U_Y(1)$ generator of the SM could kinetically mix with with generator of the dark $U(1)$ group. This would give rise to a particle, often called a *dark photon* (A'), interacting via the vector portal.

In principle, the following analysis is sensitive to any dark sector particle. Practically, other experiments have searched directly for dark bosons with mass-independent couplings using much larger data samples. For example, the NA48/2 collaboration has searched for a dark photon directly in the decay $\pi^0 \to \gamma A'$ [97], and the BaBar collaboration have searched for evidence in the decay $e^-e^- \to \gamma A'$ [98]. The coupling here is mass independent, and therefore this search is less sensitive to dark photons in comparison to these direct production searches.

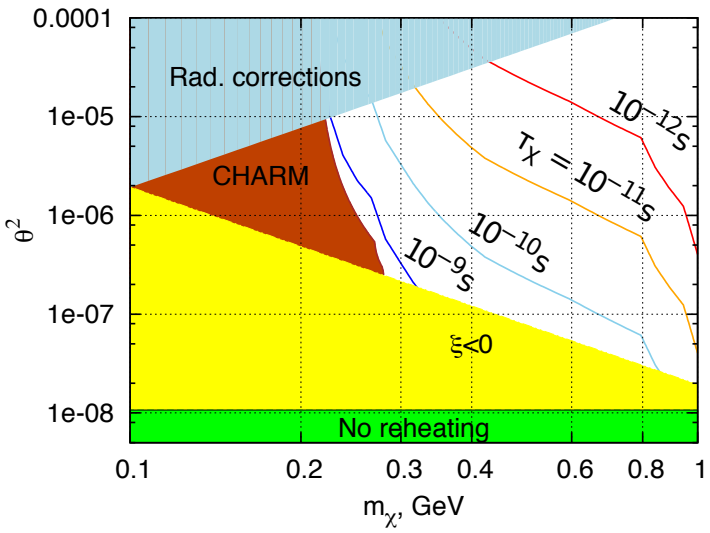


Figure 7.1: Plot taken from Refs. [93, 94], showing the allowed and excluded regions of the Higgs-inflaton mixing parameter squared, θ^2, as a function of the inflaton mass. Shaded areas indicates the parameter spaces excluded by: (blue) large radiative corrections to the inflaton potential; (red) an experiment by the CHARM collaboration [95]; (yellow) the coupling constant between the inflaton and gravity, ξ, is required to be greater than zero; and (green) there being insufficient mixing between the inflaton and Higgs field for there to be any reheating. Lines indicate specific inflaton lifetime contours for this model.

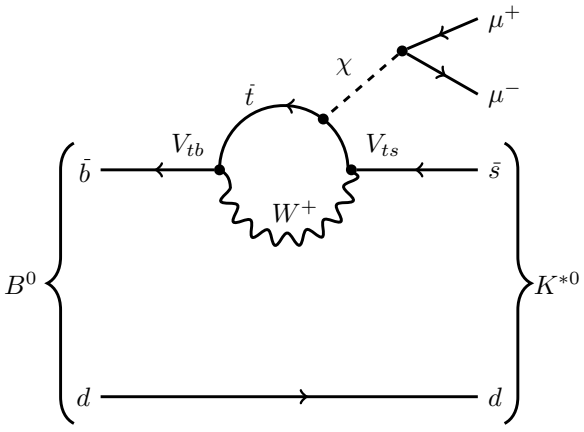


Figure 7.2: Feynman diagram showing the decay $B^0 \to K^{*0}\chi$, and $\chi \to \mu^+\mu^-$. Depending on the portal through which the χ acts, it couples directly to the muons, or mixes with a SM Higgs, Z, or photon.

SUSY could also be restricted, at energies reached thus far, to a dark sector. It is known that SUSY is a broken symmetry and thus must have associated Goldstone particles: a fermionic *goldstino* and associated super-partners called *sgoldstinos*, which are scalar and pseudoscalar. In some models, the sgoldstinos are the messenger particles between the SM and the dark SUSY sector. After SUSY breaking, the goldstino becomes the longitudinal component of the gravitino, and the sgoldstinos are massless particles, which gain mass from corrections at higher orders. Then, after electroweak symmetry breaking, these sgoldstinos interact with SM fermions via Yukawa-like interactions, but suppressed by the SUSY breaking parameter, F, [99]. This is interesting because although the sgoldstino masses are unknown, a measurement of its coupling to fermions would give access to F and the scale of SUSY breaking since $\sqrt{F} \sim \Lambda_{\rm SUSY}$. The suppression of the coupling between the sgoldstinos and fermions means that the larger the scale of $\Lambda_{\rm SUSY}$, the longer the lifetime of the sgoldstinos.

Naïvely, one might expect that in the case that χ is a scalar or pseudoscalar, the decay $B^0 \to K^+\mu^+\mu^-$ to be more sensitive than $B^0 \to K^{*0}\mu^+\mu^-$. The latter mode for a spin-0 χ requires an orbital angular momentum of one, because the B^0 is a pseudoscalar and the K^{*0} is a vector, and therefore leads to some suppression due to a barrier factor. However, this suppression is only significant at high dimuon masses, close to threshold, where the phasespace is restricted. A further complication of using the decay $B^+ \to K^+\mu^+\mu^-$ for this analysis is the lack of a good quality B^0 decay vertex. Decay rate predictions for a χ operating through the axion portal for decays of the type $B \to K\chi$ to be [100]:

$$\Gamma\big(B \to K\chi\big) = \Gamma_0 \frac{\lambda_K \big(m_B^2 - m_K^2\big)^2}{m_B^6} \big[f_0\big(m_\chi^2\big)\big] \tag{7.1}$$

$$\Gamma\big(B \to K^*\chi\big) = \Gamma_0 \frac{\lambda_{K^*}^3}{m_B^6} \big[A_0\big(m_\chi^2\big)\big] \tag{7.2}$$

where the phasespace factor is

$$\lambda_\kappa = \Big[\big(m_B^2 - m_\chi^2 - m_\kappa^2\big)^2 - 4m_\chi^2 m_\kappa^2\Big]^{\frac{1}{2}}, \tag{7.3}$$

and form factors are denoted as f_0, A_0, and Γ_0 is a constant. Figure 7.3 shows that the phasespace factor is the dominant factor in the shape for all the decays, and that searching for a new scalar or vector particle in the decays $B^0 \to K^{*0}\chi$ and $B^+ \to K^+\chi$ is similarly sensitive for $m_\chi \lesssim 2000\,\text{MeV}$, but at high masses $B^+ \to K^+\chi$ is more sensitive.

The following chapter introduces the analysis strategy, an overview of the selection, and how the discrete samples of $B^0 \to K^{*0}\chi$ are used to parameterise various distributions at all masses. The analysis is performed blindly, and some results of the unblinding procedure are given in the final selection along with a calculated p-value.

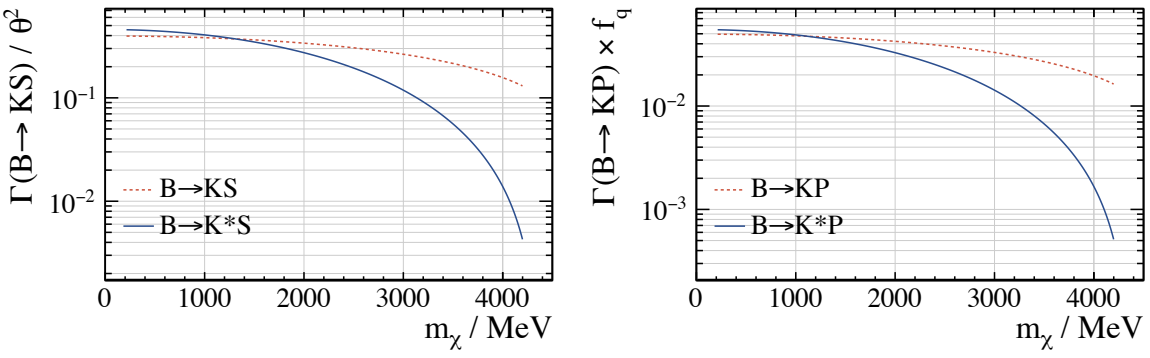


Figure 7.3: Decay rate predictions, from Eq. 7.3, for decays of the form $B \to KX$, where X is either (left) a scalar (S) or (right) an axial-vector (P). The parameters θ and f_q are parameters of the model in Ref [100]. The shapes of the curves are dominated by the available phasespace, and sensitivity is comparable for $m_\chi \lesssim 2000\,\text{MeV}$.

7.2 Analysis strategy

This analysis is a completely general search for a particle with unknown mass, m_χ, and lifetime, τ_χ. Exhaustive details of the analysis strategy can be found in Ref. [101], but the following section will outline the important points. Broadly, the search involves a frequentist scan of the dimuon invariant mass spectrum, separated into two bins in decay time, from $B^0 \to K^{*0}\mu^+\mu^-$ for an excess of events consistent with $\chi \to \mu^+\mu^-$. Firstly, an explanation of the search in the mass dimension will be presented, and then extended to deal with the decay time dimension.

7.2.1 Searching in the mass dimension

A scan in mass is performed where the test mass, m_t, is incremented in steps of $\frac{1}{2}\sigma_m$, where σ_m is the local mass resolution defined at m_t. In this analysis, σ_m lies in the range $1 \lesssim \sigma_m \lesssim 6\,\text{MeV}$. At each m_t signal and background regions are defined as

$$|m - m_t| < 2\sigma_m \tag{7.4}$$

and

$$3\sigma_m < |m - m_t| < (2x + 3)\sigma_m, \tag{7.5}$$

respectively. The width of the sideband region with respect to the signal region is defined by x, and the signal region extends $2\sigma_m$ either side of the test mass, and therefore will contain 95% of all signal events. A gap of $1\sigma_m$ separates the signal and background regions to allow for leakage of signal candidates beyond $2\sigma_m$. Using two sideband regions, one

either side of the signal, means the background contribution to the signal region can be estimated by assuming that the background is, on average, linear over the range of interest.

At each mass point in the search the number of events in the signal region, n_s, and the number of events in the background region, n_b, are counted. If the background is, on average, linear in the local mass region, and if there is no signal contribution: $\langle n_s \rangle = \frac{1}{x}\langle n_b \rangle$. However, if there is evidence of signal, then $n_s = s + b$, where s is the number of signal events, and b is the background component in the signal region, as estimated from the sideband regions. Thus, one can construct a likelihood:

$$\mathcal{L}(n_s, n_b|s, b) = \mathcal{P}(n_s, s + b) \cdot \mathcal{P}(n_b, xb), \tag{7.6}$$

where $\mathcal{P}(n, \lambda)$ is the probability of observing n from a Poisson distribution parameterised by λ. This is simply the likelihood of the background estimate from the sidebands fluctuating to the observed number of events in the signal region.

The likelihood given in Eq. 7.6 assumes that the background is, on average, exactly linear, and that the sideband region is always factor of x larger than the signal region. In reality, this is not precisely true, and local deviations mean that the actual scaling factor is an unknown y, with an uncertainty σ_y, which is accounted for by modifying the likelihood function to be

$$\mathcal{L}(n_s, n_b, x|s, b, y) = \mathcal{P}(n_s, s + b) \cdot \mathcal{P}(n_b, yb) \cdot \mathcal{G}(y, x, \sigma_y). \tag{7.7}$$

Here, $\mathcal{G}(n, \mu, \sigma)$ is the probability of observing n given a Gaussian distribution with a mean, μ, and standard deviation, σ. This modification allows the uncertainty on the background shape to be immediately accounted for in the method, meaning that no additional systematic uncertainty is required.

The profile likelihood, Λ, is used to extract the p-values, it defined as:

$$\Lambda(s|n_s, n_b) = \frac{\mathcal{L}(s, \hat{b}(s), \hat{y}(s)|n_s, n_b, x)}{\mathcal{L}(\hat{s}, \hat{b}, \hat{y}|n_s, n_b, x)}, \tag{7.8}$$

where $\hat{s}$, $\hat{b}$, and $\hat{y}$ are chosen to maximise the likelihood; the functions $\hat{b}(s)$ and $\hat{y}(s)$ maximise the likelihood for a given s. The function $-2\ln\big(\Lambda(s|n_s, n_b)\big)$ behaves like a χ^2 distribution with one degree of freedom.

7.2.2 Searching in the lifetime dimension

Sensitivity to the lifetime of the χ, τ_χ, is introduced by splitting the data at each test mass into two regions: a prompt and a displaced region, defined by $\tau < 3\sigma_\tau$ and $\tau > 3\sigma_\tau$

respectively; where σ_τ is defined to be the local decay time resolution. The joint likelihood is the product of the two individual likelihoods:

$$\mathcal{L}(n_s^p, n_b^p, n_s^d, n_b^d, x | s^p, b^p, y^p, s^d, b^d, y^d) = \mathcal{L}(n_s^p, n_b^p, x | s^p, b^p, y^p) \times \mathcal{L}(n_s^d, n_b^d, x | s^d, b^d, y^d), \tag{7.9}$$

where superscript $p(d)$ denotes the prompt(displaced) region. The combined profile likelihood is:

$$\Lambda_{\text{Tot}} = \Lambda(s^p | n_s^p, n_b^p) \cdot \Lambda(s^d | n_s^d, n_b^d). \tag{7.10}$$

Then, $-2\ln\left(\Lambda_{\text{Tot}}\right)$ behaves as a χ^2 distribution with two degrees of freedom. Since this is analytically solvable, a p-value can be calculated quantifying how incompatible the number of events observed in the signal region is with the null hypothesis of zero signal events. This is done for each test mass.

The information supplied by the addition of two bins in the lifetime dimension is approximately optimal for all χ lifetimes, except for when $\tau \sim 3\sigma_\tau$. In this case, it is marginally more optimal to include shape information of the background distribution from the sidebands; but this introduces significantly more complications to the analysis. Therefore background shape information is not used.

7.2.3 Calculation of a global p-value

At each m_t, a p-value is calculated using the profile likelihood of the joint likelihood given in Eq. 7.9. The *look-elsewhere effect* is the phenomenon whereby an experiment yields a statistically significant result by chance, as a simple consequence of making a number of measurements. In the case of this analysis the look-elsewhere effect must be accounted for because around a thousand local p-values are measured in the course of the single experiment. This is accounted for by calculating the conversion of the minimum local p-value to a global p-value using an ensemble of toys.

After the p-value at each m_t is calculated, the region in which the lowest p-value is observed will be isolated and removed from the sample; leaving data which is entirely background-like. This is assuming that there is only one NP particle to be observed in this analysis. The remaining background-like distribution is turned into a Probability Density Function (PDF), where the region that has been removed is interpolated across. Toy datasets can then be generated from this PDF.

This method requires the generation of $\mathcal{O}(10^7)$ datasets to probe down to a 5σ global p-value. If the minimum local p-value observed is less than any of the local p-values in the toy datasets then the asymptotic formula will be used [102]. The asymptotic formula uses a few toy datasets to set an upper bound on the conversion of local to global minimum

p-value. This bound is given by

$$P_{\text{global}} \leqslant P(\chi^2_s > c) + \langle N(c) \rangle, \tag{7.11}$$

where s is the number of degrees of freedom of the χ^2. In this case $s = 2$. The function $N(c)$ returns the number of *upcrossings* of a dataset above the level of c, where an upcrossing is defined to be when the p-value crosses from below the value of c, to above it [103]. The value of c for a given p-value is its significance squared. Clearly, for large p-values, the ability to obtain a reliable value of $N(c)$ will be highly dependent on the toy datasets generated, and therefore an approximation is needed. Reference [102] shows that Eq. 7.11 can be written as

$$P_{\text{global}} \leqslant P(\chi^2_s > c) + \langle N(c_0) \rangle \exp\left(-\frac{c - c_0}{2}\right) . \left(\frac{c}{c_0}\right)^{\frac{s-1}{2}}, \tag{7.12}$$

where c_0 is the number of upcrossings for a small value of c, optimally $c = s - 1$. Therefore, once $\langle N(c_0) \rangle$ has been determined using a few hundred toy datasets, the asymptotic boundary is trivial to calculate. This can be used to check the final conversion, and used to calculate the global p-value should a significance of greater than $5\,\sigma$ be observed.

7.2.4 Limit setting

Upper limits will be set as a function of m_χ and τ_χ. This requires further modification to the likelihood function to account for the relationship between the number of signal events in the prompt and displaced regions for a given value of decay time. Additional Gaussian factors are added: one to account for the uncertainty in the fraction of signal events that are observed in the two lifetime regions, and another to account for the uncertainty in the efficiency ratio with respect to the normalisation channel. The resulting likelihood for a given m_χ and τ_χ is

$$\begin{aligned}\mathcal{L}\big(n^p_s, n^p_b, n^d_s, n^d_b, x, \tau | \dots\big) = \mathcal{L}\big(n^d_s, n^d_b, x | \varepsilon s f, b^p, y^p\big) \\ \times \mathcal{L}\big(n^p_s, n^p_b, x | \varepsilon s (1-f), b^d, y^d\big) \\ \times \mathcal{G}\big(f, \overline{f}(\tau_\chi), \sigma(f)\big) \times \mathcal{G}\big(\varepsilon, \bar{\varepsilon}(\tau_\chi), \sigma(\varepsilon)\big),\end{aligned} \tag{7.13}$$

where the fraction of signal events in the prompt region is given by f, which has an expected value from simulation $\overline{f}$ and an uncertainty $\sigma(f)$. The same nomenclature is used for the efficiency measured relative to the normalisation channel, ε.

The normalisation channel used will be the SM decay $B^0 \to K^{*0}\mu^+\mu^-$, restricted to the region $1.1 < q^2 < 6.0\,\text{GeV}^2$. This range is chosen to minimise theoretical uncertainties at low and high q^2 and reduce experimental uncertainties by removing the region around the

Meson (X)	Mass	Width	$\mathcal{B}(B^0 \to K^{*0}X)$	$\mathcal{B}(X \to \mu^+\mu^-)$	$\mathcal{B}_{\rm tot}(B^0 \to K^{*0}\mu^+\mu^-)$
η	547.9	0.001	$(1.59 \pm 0.10) \times 10^{-5}$	$(5.8 \pm 0.8) \times 10^{-6}$	$(9.2 \pm 1.4) \times 10^{-11}$
ρ	775.3	147.8	$(3.9 \pm 1.3) \times 10^{-6}$	$(4.55 \pm 0.28) \times 10^{-5}$	$(1.8 \pm 0.6) \times 10^{-10}$
ω	782.7	8.5	$(2.0 \pm 0.5) \times 10^{-6}$	$(9.0 \pm 3.1) \times 10^{-5}$	$(1.8 \pm 0.8) \times 10^{-10}$
ϕ	1019.5	4.3	$(1.00 \pm 0.05) \times 10^{-5}$	$(2.87 \pm 0.19) \times 10^{-4}$	$(2.9 \pm 0.2) \times 10^{-9}$
D^0	1864.8	10.1	$(4.2 \pm 0.6) \times 10^{-5}$	$< 6.2 \times 10^{-9}$	$< 2.6 \times 10^{-14}$
J/ψ	3096.9	0.093	$(5.96 \pm 0.03) \times 10^{-2}$	$(1.32 \pm 0.06) \times 10^{-3}$	$(7.9 \pm 0.4) \times 10^{-5}$
$\psi(2S)$	3686.1	0.299	$(7.9 \pm 0.9) \times 10^{-3}$	$(6.0 \pm 0.4) \times 10^{-4}$	$(4.7 \pm 0.6) \times 10^{-6}$

Table 7.2: Selected particle properties for mesons which decay into a dimuon pair final state, and the branching fractions for the relevant decays [14]. Central values of mass and width are given in MeV.

ϕ meson which is centred at $q^2 \simeq 1.04\,\text{GeV}^2$.

7.2.5 Resonant backgrounds

In the absence of resonances, the dimuon background from the SM decay $B^0 \to K^{*0}\mu^+\mu^-$ is expected to be locally-linear to within 1 % over the entire mass range. However, the inclusion of resonances can lead to significant departures from linearity, dependent upon the resonance's width (Γ), magnitude, and the value of x. Table 7.2 lists a number of mesons that decay to a dimuon final state and could contribute as background to the decay $B^0 \to K^{*0}\mu^+\mu^-$. Wide resonances, $\Gamma \gtrsim 20\sigma_m$, such as the ρ, are sufficiently wide not to be problematic, even if they dominate the local background distribution. Conversely, narrow resonances where $\Gamma \lesssim 5\sigma_m$ lead to significant deviations from a locally-linear background and must be vetoed. Dimuon decays of the ϕ, J/ψ, and $\psi(2S)$ mesons have the highest branching fractions, and are also among the narrowest resonances; they are therefore vetoed. Intermediate resonances, $5 \lesssim \Gamma \lesssim 20\sigma_m$, are considered on a case-by-case basis since they are only troublesome if they account for a large fraction of the local background. For this analysis, these ranges roughly translate to requiring that resonances with $\Gamma < 25\,\text{MeV}$ will be vetoed, and those with $\Gamma > 125\,\text{MeV}$ shall be ignored. Other resonances in Table 7.2 are broad, and contribute to the dimuon structures at low mass.

It is shown in Ref. [101] that the local-linearity approximation is accurate to $\sim$5 % in regions where there may be contributions from wide resonances, below the mass of the J/ψ. This means that a value of $\sigma_y = 0.05$ is appropriate. The results are rather insensitive to the choice of x, in this region a value of $x = 5$ is used.

The resonances with natural widths that fall in the intermediate region $5 \lesssim \Gamma \lesssim 20\sigma_m$, comprise of various $c\bar{c}$ states with masses above that of the $\psi(2S)$. All these resonances, from the $\psi(3770)$ up to higher masses, are observed to be relatively wide, with $\Gamma \sim 70\,\text{MeV}$; for example, there is contribution from the decay $\psi(4160) \to \mu^+\mu^-$ which was first observed

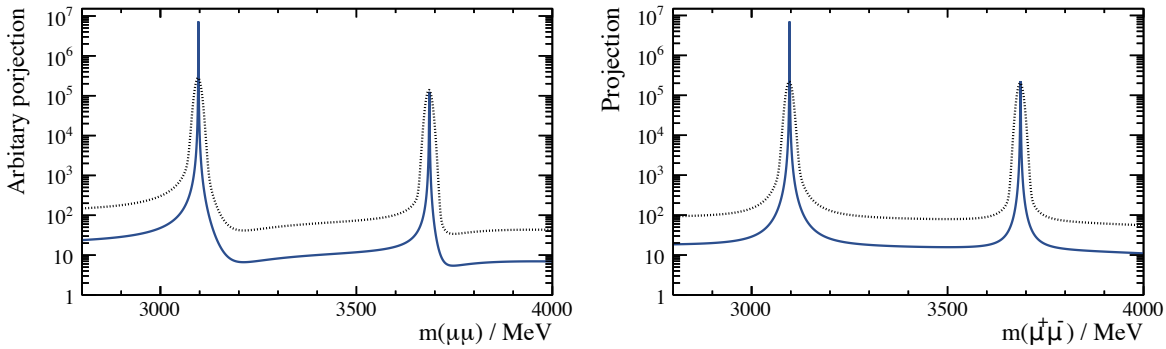


Figure 7.4: Theoretical distributions of the decays $J/\psi \to \mu^+\mu^-$ and $\psi(2S) \to \mu^+\mu^-$ interfering with a non-resonant dimuon component using a model in Ref. [105] with (left) no phase difference, and (right) maximal phase difference. The solid blue line shows the raw model, and the dotted line is the same distribution which has been convolved with a Gaussian to account for detector resolution effects.

by LHCb in the decay $B^+ \to K^+\mu^+\mu^-$, and has a width of $(65\,^{+22}_{-16}\,\text{MeV}$ [104] . Since there are many resonances in this region the background shape is more uncertain. This is accounted for by reducing the size of the sidebands by taking $x = 1$. With this choice the value of σ_y was chosen by applying the method to candidate data from the decay $B^+ \to K^+\mu^+\mu^-$, giving a minimum local p-value of 0.3(0.4) for $\sigma_y = 0(0.1)$, thus no false discovery claim could be made with either choice. A value of $\sigma_y = 0.1$ is taken.

Interference effects between $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to K^{*0}\rho$, where $\rho \to \mu^+\mu^-$, could lead to non-linearity in the background shape, despite the fact that the total branching fractions of $B^0 \to K^{*0}\rho(\to \mu^+\mu^-)$ decays are so small that less than one event is expected in the data sample. However, this deviation can only be a maximum of $\sim$5 %, which is already accounted for in the choice of $\sigma_y = 0.05$.

The contribution from the decay $\phi \to \mu^+\mu^-$ to the $K^{*0}\mu^+\mu^-$ final state is removed by excluding dimuon candidates in the range $1000 < m_{\mu^+\mu^-} < 1040\,\text{MeV}$. Vetoed regions for the J/ψ and $\psi(2S)$ are calculated using a theoretical model which can account for the interference effects between the $c\bar{c} \to \mu^+\mu^-$ resonances and the non-resonant $\mu^+\mu^-$ component [105]. The model is used to generate PDFs with different phase differences between the resonant and non-resonant components, which is then convolved with a double Gaussian function to account for detector resolution effects; these are shown in Fig. 7.4. After smearing, the PDFs are then used to calculate the regions around the J/ψ and $\psi(2S)$ where the background is locally linear to better than 5 % for various values of x. The resulting veto regions are very similar to those used in Ref. [76], except that the upper boundary of the $\psi(2S)$ veto is extended to cover the $\psi(3770)$. All the vetoed regions to remove narrow resonances are summarised in Table 7.3.

Resonance(s)	Vetoed region (MeV)
ϕ	$1000 < m_{\mu^+\mu^-} < 1040$
J/ψ	$2960 < m_{\mu^+\mu^-} < 3204$
$\psi(2S)$, $\psi(3770)$	$3614 < m_{\mu^+\mu^-} < 3875$

Table 7.3: Summary of the vetoes regions in $m_{\mu^+\mu^-}$ to remove the contributions from the decays $\phi \to \mu^+\mu^-$ and various narrow $\psi \to \mu^+\mu^-$ to the dimuon distribution.

τ_χ (ps)	m_χ (MeV)										
10									2500		
100	214	220	235	250	500	800	1000	1500	2000	2500	4000
1000				250					2500		

Table 7.4: Samples of simulated $B^0 \to K^{*0}\chi$ generated with given mass and lifetime. A total of 1.5 million events are generated for each sample, but only 150 000 for the samples with a m_χ of 220 and 235 MeV.

7.3 Selection

This analysis uses the full $3.0\,\text{fb}^{-1}$ of data collected by the LHCb experiment [45] in the years 2011 and 2012. Since the properties of the χ are unknown, a variety of simulated samples of the decay $B^0 \to K^{*0}\chi$, where $\chi \to \mu^+\mu^-$, were generated with a range of m_χ and τ_χ, a summary of these are shown in Table 7.4. Studies are also performed using simulated events of the decays $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to J/\psi\, K^{*0}$.

Reconstructed decays of $B^0 \to K^{*0}\chi$ that must be selected by the L0 triggers for muon, dimuon, or hadronic candidates. Subsequent trigger levels require that the decay has a topology consistent with a B-meson decaying in to a multi-body final state which includes muons. Only `TOS` candidates are used in this analysis for two reasons: firstly, the ratio of trigger efficiency for the SM $B^0 \to K^{*0}\mu^+\mu^-$ to that of (possibly displaced) χ mode enters into the limits and, thus, must be precisely determined; and the use of `TIS` events would come with a substantial enhancement of the di-b-hadron backgrounds. The few percent gain in signal efficiency obtained using `TIS` is not worth the increase in these backgrounds di-b-hadron backgrounds.

Since the χ could be long lived, its decay vertex can lie downstream of the B^0 vertex, and could fly far enough to leave acceptance of the VELO before it decays. It is therefore wise to consider different *track types*. For most LHCb analyses, the only appropriate tracks to use are *long* tracks — as have been used implicitly throughout this thesis. Long tracks are fitted using hits in the tracking stations T1–3, the VELO, and can use TT information. But, it is also possible to reconstruct tracks using only some of the tracking stations; for example, a *downstream* track is reconstructed using hits in the TT and

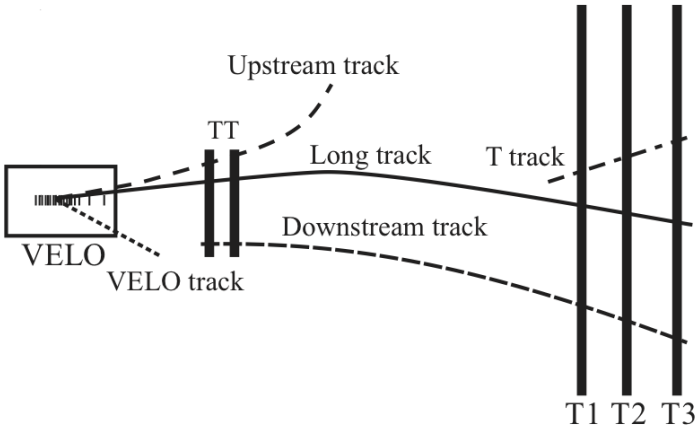


Figure 7.5: Schematic diagram of the LHCb tracking system, showing the VELO, TT, and tracking stations T1–3. Track types at LHCb are defined by the regions through which they travel, as labelled in the diagram. Most tracks used in analyses are classified as long tracks, but for long-lived particles (such as the K^0_S) downstream tracks are often also used to increase statistics. This analysis focused on long tracks.

tracking stations T1–3, but not the VELO. Figure 7.5 shows the various definitions of track types. The problem with using downstream tracks is that they are not triggered efficiently in HLT2. For example, a simulated χ with a mass of 250 MeV and a lifetime of 100 ps has a reconstruction and stripping efficiency of $\sim$0.9 % if the muons are both long tracks, but the equivalent number for downstream tracks is $\sim$2.5 %; due to the boost of a light object from a decaying B^0. However, the trigger efficiency for this sample is $\sim$45 % and 8 % for long and downstream candidates respectively. There is therefore a factor two more long-track candidates for the 250 MeV 100 ps sample. This is factor grows to a few hundred as the dark boson mass increases or lifetime decreases. For this reason, this analysis deals only with χ candidates formed from long-track muons.

The offline selection criteria applied in the stripping are outlined in Table 7.5. This table lists the variables $\mathcal{P}_\pi$ and $\mathcal{P}_K$ which are MVA algorithms giving a probability-like response between zero and one, quantifying how pion or kaon like a particle is. A similar parameter is the ghost probability, $\mathcal{P}_{\text{gh}}$, which quantifies the chance that a given track is actually a ghost — in other words, an erroneous track made from random combinations of hits [106]. The full decay chain $B^0 \to K^{*0}\chi(\to \mu^+\mu^-)$ is reconstructed using a fit in which the B^0 mass is constrained to its known value [14]. All references to m_χ, $m_{K^+\pi^-}$ or τ_χ are to the values after this vertex fit has been performed.

7.3.1 Preselection

After the stripping and triggering stages, a loose preselection is applied, cutting on both topological and PID quantities; all these cuts are summarised in Table 7.6. The topological selection requirements are approximately 90 % signal efficient on all simulated samples.

Candidate	Selection criterion			
B^+	$\chi^2_{\rm vtx}/{\rm ndf}$	$<$	25	
	$\chi^2_{\rm IP}$	$<$	50	
	τ	$>$	0.2	ps
	m	$\in$	$[4800, 5800]$	MeV
	$p_{\rm T}$	$>$	1000	MeV
	$\cos\theta_{\rm dir}$	$>$	0	
χ	$\chi^2_{\rm vtx}/{\rm ndf}$	$<$	10	
	$\chi^2_{\rm FD}$	$<$	25	
	$p_{\rm T}$	$>$	250	MeV
	DOCA	$<$	0.2	mm
	DOCA χ^2	$<$	25	
Tracks	$\chi^2_{\rm trk}/{\rm ndf}$	$<$	3	
	$\min(\chi^2_{\rm IP})$	$>$	9	
	$\mathcal{P}_{\rm gh}$	$<$	0.3	
K^+, π^+	$p_{\rm T}$	$>$	250	MeV
	p	$>$	2000	MeV
	$\chi^2_{\rm IP}$	$>$	9	
K^+	$\mathcal{P}_K$	$>$	0.1	
π^+	$\mathcal{P}_\pi$	$>$	0.2	
μ^+	$p_{\rm T}$	$>$	100	MeV
	`PIDmu`	$>$	-5	
	`isMuon`		True	

Table 7.5: Selection criteria applied to signal candidates during the LHCb stripping phase. Criteria definitions are defined in text. While the B^0 mass is constrained in the fit, the selection makes a cut on the unconstrained mass.

Selection requirements are applied to the K^* and its daughters to remove candidates that are inconsistent with the decay $K^{*0} \to K^+\pi^-$. Firstly, the $K^+\pi^-$ invariant mass must within 100 MeV of the known $K^*(892)^0$ mass, which has been measured to be (895.81 ± 0.19) MeV [14]. PID constraints are applied to each K^* daughter: ${\rm DLL}_{K\pi}(K^+) > -5$, ${\rm DLL}_{K\pi}(\pi^+) < 25$; and to ensure that the kaon candidate is more consistent with a kaon than the pion candidate a cut on the difference in DLLs is also applied: ${\rm DLL}_{K\pi}(K^+) - {\rm DLL}_{K\pi}(\pi^+) > 10$.

The decay $K^0_{\rm S} \to \pi^+\pi^-$ has a branching fraction of $(69.20 \pm 0.05) \times 10^{-2}$ [14], and is removed in the preselection by requiring that $|m_{\pi^+_\mu\pi^-_\mu} - m^{\rm PDG}_{K^0_{\rm S}}| < 25$ MeV. This roughly translates to a cut in the dimuon invariant mass spectrum of $436 < m_{\mu^+\mu^-} < 490$ MeV.

Since the χ can be displaced from the B^0 decay vertex, a potential background for the decay $\chi \to \mu^+\mu^-$ is from a $\mu^+\mu^-$ pair directly from a PV. This is suppressed by requiring that the transverse flight distance (FD_T) of the χ vertex, with respect to the PV, is greater than 0.1 mm.

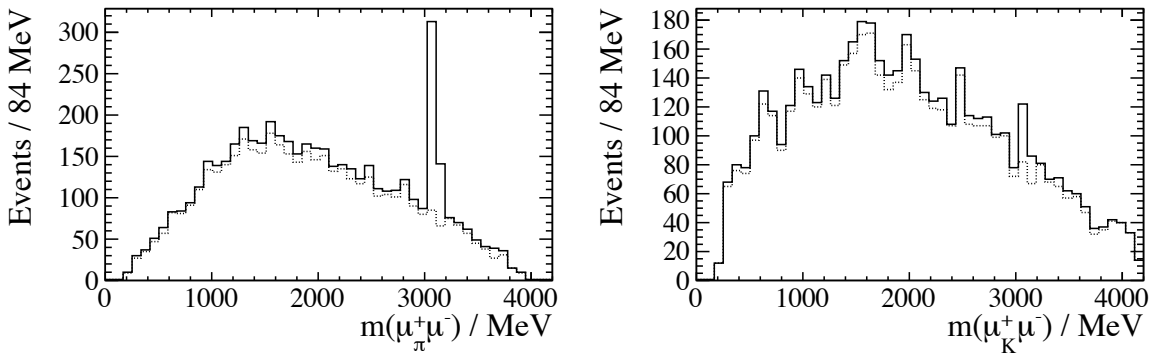


Figure 7.6: Background contributions from $B^0 \to J/\psi K^{*0}$, where both a muon and (left) pion and (right) kaon are misidentified as one another. This background is very effectively removed by requiring that the hadron does not satisfy the `isMuon` criteria; the effect of this veto is shown with a dotted line.

Candidate	Selection criterion			
B^0	θ_{dir}	$<$	0.03	rad
	χ^2_{vtx}	$<$	15	
	χ^2_{IP}	$<$	10	
K^{*0}	$\lvert m_{K^+\pi^-} - m^{\text{PDG}}_{K^{*0}}\rvert$	$<$	100	MeV
	$\text{DLL}_{K\pi}(K) - \text{DLL}_{K\pi}(\pi)$	$>$	10	
K^+	`isMuon`		`False`	
	$\text{DLL}_{K\pi}$	$>$	-5	
π^+	`isMuon`		`False`	
	$\text{DLL}_{K\pi}$	$<$	25	
χ	FD_T	$>$	0.1	mm
	m	$\notin$	$[436, 490]$	MeV

Table 7.6: Preselection criteria applied to $B^0 \to K^{*0}\chi$ candidates.

The high branching fraction of $B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}$ means that there is also contamination from decays where a hadron is misidentified as a muon, and vice versa. This background can be suppressed by requiring that neither of the K^{*0} daughters satisfy the `isMuon` criteria. Figure 7.6, shows the invariant mass of the hadron-muon pair before and after the requirement that the hadrons fail the `isMuon` criterion. A summary of all preselection cuts is shown in Table 7.6.

7.3.2 Other backgrounds from particle misidentification

Beyond backgrounds from $q\bar{q}$ states and the K^0_S decay there are many mesons that decay into a two-body final state, which could mimic the signal decay given some particle misidentification. A meson that decays via $X \to hh'$ which is then reconstructed under the

Mass criteria (MeV)		PID requirement
$\left\|m(K^+K^-_\pi) - m^{\rm PDG}_\phi\right\|$	< 10	$\mathcal{P}_\pi(\pi) > 0.3$ and $\mathcal{P}_K\ (\pi) < 0.3$
$\left\|m(K^+_\mu\pi^-_\mu) - m^{\rm PDG}_{D^0}\right\|$	< 25	$\mathcal{P}_\mu(\mu) > 0.3$
$\left\|m(p_\mu\pi^-_\mu) - m^{\rm PDG}_{\Lambda^0}\right\|$	< 10	$\mathcal{P}_p(\mu) < 0.3$
$\left\|m(p_\pi K^-\mu^+\mu^-) - m^{\rm PDG}_{\Lambda^0_b}\right\|$	< 50	$\mathcal{P}_p(\pi) < 0.2$

Table 7.7: Veto conditions to suppress double and single misidentification of particles. If, under the alternate hypothesis, the χ or K^{*0} candidate mass falls within the range indicated, the candidates are subject to the given PID requirements.

incorrect mass hypothesis could pass the selection criteria as either the χ or K^{*0} candidate. This type of contamination is studied by assigning different mass hypotheses to each final state particle and calculating the invariant mass of the $\mu^+\mu^-$ and $K^+\pi^-$ candidates. If the mass of one of these objects, after mass reassignment, is seen to peak at the mass of a known particle, then the contamination is removed by applying PID criteria.

Since a selection has been made on the $K^{*0} \to K^+\pi^-$ candidate using both PID criteria and constraints on the $K^+\pi^-$ invariant mass it is expected that there will be little contamination from background sources. To test this, candidate K^{*0} mesons coming from a B^0 candidate with an invariant mass within 80 MeV of the known B^0 mass are assigned different mass hypotheses to check for peaking components in the new $m_{K^+_h\pi^-_{h'}}$ mass spectrum. The only background that must be removed from this category is from a real $\phi \to K^+K^-$ where a kaon in the final state is misidentified as being a pion. If the mass of the $K^+K^-_\pi$ candidate lies within 10 MeV of the known ϕ mass, the ambiguous pion is subject to the requirements that $\mathcal{P}_\pi < 0.3$ and $\mathcal{P}_K > 0.3$.

Resonances decaying into a pair of hadrons which are mistaken as a pair of muons are more problematic. Weak Decays of mesons can contribute to background, especially $D^0 \to K^+\pi^-$ and $\Lambda^0 \to p\pi^-$, which are dealt with in a similar way to the vetoes described in Chap. 5. If the invariant mass of the $K^+_\mu\pi^-_\mu(p_\mu\pi^-_\mu)$ candidate falls within 25(10) MeV of the nominal $D^0(\Lambda^0)$ mass, then the muons are subject to the requirement that $\mathcal{P}_\mu(K^+_\mu, \pi^+_\mu) > 0.3(\mathcal{P}_p(p_\mu) < 0.3)$.

Misidentifying the proton as a pion in the decay $\Lambda^0_b \to pK^-\mu^+\mu^-$ can contaminate the selected $B^0 \to K^{*0}\chi$ candidates. Figure 7.7 shows the invariant mass distribution of the $p_\pi K^-\mu^+\mu^-$ system for candidates where $\mathcal{P}_p(p_\pi) > 0.2$, a clear peak is observed at the known mass of the Λ^0_b. Candidates are removed by if the mass of the Λ^0_b candidate falls within 50 MeV of the nominal Λ^0_b mass and satisfies $\mathcal{P}_p(p_\pi) > 0.2$.

The sidebands are used to estimate the level of background in the signal region. Therefore, background contributions are only problematic if they produce a narrow peaking structure in the dimuon mass, because misidentification causes the $m_{\mu^+\mu^-}$ distribution to be smeared.

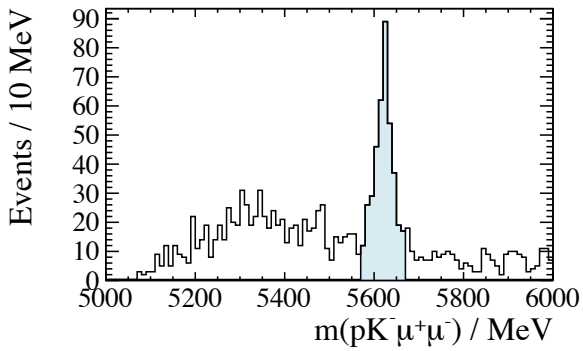


Figure 7.7: Contamination from the decay $\Lambda_b^0 \to pK^-\mu^+\mu^-$, where the proton is misidentified as a pion, here a cut of $\mathcal{P}_p(p_\pi)$ has been applied and a clear peak at the known Λ_b^0 mass, 5219.4 MeV, is observed. Candidates are vetoed if they lie within 50 MeV of the Λ_b^0 mass, which is shown as the shaded region.

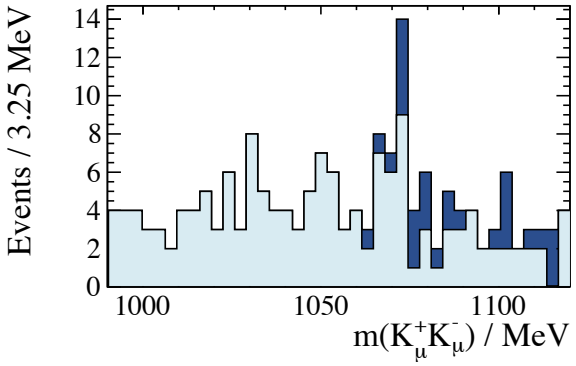


Figure 7.8: Invariant mass distribution of the $K_\mu^+ K_\mu^-$ candidates in data, showing a peak at ~1072 MeV in data. The dark and light blue regions show the distribution before and after vetoing the decay $K_S^0 \to \pi^+\pi^-$ in the preseletcion.

In general, misidentification is only problematic if the decaying particle has a very narrow natural width, so any remaining misidentification-type backgrounds have a negligible effect in the analysis.

7.3.3 Possible contamination from the $X(1070)$

While searching for potential backgrounds resulting from misidentifying two hadrons as muons, a peak is observed in the invariant mass spectrum of the $K_\mu^+ K_\mu^-$ candidates. This peak was consistent with the $X(1070)$ listed in Ref. [14], which has a mass of (1072 ± 1) MeV with a width of (3.5 ± 0.5) MeV and was observed in the $K_S^0 K_S^0$ distribution from a pion beam interacting with a liquid hydrogen target [107].

Initially the $\mu^+\mu^-$ pair under the K^+K^- mass hypothesis appears to have a contribution from a decaying $X(1070)$. Figure 7.9 shows a comparison of simulated $K_S^0 \to \pi^+\pi^-$ decays with the observed data near this excess. It is clear that $K_S^0 \to \pi^+\pi^-$ decays produce a

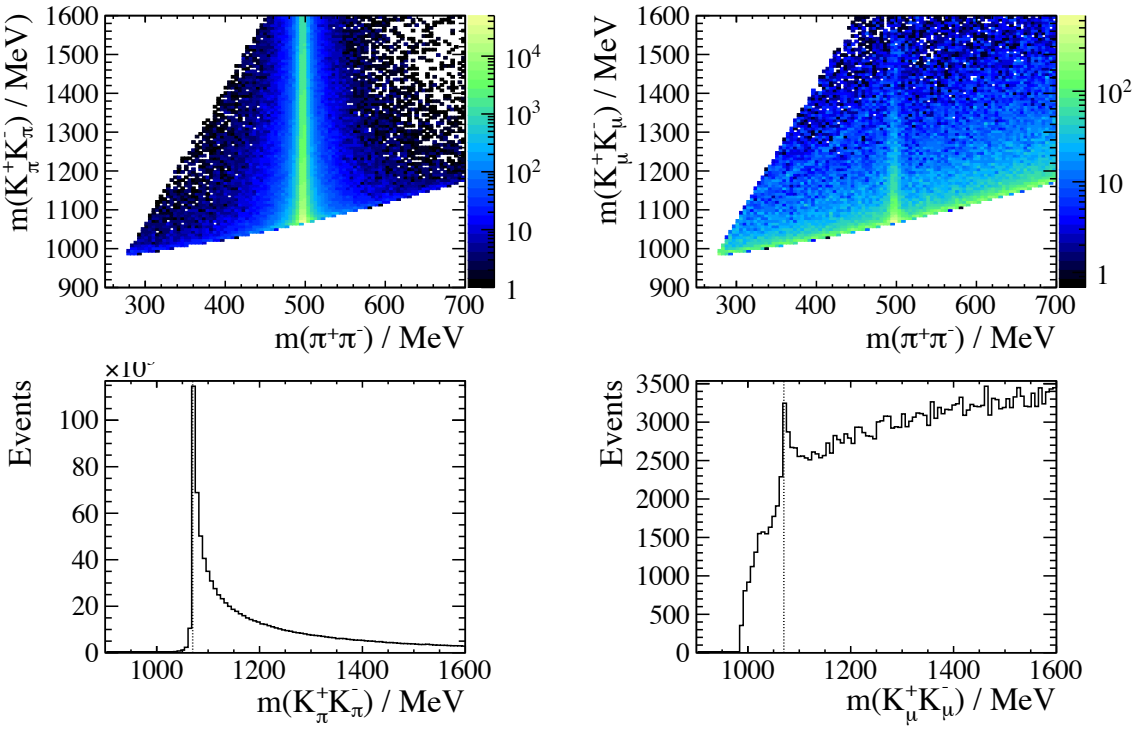


Figure 7.9: A comparison of $K^0_S \to \pi\pi$ under different mass hypotheses, for (left) simulated events, and (right) events from data. The (top) plots show the two dimensional distributions of the invariant mass distributions of a $\pi^+\pi^-$ pair and the same candidates in the K^+K^- mass hypothesis, the (bottom) plots show the projections of the K^+K^- systems. Vertical lines in the lower plots indicate 1072 MeV.

peak around 1072 MeV under the K^+K^- hypothesis. There is also a long tail but with low statistics and with a roughly uniform background this tail would not be expected to be visible in the data after the K^0_S veto in the preselection.

It is clearly not the case that the $X(1070)$ is causing the peak at $m_{K^+_\mu K^-_\mu} = 1072$ MeV, which is actually due to the decay $K^0_S \to \pi^+\pi^-$. Removing events satisfying $|m_{\pi^+_\mu \pi^-_\mu} - m^{\text{PDG}}_{K^0_S}| < 25$ MeV removes much of the peak at 1072 MeV, bringing it in line with the background. Also, the fact that no $\phi \to K^+K^-$ is observed in the $m_{K^+_\mu K^-_\mu}$ spectrum, and that the peak is narrower than the resolution of the LHCb detector, indicate that there this is a false peak and need not be accounted for further.

7.3.4 Combinatorial background and multivariate selection

The data sample is purified from combinatorial background using a multivariate selection technique. Section 4.3 outlines the uBDT algorithm, which trains a BDT whereby events are boosted not only based on misclassification, but also on how uniform the local response of the BDT is for a given set of variables. For the case of this analysis, it is important that the BDT does not bias the sample towards a given mass and lifetime, which makes

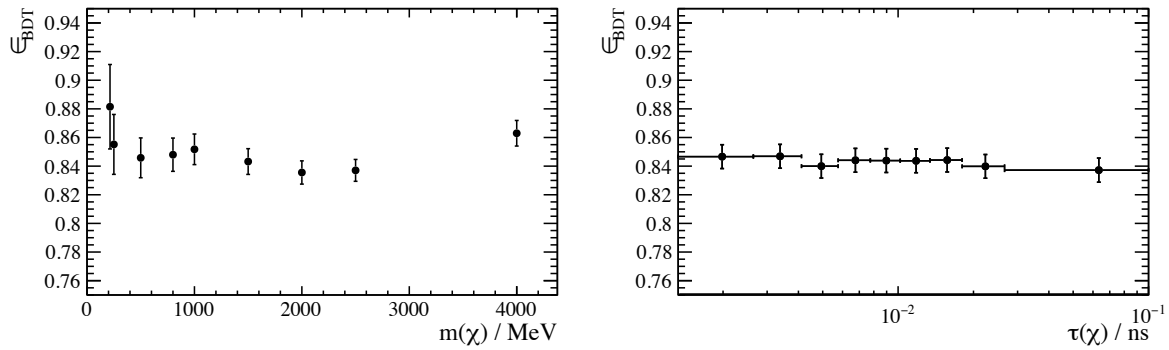


Figure 7.10: Signal efficiencies from simulation for a uBDT cut which is approximately 85 % efficient, the distributions are are approximately flat in both the (left) mass and (right) lifetime dimensions.

the uBDT ideal for this analysis.

A uBDT is trained using a signal-proxy from simulated events and background taken from the upper B^0 mass sideband where the B^0 candidate has a mass of over 150 MeV above the nominal B^0 mass. Specifically, the signal-proxy is a concoction of three different simulated samples in which the χ has a mass of: 214, 1000, and 4000 MeV; and each has a lifetime of 100 ps. These samples are chosen to give the uBDT algorithm input the largest range of masses possible; particularly a $m_\chi = 214$ MeV sample was chosen because it is close to threshold and equal to the mass of the P^0 evidenced in the Hyper-*CP* experiment [32]. It is observed that the response after a uBDT selection does result in a uniform signal efficiency in both the mass and lifetime distributions; this is shown in Fig. 7.10.

The uBDT cut is optimised by maximising the Punzi figure-of-merit [108], which is defined as

$$\text{Punzi}_{\sigma_\text{p}} = \frac{S}{\sqrt{B} + \frac{1}{2}\sigma_\text{p}}, \tag{7.14}$$

where S and B are the signal and background yields, respectively; and σ_p is the desired level of observation. The signal and background yields are calculated as follows: S is the number of simulated signal events which survive a uBDT cut; B is the background yield as estimated using $B^0 \to K^{*0}\chi$ candidates that fall outside 80 MeV from the nominal B^0 mass [14]. The invariant mass distributions of these candidates in data are fit to a decaying exponential, and the background yield is taken to be the integral of this exponential within 60 MeV (approximately 3σ) of the known B^0 mass.

The figure-of-merit is considered separately for both the prompt and displaced candidates. The optimal working point varies from sample to sample due to the fact that the background yield depends on (m, τ). However, the optimal point in the prompt case is approximately uBDT > 0.15 for all samples, a working point that is approximately 85 % efficient by construction. For this efficiency there os am estimated zero background in for a given value of m_χ. This point was calculated using a value of $\sigma_\text{p} = 5$, Eq. 7.14, although the

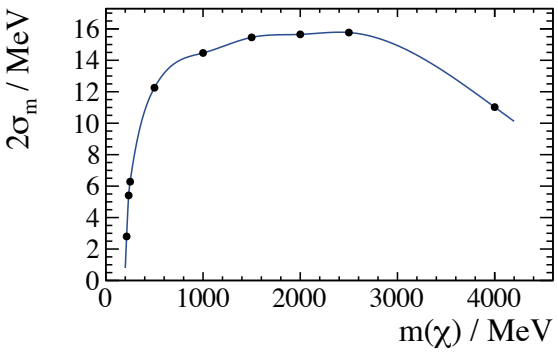


Figure 7.11: Evolution of $2\sigma_m$ as a function of mass. The mass distribution of each simulated signal sample is fit to a distribution constructed of two Gaussian functions with the same mean, the value of $2\sigma_m$ is then extracted by finding the 2σ interval of the fitted distribution. Each black point shows the value of $2\sigma_m$ for a given simulated $B^0 \to K^{*0}\chi$ sample, and the red line is the cubic spline intersecting each point.

result is seen to be insensitive to its exact value.

7.4 Evolution of parameters as a function of mass

Searching for a χ of unknown mass and lifetime requires knowledge of parameters such as: mass resolution, lifetime resolution, and efficiency; at every point on the (m_χ, τ_χ) plane that is searched. This is clearly not possible by generating an arbitrary number of simulated signal events. Instead, the available simulated samples are used to interpolate (and extrapolate) to all masses and lifetimes of interest.

7.4.1 Mass resolution of the χ candidate

The size of the signal and background regions are defined in terms of the local mass resolution, which varies across the whole mass range. To understand the evolution of σ_m, the mass distribution of various signal samples are fitted to a function constructed as the sum of two Gaussian distributions with the same mean, this is known as a double Gaussian. Fitted distributions are used to define the 2σ intervals, and then each point is intersected with a cubic spline. Figure 7.11 shows the resulting function. The mass resolution is observed to be $\sim$1 MeV for very low m_χ and quickly increases to a plateau around $2\sigma = 15$ MeV before dropping off again, because the invariant mass of the $K^+\pi^-\mu^+\mu^-$ system is constrained to the known B^0 mass.

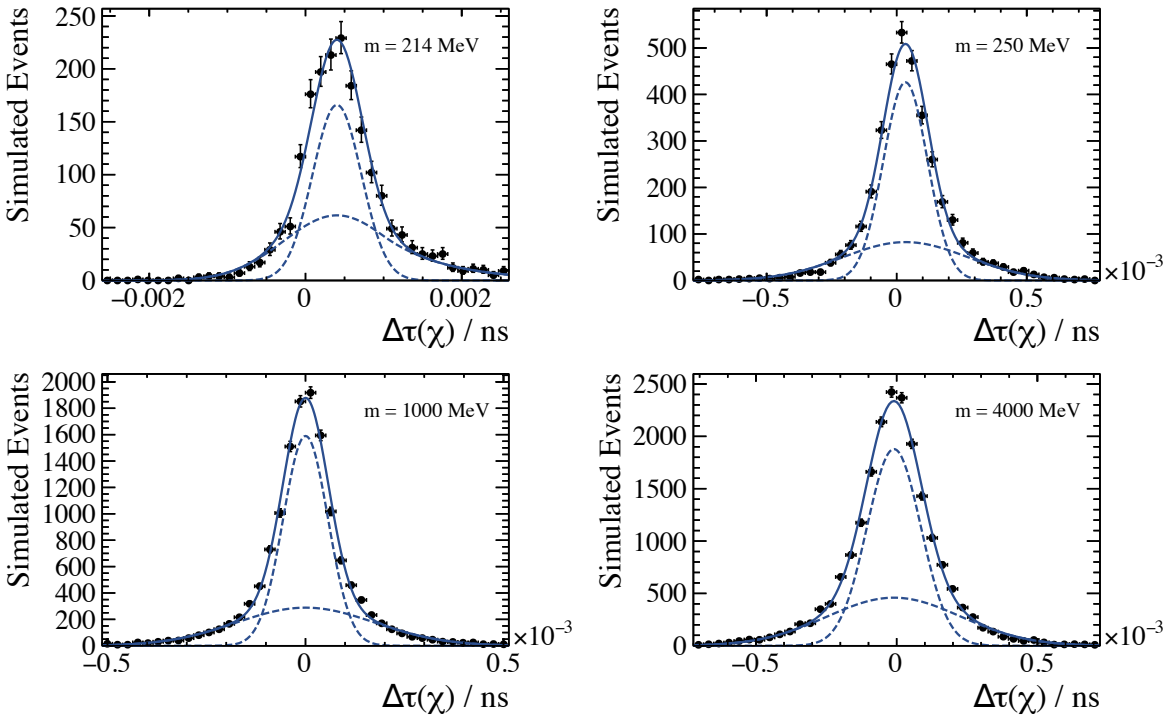


Figure 7.12: Fits to the lifetime resolution parameter, $\Delta\tau$, for individual mass samples. Each fit for $m \geqslant 250\,\text{MeV}$ is made using a double Gaussian function, and for $m < 250\,\text{MeV}$ the wider Gaussian has an exponential tail on the right-hand side. The solid line shows the total fit, and the dashed lines indicate the two components.

7.4.2 Lifetime resolution of the χ candidate

Similar to finding the mass resolution as a function of mass, the evolution of the lifetime distribution is obtained by extracting the $3\sigma_\tau$ limits from fitted distributions at known χ mass points. Linear splines are then used to interpolate to all values of m_χ, cubic splines led to massive deviations from the obvious trend, so linear interpolation is used, but for the majority of m_{DM}, the lifetime resolution is flat and this approximation has negligible effect. For each $B^0 \to K^{*0}\chi$ sample with $m \leqslant 250\,\text{MeV}$, $\Delta\tau = \tau_\chi^{\text{meas}} - \tau_\chi^{\text{true}}$ distribution is fit to a double Gaussian function. In samples with $m_\chi < 250\,\text{MeV}$ the $\Delta\tau$ distribution is observed to be significantly distorted from a simple double Gaussian, as shown in Fig. 7.12.

The distortion from a double Gaussian is explained by the narrow opening angle, θ_{open}, between muons for low mass dark bosons. When a χ is produced near the dimuon mass threshold, the daughter particles are produced nearly at rest in the frame of the χ. Therefore, the two muons have a very small opening angle in the lab frame and so the separation of muon hits in the VELO are comparable to the resolution of the VELO strips. This leads to poor spatial resolution of the χ decay vertex. By the time that the two muons are resolved from one another the measured vertex is further downstream than it really is. This effect causes a tail to extend upwards in the $\Delta\tau$ distributions for low m_χ.

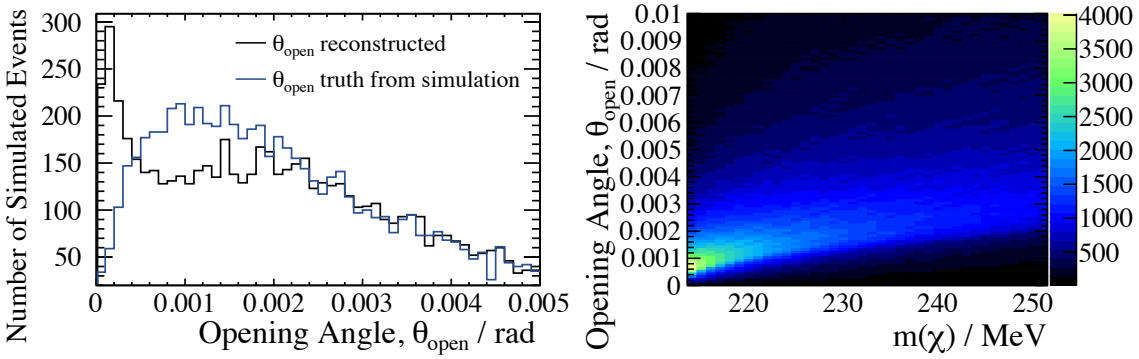


Figure 7.13: The opening angle between the two muons decaying from the χ is shown on the left for simulated events with $m_\chi = 214\,\text{MeV}$. A comparison between the reconstructed opening angle and the true opening shows that the greatest difference between the two is below $\theta_{\text{open}} = 0.002$. On the right, generator level distributions of opening angles for a range of masses, show that most χ decays have $\theta_{\text{open}} > 0.002$ for $m_\chi \gtrsim 235\,\text{MeV}$. Therefore, for higher masses, this bias does not affect the opening angle distributions.

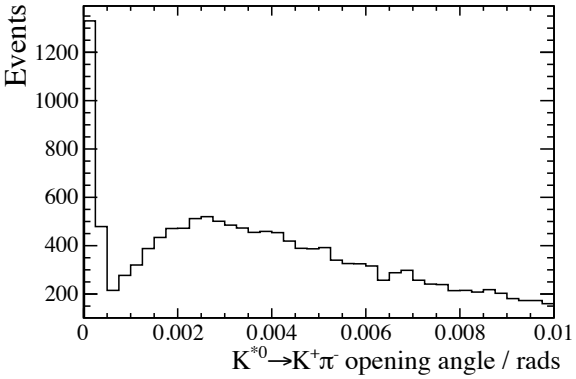


Figure 7.14: Opening angle of the $K^+\pi^-$ system in $B^0 \to K^+\pi^-\mu^+\mu^-$ taken from data, where the invariant mass of the $K^+\pi^-$ system is close to threshold.

The effect of a small opening angle can be seen by comparing the measured and true opening angle distributions for simulated decays of $B^0 \to K^{*0}\chi$, where $m_\chi = 214\,\text{MeV}$. Figure 7.13 shows a significant discrepancy between the true and measured opening angle distributions for $\theta_{\text{open}} \lesssim 0.002\,\text{rad}$. In the same figure, the evolution of θ_{open} with m_χ at generator level is shown. It can be seen that when $m_\chi = 250\,\text{MeV}$, the opening angle is predominantly larger than $0.002\,\text{rad}$.

The discrepancy between real and true opening angle can be verified using data from a decay channel with very high statistics, such as the opening angle of the $K^+\pi^-$ system in the decay $B^0 \to K^+\pi^-\mu^+\mu^-$ for for $m_{K^+\pi^-} < 640\,\text{MeV}$. Figure 7.14 shows that the opening angle of the $K^+\pi^-$ system is very low, and a similar peak at $\theta_{\text{open}} = 0$ when $m_{K^+\pi^-}$ is near threshold, $(m_{K^+} + m_{\pi^-} = 633.3\,\text{MeV})$.

Applying a cut to remove events where θ_{open} is small would be very inefficient for a low

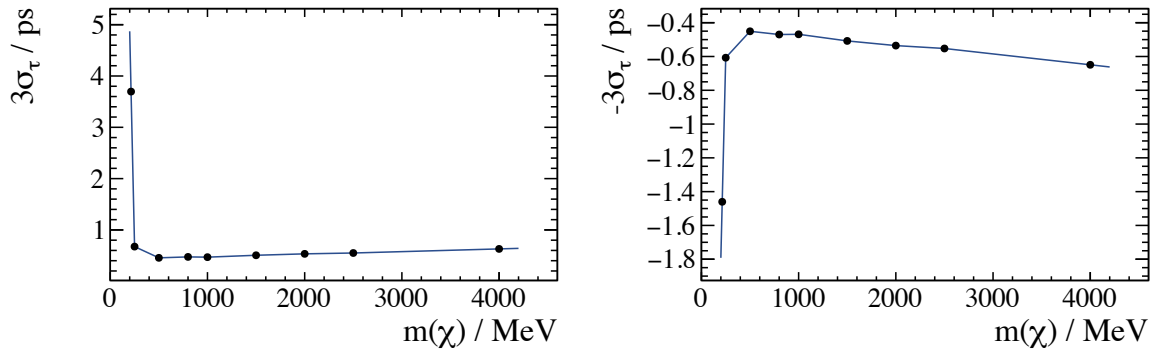


Figure 7.15: Resolutions for the lifetime of a particle as a function of mass in the (upper left) positive and (upper right) negative directions at the $3\,\sigma$ levels, used to define prompt and displaced regions. The black points are from fitted values and the blue lines show spline interpolation.

mass dark bosons, and therefore the function used to fit $\Delta\tau$ is modified to account for the positive skew. For simulated samples generated with $m_\chi < 250\,\text{MeV}$ the mean of the double Gaussian is allowed to float away from zero, and the wider of the two Gaussian functions is modified to incorporate an exponential tail extending to high $\Delta\tau$. Accounting for this effect allows the $\pm 3\sigma_\tau$ values to be taken directly from the fits in Fig. 7.12, and takes into account the bias in decay time resolution.

The prompt region is defined by $-3\sigma_\tau < \tau_\chi < 3\sigma_\tau$ and the displaced region by $\tau_\chi > 3\sigma_\tau$. Figure 7.15 shows how σ_τ varies with mass, where each point comes from the fits to $\Delta\tau$ described above, and linear spline interpolation is used to access σ_τ at all values of m_t.

7.4.3 Parameterizing the efficiency of the χ selection

Setting limits as a function of mass and lifetime requires knowledge of the efficiency for any arbitrary dark boson, as outlined in Eq. 7.13. To do this, the lifetime distribution for a given mass is fit to a simple function:

$$\mathcal{T}(\tau) = f\mathcal{G}(\mu = 0, \sigma) + (1 - f)\exp(-\alpha\tau), \tag{7.15}$$

which depends upon only a few parameters: a f, σ, and α. There is excellent agreement between τ_χ distributions from simulated samples of $B^0 \to K^{*0}\chi$ and this simple parameterisation, as shown in Fig. 7.16.

The parameters σ and α evolve smoothly as a function of mass. This allows spline interpolation to determine their values for arbitrary masses. Linear regression is used to evolve f, because of the large errors on the values of f yielded by these fits. So, σ, α, and f, become functions of mass, therefore $\mathcal{T}$ becomes a function of mass and lifetime: $\mathcal{T}(m, \tau)$. Figure 7.17 shows how the aforementioned parameters evolve as a function of

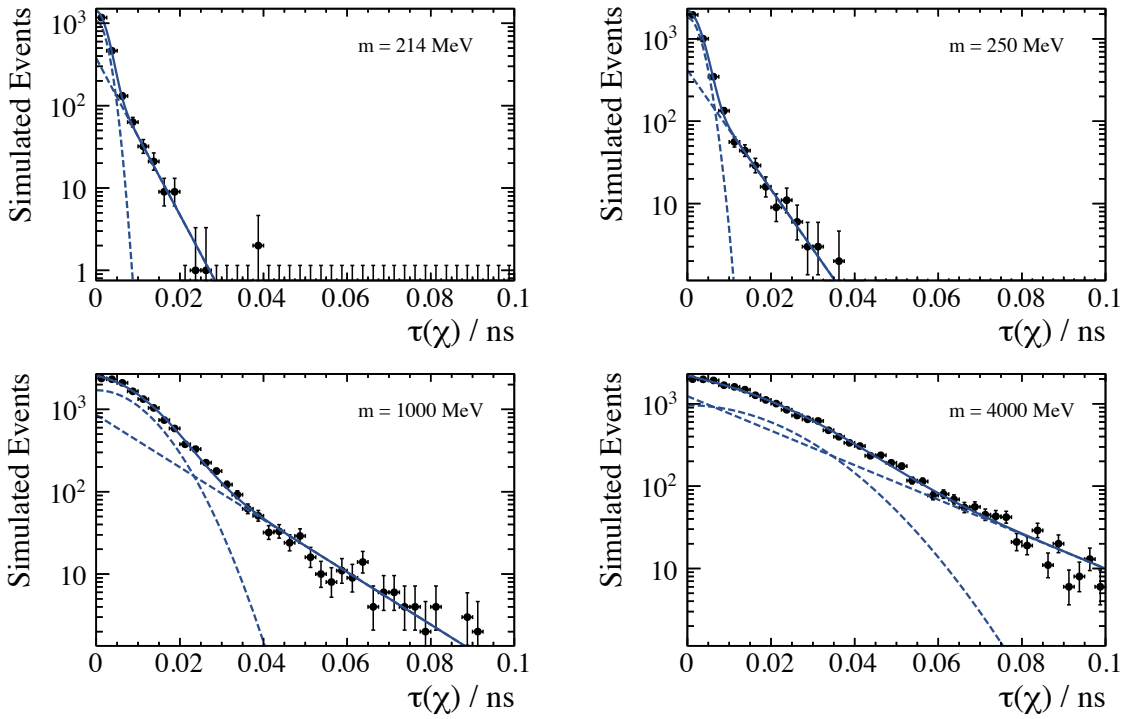


Figure 7.16: Fits, using the function in Eq. 7.15, to the decay time distributions of simulated events produced with the masses: 214, 250, 1000, and 4000 MeV (as indicated) each with $\tau_\chi = 100\,\text{ps}$. The black points are the simulated events, the solid line is the total fit, with dashed lines indicating the Gaussian and exponential components.

mass.

Once the two-dimensional map of $\mathcal{T}(m,\tau)$ has been produced, it is scaled such that for all values of m,

$$\mathcal{T}(m,\tau = 0\,\text{ns}) = \frac{\varepsilon_{\tau=0}(B^0 \to K^{*0}\mu^+\mu^-)}{\varepsilon_{1.1<q^2<6.0}(B^0 \to K^{*0}\mu^+\mu^-)}\mathcal{T}(m,\tau). \tag{7.16}$$

Here, the numerator is q^2 dependent and is calculated using simulated events, and the denominator is the total efficiency for candidates in the indicated q^2 region, which is calculated to be 3.93 %. This efficiency ratio is approximately equal to one for all values of m_χ. Figure 7.18 shows both the efficiency for a prompt dimuon pair — or equivalently a χ — as a function of q^2, and the rescaled distribution of $\mathcal{T}(m,\tau)$, which shall be called $\mathcal{R}(m,\tau)$.

In summary, $\mathcal{R}(m,\tau)$ is the lifetime distribution for a dark boson where $\tau_\chi = 100\,\text{ps}$, scaled such that at $\tau = 0$ the distribution is equal to $\varepsilon^{\tau=0}_{B^0\to K^{*0}\mu^+\mu^-}/\varepsilon^{1.1<q^2<6.0}_{B^0\to K^{*0}\mu^+\mu^-}$. To assess limits for all values of mass and lifetime initial lifetime distribution must be taken into account by convolving $\mathcal{R}(m,\tau)$ with a decaying exponential to derive an efficiency map:

$$\varepsilon(m,\tau) = \frac{1}{\tau}\int_0^{100\,\text{ps}} \mathcal{R}(m,\tau')\exp\left(-\frac{\tau'}{\tau}\right)\mathrm{d}\tau'. \tag{7.17}$$

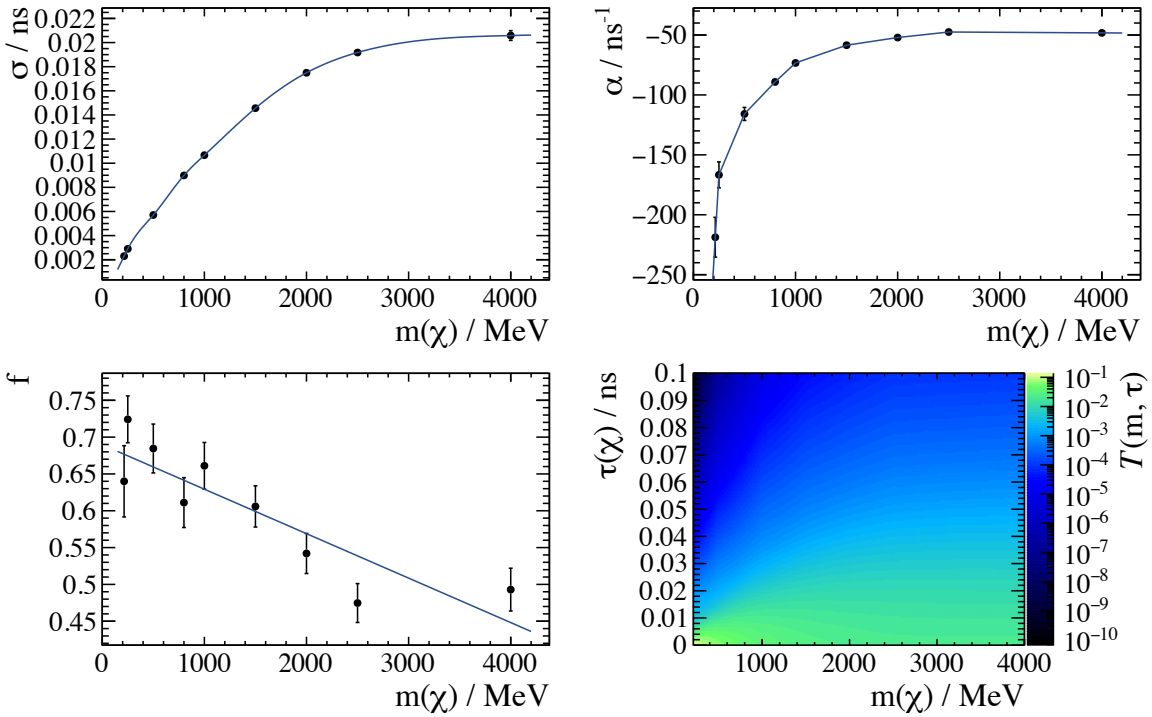


Figure 7.17: Parameters from Eq. 7.15 as a function of mass from simulated events. The parameters are (upper left) σ, (upper right) α, (lower left) f, and (lower right) the two-dimensional projection of the total parameterisation. Splines are used to parameterise each shape, except for the parameter f, where a linear fit is used.

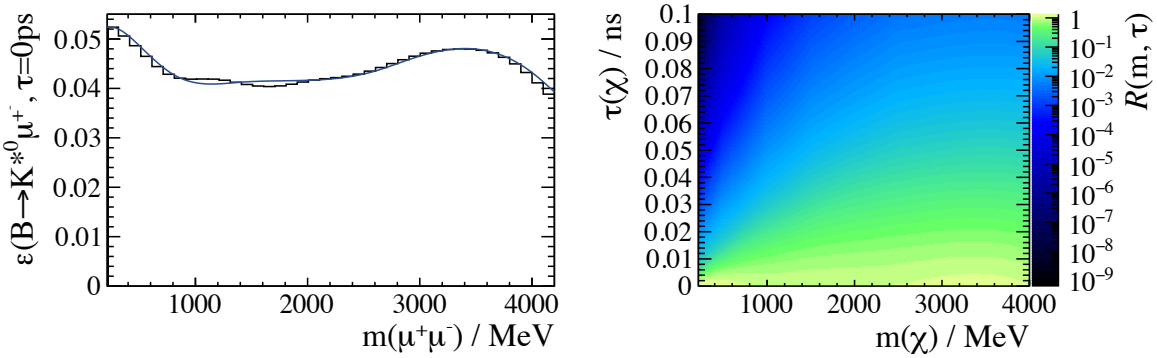


Figure 7.18: Efficiency $B^0 \to K^{*0}\mu^+\mu^-$ for prompt muons is shown on the left, in bins, and a spline is used to parameterise the shape, the deviations between 1000 and 2000 MeV are too small to have any significant impact on the limits. This is divided by the efficiency in the normalisation region, and the $\tau = 0$ lifetime bin from Fig. 7.17 is set to this value for all masses; the decay time dimension is scaled.

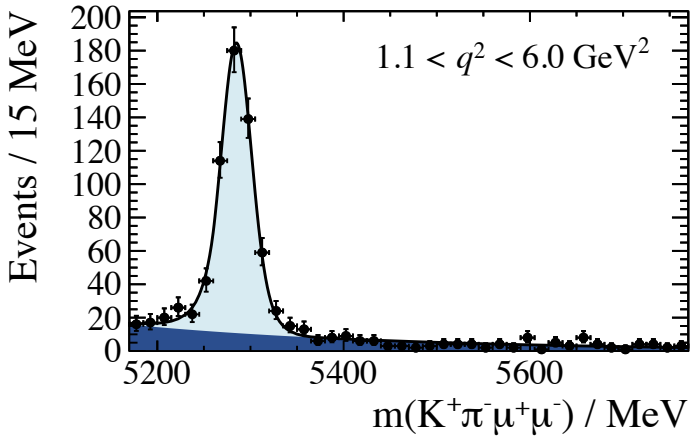


Figure 7.19: Fit to the invariant mass spectrum of the B^0 candidates in selected data in the range $1.1 < q^2 < 6.0\,\text{GeV}^2$. The signal model is the sum of two Gaussian functions with power-law tails on the low-mass side with parameters taken from the analysis described in Ref. [76],the background model is a decaying exponential. This fit results in a signal yield of (527 ± 26) compared to approximately 625 in the SM analysis.

The upper lifetime acceptance is chosen to be 100 ps, because the efficiency at longer lifetimes, for all values of m_χ, is very poor.

7.5 Unblinding results

This search for a new dark boson is done blindly to ensure that no bias is introduced in the course of the analysis. The data is unblinded in stages to ensure that the selection is behaving as expected on real data.

Selected $B^0 \to J/\psi K^{*0}$ events are used to check that the selection was not biased based on neither the year, nor the polarity of the LHCb magnet. No bias is observed; in fact efficiencies for the uBDT are observed to be consistent to the 10^{-4} level in all four regions.

The unblinding procedure begins by checking the yield of the normalisation channel $B^0 \to K^{*0}\mu^+\mu^-$ in the range $1.1 < q^2 < 6.0\,\text{GeV}^2$ and comparing it with the yield from Ref. [76], which details an angular analysis of this decay. The yield is taken from an unbinned fit to selected prompt $B^0 \to K^{*0}\chi$ candidates using a mass model of two Gaussian functions sharing the same mean, and with a power-law tail on the low mass side. A simple exponential models the background component. This yields 527 B^0 candidates, which can be compared to approximately 625 events in the SM selection, where the drop in signal is expected given the search is for a rare process. Together with the drop in signal, comes a drop in background yield, from ∼630 background events over the full mass range, to only 290. Figure 7.19 shows the B^0 candidate mass spectrum for the normalisation channel, and the fitted distribution overlaid.

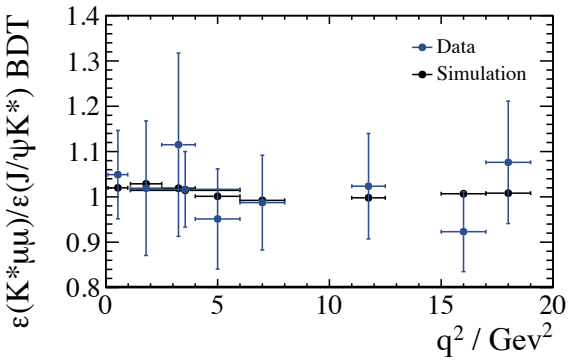


Figure 7.20: Ratio of the efficiencies of the BDT selection for a range of q^2 bins, as used in the SM $B^0 \to K^{*0}\mu^+\mu^-$ analysis, with respect to the efficiency for $B^0 \to J/\psi K^{*0}$. Data and simulation are shown to be in good agreement.

After unblinding the region $1.1 < q^2 < 6.0\,\text{GeV}^2$, other prompt q^2 regions were also used to confirm that the ratio of BDT selection efficiencies $\varepsilon_{\text{BDT}}^{\text{q}^2\text{bin}}(B^0 \to K^{*0}\mu^+\mu^-)/\varepsilon_{\text{BDT}}(B^0 \to J/\psi K^{*0})$ are approximately the same in data and simulation. To determine these efficiencies, the full selection without the BDT is taken, and a fit performed and the signal yield is extracted. The same procedure is followed with the BDT applied. Figure 7.20 shows the ratio of these yields in different regions of q^2. It is observed the distributions for data and simulation are in good agreement, centred around unity with about 5–10 % precision.

Using the unblinded distributions it is possible to estimate the amount of combinatorial background that will remain in the final selection. Fitting an exponential to model the background across the signal region allows an approximate number of events in the combinatorial background to be deduced. This number of events can be taken from the upper-mass sideband, and the invariant mass of the dimuon pair can be plotted. With this method, it expected that a maximum of 10(2) events will contribute to the background in the prompt(displaced) region at a single test mass, but on average the value is 1.8(0.2) events per bin. Figure 7.21 shows the shape and scale of the combinatorial background, and the fits used to calculate it.

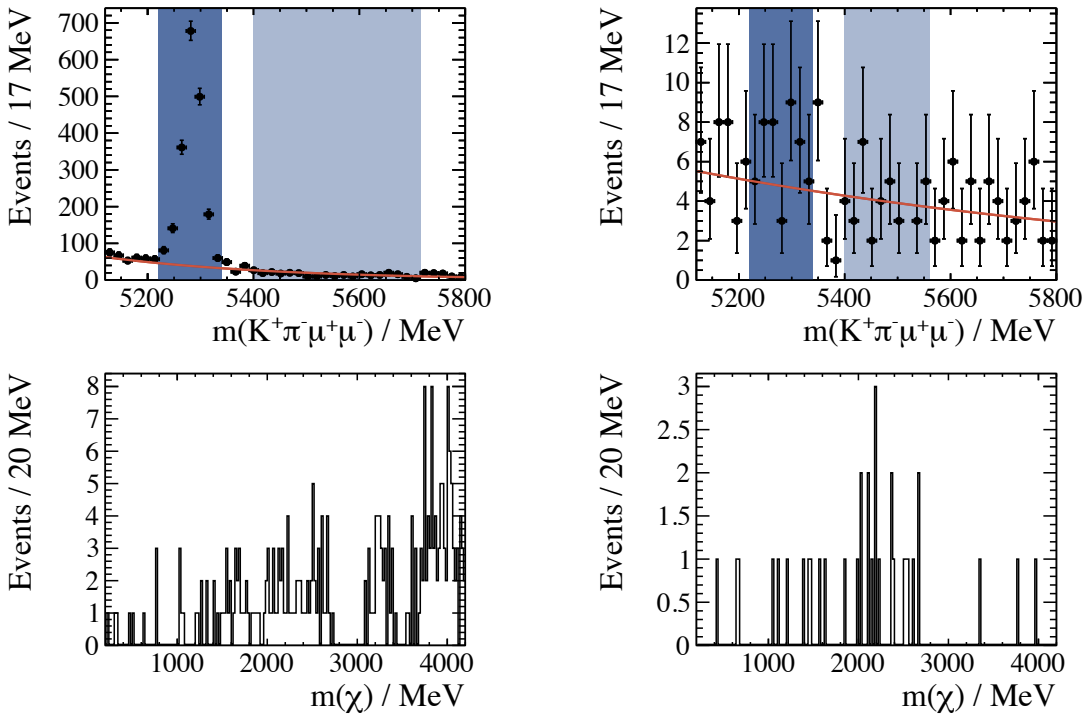


Figure 7.21: Fits to the background component of selected $B^0 \to K^{*0}\mu^+\mu^-$ events are shown in the upper plots. The the dark blue signal regions are not used in the fit, but the integral of the background exponential function is used to determine the combinatorial background contribution. The light blue region in the upper tail contains the same number of events as in the signal region. Since the source of combinatorial background varies little, the events in the high-mass (light blue) region will have a similar mass distribution to those in the signal region. These mass distributions are shown in the lower plots, giving an idea of the contribution from combinatorial background at each test mass. A bin width of 20 MeV is comparable to the width of the sigal window in a single test for most of values of $m_{\mu^+\mu^-}$.

7.5.1 Calculation of the *p*-value

The unblinded distribution in $m_{\mu^+\mu^-}$ is shown in Fig. 7.22. Using the statistical method described in Sec. 7.2 the following ranges in m_t are scanned:

$$\begin{aligned} 253.4 < m_t < \ \ 369.5\,\text{MeV}, \\ 574.5 < m_t < \ \ 906.5\,\text{MeV}, \\ 1136.0 < m_t < 2813.0\,\text{MeV}, \\ 3246.5 < m_t < 3576.5\,\text{MeV}, \\ 3796.0 < m_t < 4356.0\,\text{MeV}. \end{aligned}$$

Values of m_t do not reach threshold boundaries because of the sidebands extending in either direction. The minimum local p-value is found to be 3.6×10^{-3} at $m_t^{\text{min}} = 4285.0\,\text{MeV}$.

The look-elsewhere effect must be considered. To convert the p-value to a global one, a PDF is fit to the $m_{\mu^+\mu^-}$ distribution of the prompt and displaced $B^0 \to K^{*0}\chi$ candidates that lie outside of the vetoed regions, and outside of the signal region centred at m_t^{min}. A fourth(second) order Chebychev polynomial is fit to candidates in the prompt(displaced) region. From these PDFs 1.5×10^7 toy datasets are generated, and the minimum p-value for each one is calculated. Constructing a cumulative distribution of these p-values makes an easy conversion from local to global p-values, this conversion is shown in Fig. 7.22. Shown alongside the cumulative histogram is the asymptotic approximation, which is seen to be in excellent agreement for local p-values less than about 10^{-4}. The histogram converts the local p-value to a global p-value of 0.63, equivalent to a shift in significance of $2.9\,\sigma$ to $0.48\,\sigma$. These results show no evidence for a new dark boson in the mass ranges given above.

7.5.2 Systematic uncertainties

Systematic uncertainties must be assessed in order to set limits. Sources of systematic uncertainties are from: the ratio of efficiencies $\varepsilon(B^0 \to K^{*0}\chi)/\varepsilon(B^0 \to K^{*0}\mu^+\mu^-)$; the of prompt and displaced regions as defined by the lifetime resolution; and the uncertainty on the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction in the range $1.0 < q^2 < 6.0\,\text{GeV}^2$. These systematic uncertainties do not affect the significance of the minimum local p-value.

In the prompt region the efficiency ratio $\varepsilon_{\text{B}^0 \to \text{K}^{*0}\chi(\mu^+\mu^-)}/\varepsilon_{\text{B}^0 \to \text{K}^{*0}\mu^+\mu^-}$ is constructed to be close to unity. The selection should not change this, and the ratio should remain the same for all lifetimes. This validated by comparing $B^0 \to J/\psi\, K^0_{\text{S}}$ in data and simulation. The full selection is applied using the K^0_{S} as a proxy for the χ, and using the variable $\mathcal{P}_\pi$ of the pions of the K^0_{S} decay in place of the $\mathcal{P}_\mu$. Mass cuts are also applied to isolated

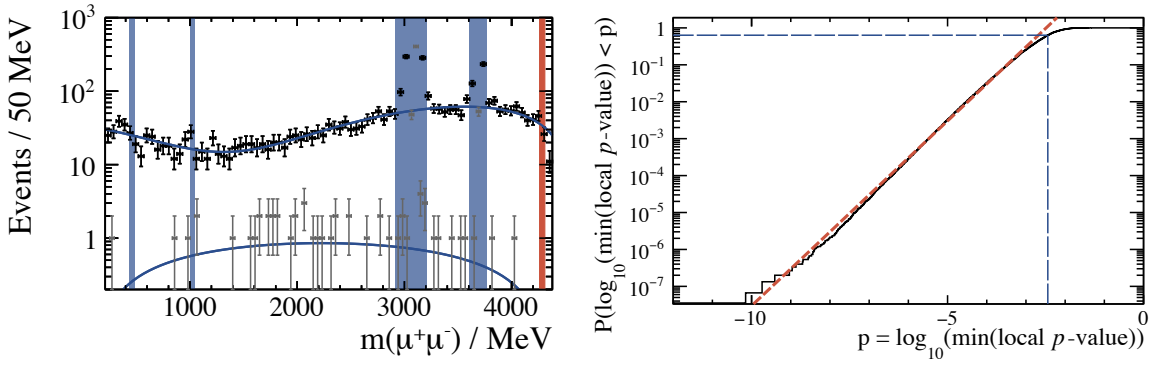


Figure 7.22: Invariant mass distributions of the dimuon pair are shown on the left, while prompt and displaced candidates are black and grey points, respectively. Chebychev polynomials fitted to the data are shown as solid blue lines. The narrow red region indicates $|m_t - 4285.0| < 5\sigma_m$, which encompasses the test mass with the minimum p-value, and everything withing the sideband regions. On the right is the result of running the method on 1.5×10^{-7} toy datasets, used to convert the local p-value (3.6×10^{-3}) to a global one (0.63). This conversion is indicated by the blue dashed line. The result of the asymptotic formula is shown by the red dashed line, and is seen to be an excellent approximation of the true distribution.

the J/ψ and K^0_S candidates, and a cut of $\tau_{K^0_S} > 0.2\,\text{ps}$ is applied to avoid contamination from prompt decays such as $B^0 \to J/\psi\,\pi^+\pi^-$. Lifetime distributions for these selected K^0_S candidates in data and simulation are shown in Fig. 7.23. These distributions would be approximately uniform without detector effects; the degree to which the distributions agree demonstrates that the simulation models the detector well and results in negligible systematic uncertainty.

Throughout the search it is assumed that the signal window, which extends $\pm 2\sigma_m$ around the test mass, contains 95% of a total signal from a decaying χ. A poorly modelled mass resolution will lead to an incorrect efficiency. The accuracy of the mass resolution is tested using $B^0 \to J/\psi\,K^{*0}$ candidates falling within 60 MeV of the nominal B^0 mass. In actuality, it is observed that 94% of $J/\psi \to \mu^+\mu^-$ candidates fall within $2\sigma_m$ of the known J/ψ mass.

The simulated events of the SM decay $B^0 \to K^{*0}\mu^+\mu^-$ were generated assuming knowledge of the true Wilson coefficients that contribute to the decay. Incorrect values of the Wilson coefficients would lead to incorrect efficiency calculations. However, results of the angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from LHCb, as described in Ref. [76], show that the Wilson coefficients used in the simulation are consistent with those observed. A systematic uncertainty of approximately is applied to account for possible mismodelling of the SM simulation. An ensemble of toy datasets are generated, each of which having input Wilson coefficients that are randomly distributed about their means within their uncertainties. The RMS deviation from the nominal value of this ensemble is taken as the systematic uncertainty. It is approximately 1 %.

The mismodelling of the lifetime resolution is assessed using events consistent with the decay

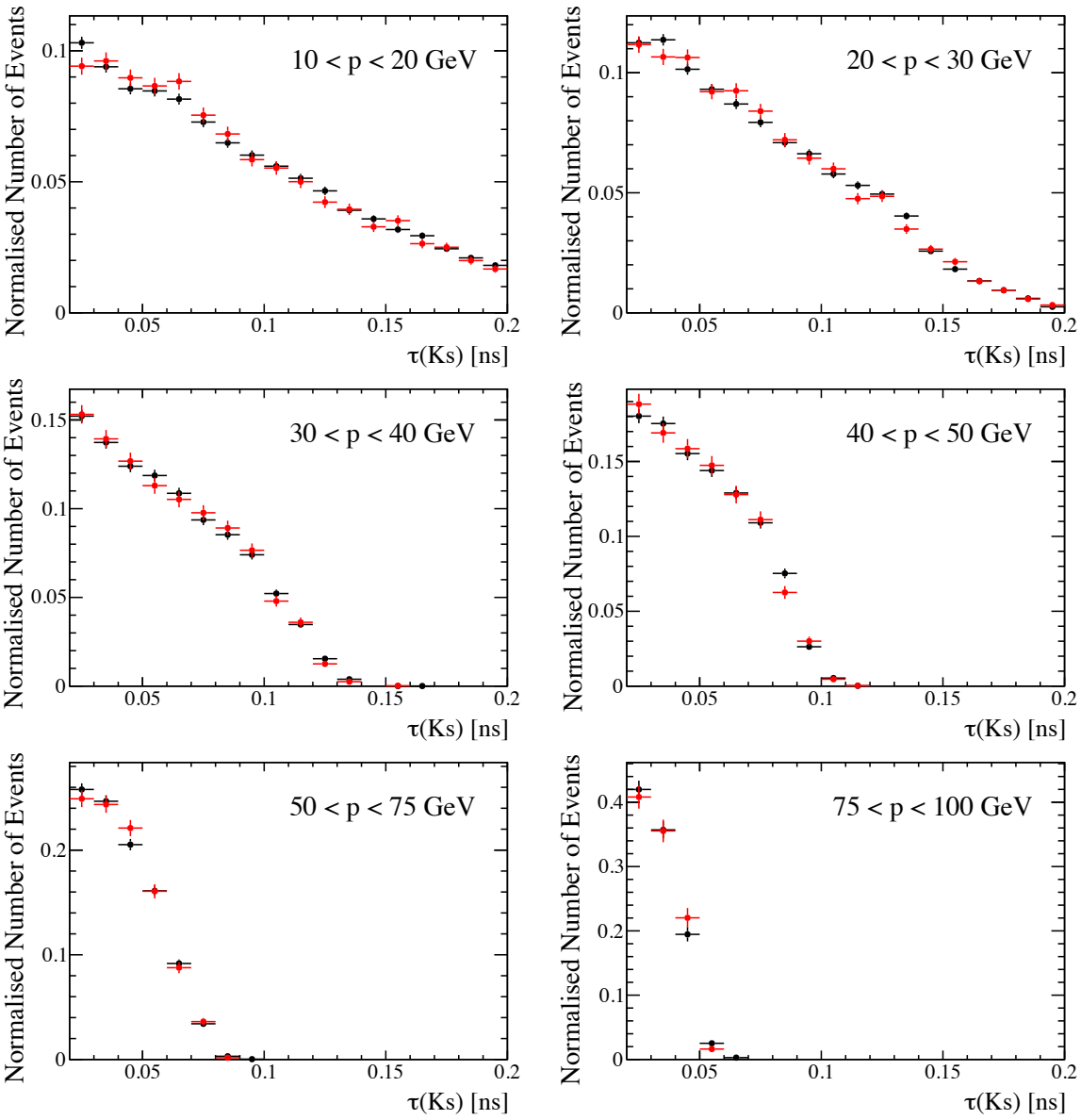


Figure 7.23: Reconstructed K^0_S lifetime distributions in bins of K^0_S momentum for (black) data and (red) simulation, where the K^0_S is part of the decay $B^0 \to J/\psi K^0_S(\to \pi^+\pi^-)$ and is a proxy χ candidate. Each plot is normalised to unit area. Good agreement is seen between the data and simulation.

$B^0 \to J/\psi K^{*0}$ in data. Events falling in the displaced region, defined to be $\tau_{\mu^+\mu^-} > 3\sigma_\tau$, should constitute 0.13% of all events, but actually make up $(0.18 \pm 0.01)\%$. To this end, a 3% systematic uncertainty is applied, which is calculated by altering the lifetime resolution as calculated using simulation until 0.18% of events fell in the displaced region.

The statistical uncertainty propagated from the normalisation yield is 5%. Additionally, there is contamination from the S-wave component of the K^{*0}, which is estimated to be $(4 \pm 4)\%$ [109].

It has been noted that this analysis relies on the assumption that the background is smoothly varying, and can be approximated as being locally linear. This assumption clearly introduces an uncertainty, but this is already accounted for in the method by the addition of the Gaussian, $\mathcal{G}(y, x, \sigma_y)$, in the likelihood shown in Eq. 7.7. There is no need to add a systematic uncertainty for the chosen value of σ_y, because it is, itself, an uncertainty.

7.6 Summary

A selection designed to be uniformly efficient in all regions of the dark boson's mass-lifetime parameter space has been presented with a view to embarking on a search for a new dark boson. A frequentist method designed to search for an signal in any arbitrary mass spectrum has also been presented and used in the case of the dimuon mass spectrum from $B^0 \to K^{*0}\mu^+\mu^-$.

This strategy has extracted a p-value of a particle considering test masses which do not go all the way to the boundaries of various vetoes in the invariant dimuon mass spectrum. It is determined that the maximum deviation of the selected candidates from the null hypothesis of zero signal has a global significance of $0.48\,\sigma$ at $m_t = 4285.0\,\text{MeV}$. This is consistent with no new particle in the $\mu^+\mu^-$ distribution over the mass ranges probed. The full analysis will get much closer to these edges and probe the interesting $m_{\mu^+\mu^-} = 214\,\text{MeV}$ region. The next step in the analysis is to set limits and present them in a model independent way. Of course, they can be translated to specific models for interpretation.

7.6.1 The published LHCb analysis

The entire analysis described in this chapter, including strategy and selection, was completed as part of the published LHCb analysis after this thesis was submitted, and therefore results to include in this thesis were calculated separately, as given above. Since then the LHCb analysis has been published in Ref. [3]. There are a few differences between the two analysis. Firstly the LHCb analysis covered the entire range in $m_{\mu^+\mu^-}$, going very close to

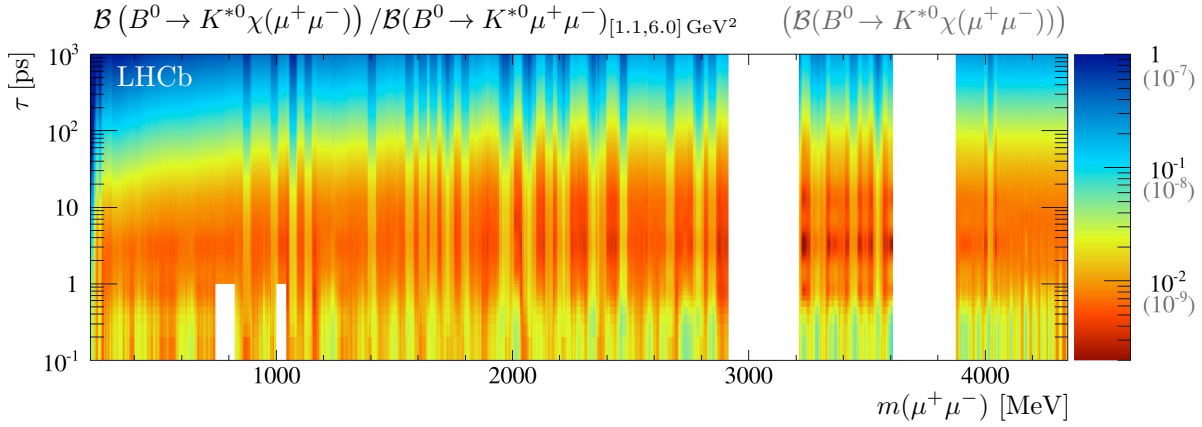


Figure 7.24: Upper limits for $\mathcal{B}\big(B^0 \to K^{*0}\chi(\mu^+\mu^-)\big)/\mathcal{B}\big(B^0 \to K^{*0}\mu^+\mu^-\big)_{[1.1,6.0]\,\mathrm{GeV}^2}$ and $\mathcal{B}\big(B^0 \to K^{*0}\chi(\mu^+\mu^-)\big)$ at 95 % CL. Vetoed areas for the J/ψ and $\psi(2S)$ can be seen for all lifetime regions, while the ω and ϕ regions are removed in only the prompt case. This figure is taken from Ref. [3].

the kinematic boundaries. The conversion of local to global p-value was made using a PDF constructed from cubic spline interpolation and generating toys in bins, rather than continuously. Finally, the decision was made in the LHCb analysis to veto the region around the ω mass, because the branching fraction $\mathcal{B}\big(B^0 \to K^{*0}\omega(\to \mu^+\mu^-)\big)$ is relatively large (as shown in Table 7.2) and therefore there is some chance of observing interference effects. The ϕ and ω vetoes also only apply in the prompt regions. The minimum local excess in the published LHCb analysis appears at $m_\chi = 253\,\mathrm{MeV}$ with a p-value of about 0.8, consistent with no new physics. This is just outside the range of the results given above. The published plane of upper limits for $\mathcal{B}\big(B^0 \to K^{*0}\chi(\mu^+\mu^-)\big)/\mathcal{B}\big(B^0 \to K^{*0}\mu^+\mu^-\big)_{[1.1,6.0]\,\mathrm{GeV}^2}$ and $\mathcal{B}\big(B^0 \to K^{*0}\chi(\mu^+\mu^-)\big)$ at 95 % CL are shown in Fig. 7.24. These limits take into account the systematic uncertainties from the previous section. A model dependent exclusion region for the inflaton is shown in Fig. 7.25 for comparison with same parameter space as shown in Fig. 7.1.

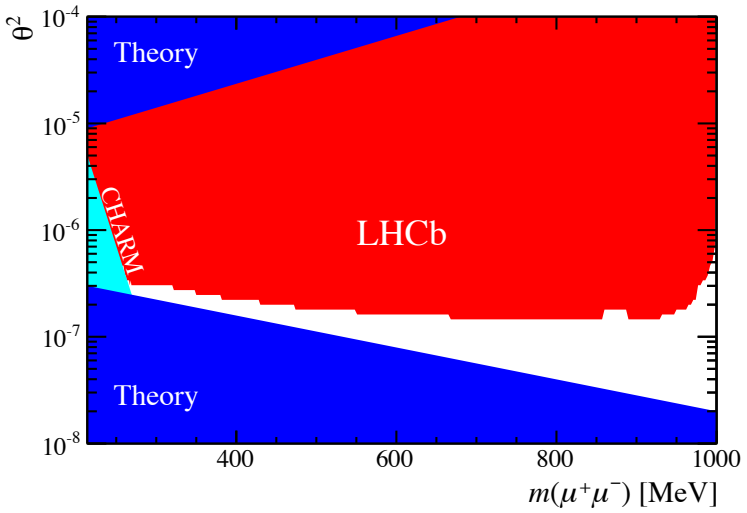


Figure 7.25: Projected exclusion regions for an inflaton model from Ref. [93], in the mass range $1 < m_\chi < 1000\,\text{MeV}$ to 95% CL, this plane is equivalent to the parameter space shown in Fig. 7.1. The regions excluded by theory [93] (red) and by the CHARM experiment [95] (cyan) are also shown.

Chapter 8

Conclusions

This thesis presents three published analyses [1–3], all undertaken using data collected by the LHCb experiment [45]. The results of two of these analyses boast first observations. Each has the objective of finding evidence of physics BSM in the decays of B mesons. The analysis techniques employed are all slightly different, most obviously Ref. [1] and Ref. [2] describe indirect searches for NP, while Ref. [3] is a direct search for a dark boson of unknown mass and lifetime.

An analysis of the decay $B^+ \to D_s^+\phi$ is presented in Chap. 5. First evidence for the decay was seen with a statistical significance of greater than 3σ, this also constitutes the first evidence for a fully hadronic decay via an annihilation-type diagram. Measurements involving the decay $B^+ \to D_s^+\phi$ are sensitive to NP effects contributing at tree level because there is only one allowed process in the SM through which it can propagate,. This single process is suppressed by the CKM matrix element V_{ub}, which is the element with the largest uncertainties in the CKM matrix. There are there also are historic tensions between measurements of V_{ub} made using inclusive an exclusive modes.

The branching fraction measurement

$$\mathcal{B}\big(B^+ \to D_s^+\phi\big) = \big(1.87\,^{+1.25}_{-0.73}(\text{stat}) \pm 0.19(\text{syst}) \pm 0.32(\text{norm})\big) \times 10^{-6}$$

is somewhat higher than SM predictions, which are of order 10^{-7} [62–65], but not incompatible considering the range of the theoretical predictions, and the large statistical uncertainties inherent in the measurement. This value of $\mathcal{B}(B^+ \to D_s^+\phi)$ sheds no light on the true value of V_{ub}. Reference [110] argues that these discrepancies cannot be explained by physics BSM, and is rather due to underestimated uncertainties in either theory or experiment.

Since the decay $B^+ \to D_s^+\phi$ is mediated by a W^+ boson, in NP scenarios another charged boson, such as a H^+ from a 2HDM, could contribute to the decay amplitude. These additional processes could alter the branching fraction significantly, which is not observed,

or introduce extra phases into the decay, causing the *CP*-asymmetry to deviate from $\mathcal{A}_{CP} = 0$, as expected in the SM. The value measured, after correcting for production and detection asymmetries, is

$$\mathcal{A}_{CP}\big(B^+ \to D_s^+\phi\big) = -\big(0.01 \pm 0.41(\text{stat}) \pm 0.03(\text{syst})\big),$$

which is consistent with SM expectations.

Chapter 6 presented an analysis with the first observations and branching fraction measurements of the decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$, as well as the differential branching fraction of $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ in bins of q^2. The integrated branching fractions of these decays are

$$\begin{aligned}\mathcal{B}\big(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-\big) &= \big(4.36\,^{+0.29}_{-0.27}(\text{stat}) \pm 0.21(\text{syst}) \pm 0.18(\text{norm})\big) \times 10^{-7},\\ \mathcal{B}\big(B^+ \to \phi K^+\mu^+\mu^-\big) &= \big(0.82\,^{+0.19}_{-0.17}(\text{stat})\,^{+0.10}_{-0.04}(\text{syst}) \pm 0.27(\text{norm})\big) \times 10^{-7},\end{aligned}$$

and both have statistical significances of greater than 5σ The decay $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ has a large branching fraction and could be used for future analyses similar to those of interest in $B^0 \to K^{*0}\mu^+\mu^-$, since both are sensitive to the same operators. An angular analysis would help constrain the scalar, pseudoscalar, and tensor amplitudes of the decay, all of which are vanishingly small in the SM. This is made difficult by the number of contributing states to the $K^+\pi^+\pi^-$ system, but with more statistics it will be possible to gain a better understanding of strange states that decay into kaons and pions.

Interpreting precise measurements of B physics observables that can be made experimentally are often made difficult by the form-factor parameterisation which must be adopted. It is these QCD effects, particularly the form-factors, that are the dominant sources of theoretical uncertainty. Difficulties in dealing with QCD has been demonstrated by a lack of consistent predictions for $\mathcal{B}(B^+ \to D_s^+\phi)$, and the absence of any predictions for either $\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)$ and $\mathcal{B}(B^+ \to \phi K^+\mu^+\mu^-)$. Techniques such as lattice QCD [111] have recently been employed in this are with encouraging results [112]. This bodes well for future measurements of this kind.

Given larger statistics, the additional channels $B^+ \to K^+K^-\pi^+\mu^+\mu^-$ and $B^+ \to \pi^+\pi^+\pi^-\mu^+\mu^-$ may be observable, and give access to the ratio of CKM matrix elements V_{td}/V_{ts}. These would be complimentary to the current measurements from B-meson oscillations.

The large uncertainties in the measurement of $\mathcal{B}(B^+ \to \phi K^+\mu^+\mu^-)$ are primarily due to uncertainties propagated from the branching fraction of the normalization channel $B^+ \to J/\psi\,\phi K^+$. The paper in Ref. [2] quotes the ratio of branching fractions in order for $\mathcal{B}(B^+ \to \phi K^+\mu^+\mu^-)$ to be calculated given an improved measurement.

Finally, a direct search for a NP particle, χ, belonging to some dark sector is presented. Using $B^0 \to K^{*0}\mu^+\mu^-$ candidates the dimuon invariant mass spectrum is searched for a signal indicative of a dark boson decaying via $\chi \to \mu^+\mu^-$. The selection is designed specifically not to bias any corner of the mass-lifetime space that the χ might inhabit, this is done primarily with the aid of the uBoost algorithm. Efficiencies and resolutions are parameterised using discrete simulated samples of $B^0 \to K^{*0}\chi$ and spline interpolation is used to understand selection and resolution effects for any value of m_χ. A novel frequentist strategy was employed to perform the search. The strategy involved a scan in regions of dimuon invariant mass. For each region, the p-value for a signal excess was calculated. Once the look elsewhere effect was accounted for the minimum local p-value is 0.48σ at $m_{\mu^+\mu^-} = 4285.0\,\text{MeV}$. This is consistent with with no new particle observed. Further studies will push tested masses to the boundaries of vetoed regions. The projected limits for an inflaton model indicate that much of the allowed parameter space will be excluded for $m_\chi < 1000\,\text{MeV}$. A similar approach could be used in any arbitrary mass spectrum to search for a multitude of particles appearing above a smoothly varying background.

In conclusion, the SM continues in its resilience. The vast majority of experimental HEP measurements appear to be in agreement with the SM to the limit of their accuracy, with the notable exceptions of the neutrino sector. There is a complementarity that exists between indirect and direct measurements. Historically, it has often been the case that indications of future discoveries were first anticipated by observations made by indirect experiments. This pattern looks set to endure, as precision measurements in the flavour sector continue to play an important role in searching for NP. As of Run 1 of the LHC, there has been no clear indication of where NP may lie. There are no results indicating NP with a significance of 5σ. However, there are many flavour physics observables with significances greater than 3σ [76, 78, 113–117] which makes the area very interesting.

It is not unreasonable to expect — or at least hope — that NP lives just around the corner. With the factor of two increase in statistics expected in LHCb's total dataset in Run 2 of the LHC, it will be very interesting to see how the flavour sector observables that differ from the SM expectation change. It is clear that the important interplay between indirect and direct searches in the arena of HEP is set to continue.

Glossary

2HDM	Two Higgs Doublet Model
BAU	Baryon Asymmetry of the Universe
BBDT	Bonsai BDT
BDT	Boosted Decision Tree
BSM	Beyond the Standard Model
C	Charge conjugation
CC	Charged Current
CKM	Cabibbo-Kobayashi-Maskawa
CP	Charge-Parity
CPV	*CP* Violation
DLL	Delta Log Likelihood
DM	Dark Matter
DOCA	Distance of Closest Approach
DT	Decision Tree
ECAL	Electromagnetic Calorimeter
EFT	Effective Field Theory
FCNC	Flavour Changing Neutral Current
HCAL	Hadron Calorimeter
HEP	High Energy Physics
HLT	High Level Trigger
HPD	Hybrid Photomultiplier Detector
HQET	Heavy Quark Effective Theory
IP	Impact Parameter
IT	Inner Tracker
L0	Level One Trigger
LHC	Large Hadron Collider
LHCb	LHC beauty
MVA	Multivariate Analysis
NC	Neutral Current
NP	New Physics
OPE	Operator Product Expansion

OT	Outer Tracker
P	Parity
PDF	Probability Density Function
PID	Particle Identification
PQ	Peccei-Quinn
PS	Preshower detector
PV	Primary Vertex
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
RICH	Ring Imaging Cherenkov
RPV	R-Parity Violation
SM	Standard Model
SPD	Scintillating Pad Detector
SPS	Super Proton Synchrotron
SSB	Spontaneous Symmetry Breaking
ST	Silicon Tracker
SUSY	Supersymmetry
T	Time reversal
`TIS`	Trigger Independent of Signal
TMVA	Toolkit for Multivariate Analysis
`TOS`	Trigger On Signal
TT	Tracker Turicensis
uBDT	Uniform BDT
UT	Unitarity Triangle
VELO	Vertex Locator
VEV	Vacuum Expectation Value

Bibliography

[1] LHCb collaboration, R. Aaij *et al.*, *First evidence for the annihilation decay mode* $B^+ \to D_s^+\phi$, JHEP **02** (2013) 043, arXiv:1210.1089.

[2] LHCb collaboration, R. Aaij *et al.*, *First observations of the rare decays* $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ *and* $B^+ \to \phi K^+\mu^+\mu^-$, JHEP **10** (2014) 064, arXiv:1408.1137.

[3] LHCb collaboration, R. Aaij *et al.*, *Search for hidden-sector bosons in* $B^0 \to K^*\chi(\to \mu^+\mu^-)$ *decays*, arXiv:1508.04094, submitted to PRL.

[4] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[5] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[6] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367.

[7] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[8] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156.

[9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585.

[10] T. W. B. Kibble, *Symmetry breaking in nonAbelian gauge theories*, Phys. Rev. **155** (1967) 1554.

[11] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press, 2003.

[12] F. Halzen and A. D. Martin, *Quarks and leptons: an introductory course in modern particle physics*, Wiley, 1984.

[13] T. Mannel, *Effective Field Theories in Flavour Physics*, Springer Tracts in Modern Physics, Springer Berlin Heidelberg, 2004.

[14] Particle Data Group, K. A. Olive *et al.*, *Review of particle physics*, Chin. Phys. **C28** (2014) 090001.

[15] E. Noether, *Invariante Variationsprobleme*, Nachr. d. König. Gesellsch. d. Wiss. zu Göttingen, Math-phys. Klasse, Seite 235-157 (1918) arXiv:www.physics.ucla.edu/~cwp/articles/noether.trans/german/emmy235.html.

[16] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2 pi Decay of the k(2)0 Meson*, Phys. Rev. Lett. **13** (1964) 138.

[17] CMS Collaboration, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[18] ATLAS Collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[19] CMS Collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, arXiv:1207.7235.

[20] ATLAS Collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, arXiv:1207.7214.

[21] Particle Data Group, J. Beringer *et al.*, *Review of particle physics*, Phys. Rev. **D86** (2012) 010001.

[22] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, *Tenth-Order QED Contribution to the Electron g-2 and an Improved Value of the Fine Structure Constant*, Phys. Rev. Lett. **109** (2012) 111807, arXiv:1205.5368.

[23] A. D. Sakharov, *SPECIAL ISSUE: Violation of CP in variance, C asymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34** (1991) 392.

[24] J. M. Cline, *Baryogenesis*, arXiv:hep-ph/0609145.

[25] P. Huet and E. Sather, *Electroweak baryogenesis and standard model CP violation*, Phys. Rev. **D51** (1995) 379, arXiv:hep-ph/9404302.

[26] Planck, R. Adam *et al.*, *Planck 2015 results. I. Overview of products and scientific results*, arXiv:1502.01582.

[27] V. C. Rubin and W. K. Ford, Jr. *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, Astrophys. J. **159** (1970) 379.

[28] V. C. Rubin, W. K. J. Ford, and N. . Thonnard, *Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/*, Astrophys. J. **238** (1980) 471.

[29] M. Markevitch *et al.*, *Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56*, Astrophys. J. **606** (2004) 819, arXiv:astro-ph/0309303.

[30] J. A. Tyson, R. A. Wenk, and F. Valdes, *Detection of systematic gravitational lens galaxy image alignments — Mapping dark matter in galaxy clusters*, ApJL **349** (1990) L1.

[31] HyperCP Collaboration, R. A. Burnstein *et al.*, *HyperCP: A High-rate spectrometer for the study of charged hyperon and kaon decays*, Nucl. Instrum. Meth. **A541** (2005) 516, arXiv:hep-ex/0405034.

[32] HyperCP Collaboration, H. Park *et al.*, *Evidence for the decay* $\Sigma^+ \to p\mu^+\mu^-$, Phys. Rev. Lett. **94** (2005) 021801, arXiv:hep-ex/0501014.

[33] J. Charles *et al.*, *Current status of the Standard Model CKM fit and constraints on* $\Delta F = 2$ *New Physics*, Phys. Rev. **D91** (2015), no. 7 073007, arXiv:1501.05013.

[34] Heavy Flavor Averaging Group (HFAG), Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and* τ*-lepton properties as of summer 2014*, arXiv:1412.7515.

[35] LHCb, R. Aaij *et al.*, *Determination of the quark coupling strength* $|V_{ub}|$ *using baryonic decays*, arXiv:1504.01568.

[36] BaBar Collaboration, J. P. Lees *et al.*, *Evidence of* $B^+ \to \tau^+\nu$ *decays with hadronic B tags*, Phys. Rev. **D88** (2013), no. 3 031102, arXiv:1207.0698.

[37] Belle, A. Abdesselam *et al.*, *Measurement of the branching fraction of* $B^+ \to \tau^+\nu_\tau$ *decays with the semileptonic tagging method and the full Belle data sample*, arXiv:1409.5269.

[38] R. D. Peccei, *The Strong CP problem and axions*, Lect. Notes Phys. **741** (2008) 3, arXiv:hep-ph/0607268.

[39] C. A. Baker *et al.*, *An Improved experimental limit on the electric dipole moment of the neutron*, Phys. Rev. Lett. **97** (2006) 131801, arXiv:hep-ex/0602020.

[40] G. V. R.C̃rewther, P. di Vecchia and E. Witten, *Chiral estimate of the electric dipole moment of the neutron in quantum chromodynamics*, Phys. Lett. **B88** (1979) 123.

[41] K. G. Wilson, *Non-lagrangian models of current algebra*, Phys. Rev. **179** (1969) 1499.

[42] A. Pich, *Effective field theory: Course*, arXiv:hep-ph/9806303.

[43] LHCb collaboration, R. Aaij *et al.*, *Observation of the resonant character of the* $Z(4430)^-$ *state*, Phys. Rev. Lett. **112** (2014) 222002, arXiv:1404.1903.

[44] LHCb collaboration, R. Aaij *et al.*, *Evidence for pentaquark-charmonium states in* $\Lambda_b^0 \to J/\psi p K^-$ *decays*, Phys. Rev. Lett. **115** (2015) 07201, arXiv:1507.03414.

[45] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[46] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, arXiv:1405.7808, submitted to JINST.

[47] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[48] 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.

[49] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, arXiv:1210.6861.

[50] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175; T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[51] I. Belyaev *et al.*, *Handling of the generation of primary events in* Gauss*, the LHCb simulation framework*, Nuclear Science Symposium Conference Record (NSS/MIC) **IEEE** (2010) 1155.

[52] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[53] 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.

[54] GEANT4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270; GEANT4 collaboration, S. Agostinelli *et al.*, *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[55] M. Clemencic *et al.*, *The LHCb simulation application,* Gauss*: design, evolution and experience*, J. of Phys. : Conf. Ser. **331** (2011) 032023.

[56] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[57] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, Nucl. Instrum. Meth. **A543** (2005) 577, arXiv:physics/0408124.

[58] L. Breiman, *Bagging predictors*, Machine Learning **24** (1996), no. 2 123.

[59] A. Krogh and J. Vedelsby, *Neural network ensembles, cross validation, and active learning*, in *Advances in Neural Information Processing Systems*, pp. 231–238, MIT Press, 1995.

[60] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning and an application to boosting*, Jour. Comp. and Syst. Sc. **55** (1997) 119.

[61] J. Stevens and M. Williams, *uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers*, JINST **8** (2013) P12013, arXiv:1305.7248.

[62] H. Zou, R. Li, X. Wang, and C. Lu, *The CKM suppressed $B(B_s) \to \overline{D}_{(s)}P, \overline{D}_{(s)}V, \overline{D}^*_{(s)}P, \overline{D}^*_{(s)}V$ decays in perturbative QCD approach*, J. Phys. **G37** (2010) 015002, arXiv:0908.1856.

[63] R. Mohanta, *Searching for new physics in the rare decay $B^+ \to D_s^{*+}\phi$*, Phys. Lett. **B540** (2002) 241, arXiv:hep-ph/0205297.

[64] R. Mohanta and A. K. Giri, *Possible signatures of unparticles in rare annihilation type B decays*, Phys. Rev. **D76** (2007) 057701, arXiv:0707.3308.

[65] C.-D. Lu, *Calculation of pure annihilation type decay $B^+ \to D^+_{(s)}\phi$*, Eur. Phys. J. **C24** (2002) 121, arXiv:hep-ph/0112127.

[66] M. Beneke, G. Buchalla, M. Neubert, and C. T. Sachrajda, *QCD factorization for exclusive, nonleptonic B meson decays: General arguments and the case of heavy light final states*, Nucl. Phys. **B591** (2000) 313, arXiv:hep-ph/0006124.

[67] I. Narsky, *Statpatternrecognition: A c++ package for statistical analysis of high energy physics data*, arXiv:physics/0507143.

[68] LHCb collaboration, *First observations and branching fraction measurements of $\overline{B}^0_s$ to double-charm final states*, LHCb-CONF-2012-009.

[69] M. Pivk and F. R. Le Diberder, *Plots: A statistical tool to unfold data distributions*, Nuclear Instruments and Methods in Physics Research A **555** (2005) 356, physics/04.

[70] K. S. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun. **136** (2001) 198, arXiv:hep-ex/0011057.

[71] LHCb collaboration, R. Aaij *et al.*, *Observation of CP violation in $B^\pm \to DK^\pm$ decays*, Phys. Lett. **B712** (2012) 203, Erratum ibid. **B713** (2012) 351, arXiv:1203.3662.

[72] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60.

[73] G. J. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Phys. Rev. D **57** (1998) 3873.

[74] LHCb collaboration, R. Aaij *et al.*, *Measurements of the branching fractions and CP asymmetries of $B^\pm \to J/\psi\pi^\pm$ and $B^\pm \to \psi(2S)\pi^\pm$ decays*, Phys. Rev. **D85** (2012) 091105(R), arXiv:1203.3592.

[75] LHCb collaboration, R. Aaij *et al.*, *Measurement of form-factor-independent observables in the decay* $B^0 \to K^{*0}\mu^+\mu^-$, Phys. Rev. Lett. **111** (2013) 191801, arXiv:1308.1707.

[76] LHCb collaboration, *Angular analysis of the* $B^0_d \to K^{*0}\mu^+\mu^-$ *decay*, Mar, 2015. LHCb-CONF-2015-002.

[77] W. Altmannshofer and D. M. Straub, *New physics in* $b \to s$ *transitions after LHC run 1*, arXiv:1411.3161.

[78] LHCb collaboration, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay* $B^0_s \to \phi\mu^+\mu^-$, arXiv:1506.08777, submitted to JHEP.

[79] Belle Collaboration, H. Guler *et al.*, *Study of the* $K^+\pi^+\pi^-$ *Final State in* $B^+ \to J/\psi K^+\pi^+\pi^-$ *and* $B^+ \to \psi' K^+\pi^+\pi^-$, Phys. Rev. **D83** (2011) 032005, arXiv:1009.5256.

[80] M. Suzuki, *Strange axial-vector mesons*, Phys. Rev. D **47** (1993) 1252.

[81] A. Tayduganov, E. Kou, and A. Le Yaouanc, *The strong decays of K1 resonances*, Phys. Rev. **D85** (2012) 074011, arXiv:1111.6307.

[82] H. Hatanaka and K.-C. Yang, $B \to K_{(1)}\gamma$ *decays in the light-cone QCD sum rules*, Phys. Rev. **D77** (2008) 094023, arXiv:0804.3198.

[83] H.-Y. Cheng, *Revisiting axial-vector meson mixing*, Phys. Lett. **B707** (2012) 116, arXiv:1110.2249.

[84] F. Divotgey, L. Olbrich, and F. Giacosa, *Phenomenology of axial-vector and pseudovector mesons: decays and mixing in the kaonic sector*, Eur. Phys. J. **A49** (2013) 135, arXiv:1306.1193.

[85] H.-Y. Cheng, *Mixing angle of* K_1 *axial vector mesons*, arXiv:1311.2370.

[86] H. Hatanaka and K.-C. Yang, $K_1(1270) - K_1(1400)$ *Mixing Angle and New-Physics Effects in* $B \to K_1\ell^+\ell^-$ *Decays*, Phys. Rev. **D78** (2008) 074007, arXiv:0808.3731.

[87] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[88] A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039.

[89] R. Essig *et al.*, *Working Group Report: New Light Weakly Coupled Particles*, arXiv:1311.0029.

[90] F. Bezrukov and D. Gorbunov, *Light inflaton Hunter's Guide*, JHEP **1005** (2010) 010, arXiv:0912.0390.

[91] M. P. Hertzberg and J. Karouby, *Generating the Observed Baryon Asymmetry from the Inflaton Field*, Phys. Rev. **D89** (2014) 063523, arXiv:1309.0010.

[92] M. P. Hertzberg and J. Karouby, *Baryogenesis from the Inflaton Field*, arXiv:1309.0007.

[93] F. Bezrukov and D. Gorbunov, *Relic Gravity Waves and 7 keV Dark Matter from a GeV scale inflaton*, arXiv:1403.4638.

[94] F. Bezrukov and D. Gorbunov, *Light inflaton after LHC8 and WMAP9 results*, JHEP **07** (2013) 140, arXiv:1303.4395.

[95] CHARM collaboration, F. Bergsma *et al.*, *Search for axion like particle production in 400 GeV proton-copper interactions*, Phys. Lett. **B157** (1985) 458.

[96] Y. Nomura and J. Thaler, *Dark Matter through the Axion Portal*, Phys. Rev. **D79** (2009) 075008, arXiv:0810.5397.

[97] CERN NA48/2, *Search for the dark photon in π^0 decays*, arXiv:1504.00607.

[98] BaBar, J. P. Lees *et al.*, *Search for a Dark Photon in e^+e^- Collisions at BaBar*, Phys. Rev. Lett. **113** (2014), no. 20 201801, arXiv:1406.2980.

[99] S. Alekhin *et al.*, *A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case*, arXiv:1504.04855.

[100] B. Batell, M. Pospelov, and A. Ritz, *Multi-lepton Signatures of a Hidden Sector in Rare B Decays*, Phys. Rev. **D83** (2011) 054005, arXiv:0911.4938.

[101] M. Williams, *Searching for a particle of unknown mass and lifetime in the presence of an unknown non-monotonic background*, JINST **10** (2015), no. 06 P06002, arXiv:1503.04767.

[102] E. Gross and O. Vitells, *Trial factors or the look elsewhere effect in high energy physics*, Eur. Phys. J. **C70** (2010) 525, arXiv:1005.1891.

[103] R. B. Davies, *Hypothesis testing when a nuisance parameter is present only under the alternative*, Biometrika **74** (1987) 33.

[104] LHCb collaboration, R. Aaij *et al.*, *Observation of a resonance in $B^+ \to K^+\mu^+\mu^-$ decays at low recoil*, Phys. Rev. Lett. **111** (2013) 112003, arXiv:1307.7595.

[105] C. Bobeth, G. Hiller, D. van Dyk, and C. Wacker, *The Decay $B \to K\ell^+\ell^-$ at Low Hadronic Recoil and Model-Independent Delta B = 1 Constraints*, JHEP **1201** (2012) 107, arXiv:1111.2558.

[106] M. Needham, *Identification of ghost tracks using a likelihood method*, tech. rep., 2008.

[107] V. V. Vladimirsky *et al.*, *Comparative analysis of the narrow resonances*, Physics of Atomic Nuclei **71** (2008), no. 12 2129.

[108] G. Punzi, *Sensitivity of Searches for New Signals and Its Optimization*, in *Statistical Problems in Particle Physics, Astrophysics, and Cosmology* (L. Lyons, R. Mount, and R. Reitmeyer, eds.), p. 79, 2003. arXiv:physics/0308063.

[109] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the decay* $B^0 \to K^{*0}\mu^+\mu^-$, JHEP **08** (2013) 131, arXiv:1304.6325.

[110] A. Crivellin and S. Pokorski, *Can the differences in the determinations of* V_{ub} *and* V_{cb} *be explained by New Physics?*, Phys. Rev. Lett. **114** (2015), no. 1 011802, arXiv:1407.1320.

[111] HPQCD, C. Bouchard *et al.*, *Rare decay* $B \to K\ell^+\ell^-$ *form factors from lattice QCD*, Phys. Rev. **D88** (2013), no. 5 054509, arXiv:1306.2384.

[112] H. J. Rothe, *Lattice Gauge Theories: An Introduction; 4th ed.*, World Scientific Lecture Notes in Physics, World Scientific, Singapore (2012).

[113] LHCb collaboration, R. Aaij *et al.*, *Measurement of the decay* $\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau$, arXiv:1506.08614, to appear in Phys.Rev.Lett.

[114] BaBar, J. P. Lees *et al.*, *Measurement of an Excess of* $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ *Decays and Implications for Charged Higgs Bosons*, Phys. Rev. **D88** (2013), no. 7 072012, arXiv:1303.0571.

[115] BaBar, J. P. Lees *et al.*, *Evidence for an excess of* $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ *decays*, Phys. Rev. Lett. **109** (2012) 101802, arXiv:1205.5442.

[116] Belle, A. Bozek *et al.*, *Observation of* $B^+ \to \bar{D}^{*0}\tau^+\nu_\tau$ *and Evidence for* $B^+ \to \bar{D}^0\tau^+\nu_\tau$ *at Belle*, Phys. Rev. **D82** (2010) 072005, arXiv:1005.2302.

[117] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using* $B^+ \to K^+\ell^+\ell^-$ *decays*, Phys. Rev. Lett. **113** (2014) 151601, arXiv:1406.6482.

Appendix A

License

License to use Fig 6.3.

American Physical Society
License Details

Jun 05, 2015

This is an Agreement between Samuel Hall ("You") and American Physical Society ("Publisher"). It consists of your order details, the terms and conditions provided by American Physical Society, and the payment instructions.

License Number	3622160100576
License date	May 04, 2015
Licensed content publisher	American Physical Society
Licensed content publication	Physical Review D
Licensed content title	${K}_{1}(1270)$-${}{K}_{1}(1400)$
Licensed copyright line	Copyright © 2008, American Physical Society
Licensed content author	Hisaki Hatanaka and Kwei-Chou Yang
Licensed content date	Oct 7, 2008
Volume number	78
Type of Use	Thesis/Dissertation
Requestor type	Student
Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs/tables/figures	1
Portion description	Figure 1
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	0 to 499
The requesting person/organization is:	Samuel Hall
Order reference number	None
Title of your thesis / dissertation	Searching for beyond the Standard Model physics using direct and indirect methods at LHCb
Expected completion date	Jul 2015
Expected size (number of pages)	150
Total	0.00 GBP
Terms and Conditions	

Terms and Conditions

The American Physical Society (APS) is pleased to grant the Requestor of this license a non-exclusive, non-transferable permission, limited to [**print** and/or **electronic** format, depending on what they chose], provided all criteria outlined below are followed.

1. For electronic format permissions, Requestor agrees to provide a hyperlink from the reprinted APS material using the source material's DOI on the web page where the work appears. The hyperlink should use the standard DOI resolution URL, http://dx.doi.org/{DOI}. The hyperlink may be embedded in the copyright credit line.

2. For print format permissions, Requestor agrees to print the required copyright credit line on the first page where the material appears: "Reprinted (abstract/excerpt/figure) with permission from [(FULL REFERENCE CITATION) as follows: Author's Names, APS Journal Title, Volume Number, Page Number and Year of Publication.] Copyright (YEAR) by the American Physical Society."

3. Permission granted in this license is for a one-time use and does not include permission for any future editions, updates, databases, formats or other matters. Permission must be sought for any additional use.

4. Use of the material does not and must not imply any endorsement by APS.

5. Under no circumstance does APS purport or intend to grant permission to reuse materials to which it does not hold copyright. It is the requestors sole responsibility to ensure the licensed material is original to APS and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover or table does not indicate that it was reprinted by APS, with permission from another source.

6. The permission granted herein is personal to the Requestor for the use specified and is not transferable or assignable without express written permission of APS. This license may not be amended except in writing by APS.

7. You may not alter, edit or modify the material in any manner.

8. You may translate the materials only when translation rights have been granted.

9. You may not use the material for promotional, sales, advertising or marketing purposes.

10. The foregoing license shall not take effect unless and until APS or its agent, Copyright Clearance Center (CCC), receives payment in full in accordance with CCC Billing and Payment Terms and Conditions, which are incorporated herein by reference.

11. Should the terms of this license be violated at any time, APS or CCC may revoke the license with no refund to you and seek relief to the fullest extent of the laws of the USA. Official written notice will be made using the contact information provided with the permission request. Failure to receive such notice will not nullify revocation of the permission.

12. APS reserves all rights not specifically granted herein.

13. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.