Measurements of the local and non-local contributions in $B^{\pm} \to K^{\pm}\mu^{\pm}\mu^{\mp}$ decays using LHCb data

By

LAKSHAN RAM MADHAN MOHAN

School of Physics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science.

JUNE 2022

Abstract

In the study of the decay of B mesons, the observed deviations from standard model predictions could potentially be explained by underestimated non-local hadronic intermediate states. To answer this question, the contribution of local and non-local contributions to $B^{\pm} \to K^{\pm}\mu^{\pm}\mu^{\mp}$ decays are measured by analysing the invariant dimuon mass distribution in the range: $300 < m_{\mu\mu} < 4700$ MeV/c^2 ($0.09 < q^2 < 22.09$ GeV$^2/c^4$). The analysis uses pp collision data collected by the LHCb experiment between 2011 and 2018 corresponding to an integrated luminosity of 9 fb^{-1}. The non-local contributions are modelled using hadronic dispersion relations including a description of intermediate one-particle and two-particle states which scatter to a pair of muons. The largest one-particle states interfering with the local contribution are J/ψ and $\psi(2S)$ and the two-particle states described are DD, D^*D^* and DD^*. The interference from J/ψ is observed to be small while a large component of the two-particle states, along with the $\psi(2S)$, is observed to destructively interfere with the local contribution. The effect of this interference on the vector current (C_9) and axial-vector current (C_{10}) Wilson coefficients, which encode heavy standard model and beyond standard model physics, are discussed. The branching fractions of the one-particle and two-particle states are also measured.

Acknowledgements

Words cannot begin to describe my gratitude for the guidance and support I have received from my supervisor, Kostas Petridis, this thesis would not have been possible without you. Thank you for teaching me and being a great friend over the past four years, your passion for physics will continue to inspire me. I would also like to thank my second supervisor, Jonas Rademacker for being a constant source of encouragement and making sure I wasn't broke. I am grateful for the Geoffrey Fertel Fund which allowed me to do this work.

The discussions I had in our weekly meetings with Andrei Golutvin, Igor Diachkov, Patrick Owen and Thomas Blake were of extreme value to me. The tips and trinkets of knowledge offered by everyone has always come through whenever I have been stuck with a problem. It has been a pleasure working with all of you. A special thanks to Igor Diachkov for calculating the data/simulation corrections I needed to perform this measurement.

I would also like to thank the LHCb group at Bristol including Alex Marshall, Jake Reich, Richard Lane and Rui Wang for being a constant source of fun especially in the last year of my PhD. In particular I would like to thank Marco Adinolfi and Keith Clarke for the great banter you offered while we were either salvaging components, cleaning boxes or gluing mirrors. I would also like to thank Marco Adinolfi for being a huge source of support during my time at CERN. The fun times I had with my cohort: Ben Westhenry, David Anthony, Nikos Stylianou, Sam Eriksen and everyone else at the Bristol particle physics group is something I will always cherish, thank you.

For all the tough times the pandemic has caused, being in Switzerland for the majority of it was an unforgettable experience. This largely would not have been possible without my incredible flatmates Adam Rennie and Nikos Stylianou. In addition, a special thanks is due to Biljana Mitreska, Katherine Rybacki and Naomi Cooke for always being up for a hike and for indulging my cooking passion and complimenting it (even at times when it was just about edible). I am also grateful for the rest of my LTA cohort whom I have had the fortune of meeting and spending time with.

The long hours of chatting and gaming with Dude, Ajeya, Athipan and Sanjay kept me going during the toughest of times, I will always be grateful to call you lot as friends. The backbone of my support is and always has been my Mum, Dad and Bharath, thank you for believing in me and encouraging me throughout.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

Author's Contribution

The measurement presented in this thesis uses data collected by the LHCb experiment which is operated by the LHCb collaboration. In Chapter 3, the author recognises the work by different sub-groups of the collaboration who develop/maintain the detectors and software which enabled this measurement to be performed. The simulation of decays used in the thesis has been produced by the LHCb simulation group. The author also recognises the CERN accelerator department for the excellent performance of the LHC.

Chapter 4 to Chapter 8 is the author's own work except for Section 4.1 which is performed centrally by the LHCb collaboration and Section 5.1 which is work by Igor Diachkov.

Preface, a summary for non-experts

Most of the known matter content of the Universe is made of subatomic particles known as up quark (u), down quark (d) and electrons (e). The up and and down quarks form the neutrons and protons which make up the nucleus of atoms and electrons are the charged particles which exist around the nucleus. At higher energies such as immediately after the Big Bang, other heavier quarks known as the strange (s), charm (c), top (t) and bottom (b) quarks existed in abundance. Along with them, heavier versions of the electron known as muons (μ) and taus (τ) were also much more common. In addition, identical copies of these particles but with opposite charges known as anti-particles were are also equally common.

Presently using machines such as the Large Hadron Collider, protons are accelerated to near the speed of light and collided with each other emulating conditions of the early Universe. In these conditions, the heavier subatomic particles are created. By studying the interaction and properties of these particles, the physics of the process which were in play 13.77 billion years ago can be directly probed. Understanding these physical processes will help explain the structure of the Universe seen today.

The collaborative work of many physicists over the past century has resulted in the Standard Model of particle physics. This model describes all the fundamental particles and their interactions. However, among a few others, it does not explain one crucial observation, the existence of the physical Universe itself. In fact, according to this model, matter and anti-matter which were both created in equal amounts by the Big Bang must have annihilated and left nothing behind. The remaining matter content of the observed Universe is the result of an asymmetry in the interaction of matter and anti-matter which is unaccounted for in our theories. This means that certain process which ocurred in the early stages of the Universe are still not fully understood.

The best possible way to search for the unknown is by scrutinizing the Standard Model which is currently the best available tool. Through the use of precision measurements the Standard Model can be tested to find the place where it breaks down. The measurement presented in this thesis aims to achieve this by studying the process of $b \to s\ell^+\ell^-$ transitions. In this process a b quark transforms to a s quark and two leptons (ℓ) which could be an electron, muon or tau. This process happens rarely and the properties of it can be predicted

precisely using the Standard Model. Moreover, the Standard Model predicts there should be no differentiation between the three leptons and that decay properties must be the same.

Observations made in the past decade indicate that the leptons may not be behaving similarly and that especially, the observed properties of $b \to s\mu^+\mu^-$ and $b \to c\tau\overline{\nu}_\tau$ is different from Standard Model predictions. These observations are the so-called "B-anomalies". This thesis studies the decay of a B meson to a kaon (K) and two oppositely charged muons ($B^\pm \to K^\pm\mu^\pm\mu^\mp$). The B meson is made of a b and u, the K is made of s and u therefore this decay proceeds via a $b \to s\mu^+\mu^-$ transition with the u being the so-called spectator quark. The decay of $B^\pm \to K^\pm\mu^\pm\mu^\mp$ could also proceed via a host of intermediate states which are difficult to theoretically estimate. Using data-driven techniques to explicitly account for the intermediate states and the data collected by the LHCb experiment from proton-proton collisions between 2011 and 2018, this thesis studies the presence of Beyond Standard Model effects.

The immense amount of theoretical and experimental work dedicated to study the B-anomalies has brought us to a turning point in our understanding of nature. The data to be collected and analysed over the next few years is expected to contain the precision necessary to be sensitive to enhanced New Physics (NP) processes. Observation of such effects would imply the existence of novel force carriers and define the future direction of particle physics research.

Table of Contents

List of Tables

List of Figures

CHAPTER

1 Introduction

To probe the fundamental interactions the differential decay rate of $B^+ \to K^+\mu^+\mu^-$ decays as a function of the invariant dimuon mass is studied. The inclusion of the charged-conjugate process ($B^- \to K^-\mu^-\mu^+$) is implied throughout this thesis. The decay of $B^+ \to K^+\mu^+\mu^-$ proceeds via the $b \to s\ell^+\ell^-$ transition as shown in Figure 1.1 which is a process where the quark 'flavour' is changed.

In this transition of b quarks, within the Standard Model (SM) of particle physics, two energy scales are involved. These are the electroweak scale defined by the W boson mass which mediates the transition at the quark level and the scale of the b quark mass (m_b) which governs the energy released by the transition. In addition, potential physics Beyond Standard Model (BSM) could exist far above the W boson mass scale. It is therefore useful to construct a low energy Effective Field Theory (EFT) where all heavy degrees of freedom such as the W, Z boson and potential NP contributions are integrated out. Such an EFT will be sensitive to all possible heavy interactions within and beyond the SM. After presenting a brief description of the SM, the EFT framework and how it is used to describe $b \to s\ell^+\ell^-$ transitions is described in

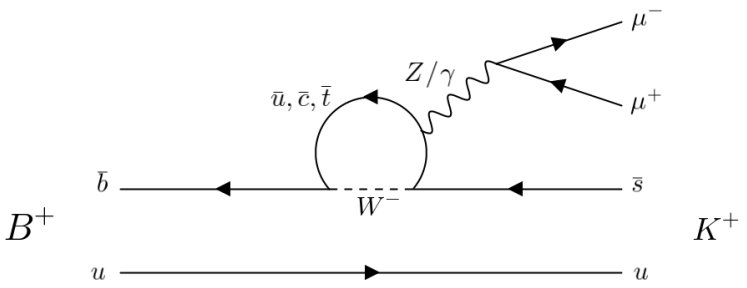

Figure 1.1: The Feynman diagram showing the electroweak decay of $B^+ \to K^+\mu^+\mu^-$.

Chapter 2.

In the past decade, measurement of the decay of B mesons with different final states have shown anomalous results when compared to SM predictions [1–7]. Global analyses of these measurements also show a deviation from the SM with a significance of 4.3σ [8] with various specific NP scenarios preferred over the SM with 4-6σ significance [9, 10]. However, there is a long standing debate as to whether the difficult-to-calculate long distance hadronic effects could be the cause of the observed anomalies [11].

The measurement presented in this thesis aims to answer this question by using a data-driven method to account for the hadronic effects and determine the strength of NP couplings. The pp collision data collected by the LHCb experiment between 2011 and 2018 with a total integrated luminosity of $9\,\text{fb}^{-1}$ is used. The previous iteration of this measurement in Ref. [12] uses LHCb data collected in 2011 and 2012 with integrated luminosity of $3\,\text{fb}^{-1}$. While the available statistics has increased, this measurement uses a better theoretically motivated model to estimate the hadronic effects and therefore offers an improved physics interpretation of the result.

The LHCb detector and its performance is discussed in Chapter 3 with the work done for the upgrade of the RICH1 discussed in Section 3.7. After the application of kinematic and Particle Identification (PID) selections, the collected data set contains 3524407 $B^+ \to K^+\mu^+\mu^-$ candidates (a factor $\approx$3.5 larger than Ref. [12]) within the fit window $300 < m_{\mu\mu} < 4700\,\text{MeV}/c^2$ and the signal region $5239.25 < m_{K\mu\mu} < 5319.25\,\text{MeV}/c^2$. A detailed account of the selection procedure is presented in Chapter 4.

In Chapter 5, the method used to account for the detector's efficiency and resolution is discussed. After estimating potential background contributions, the fit model used to measure the local and non-local contributions to $B^+ \to K^+\mu^+\mu^-$ is developed in Chapter 6. Finally the results from the fit to data and potential systematic effects are discussed in Chapter 7 before concluding in Chapter 8.

CHAPTER

2 Theoretical description of $b \to s\ell^+\ell^-$ transitions

2.1 The Standard Model of Particle Physics

This section follows the argument presented in Ref. [13]

The Standard Model (SM) of particle physics is a Quantum Field Theory (QFT) described by the gauge group,

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y. \tag{2.1}$$

Invariance of physical processes under generalised phase transformations described by these unitary groups is the principle of gauge invariance. This principle determines the nature of the fundamental constituents of matter and their interactions. The symmetry of the SM Lagrangian under a local $\mathrm{SU}(3)_C$ transformation gives rise to the strong interactions [14]. A unified description of interactions relevant to the weak and electromagnetic force is described by the gauge invariance of $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ [15]. Through the Higgs mechanism [16] (discussed in Section 2.1.2), this symmetry is spontaneously broken down to $\mathrm{U}(1)_{EM}$, which is the symmetry group of Quantum Electrodynamics (QED) and describes the electromagnetic interaction of particles.

The quantum numbers of $\mathrm{SU}(3)_C$ are called *color*, of which there are three (red, green and blue); The gauge group of weak interactions, $\mathrm{SU}(2)_L$ has the quantum number called *weak isospin* with two physical charges ($\pm\frac{1}{2}$); The quantum number of $\mathrm{U}(1)_Y$ is *hypercharge* (Y) which can be any real number. The hypercharge is normalized with[1]

$$\mathbb{Q} = T^3 + Y, \tag{2.2}$$

where T^3 is the third component of the weak isospin from $\mathrm{SU}(2)_L$ and $\mathbb{Q}$ is the electric charge.

[1] In some literature Y is replaced by Y/2

The known contents of the Universe are Fermions and Bosons. Fermions are described by Dirac spin-$\frac{1}{2}$ fields (ψ) and can be decomposed into left-handed and right-handed fields using the chiral projection operators P_L and P_R,

$$\psi = \psi_L + \psi_R, \tag{2.3}$$

$$\psi_L = P_L\psi \equiv \frac{1-\gamma_5}{2}\psi \tag{2.4}$$

$$\psi_R = P_R\psi \equiv \frac{1+\gamma_5}{2}\psi, \tag{2.5}$$

with $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ being the product of the four Dirac matrices (γ^μ).

Three generation of quarks and leptons make up the observed Fermions. The left-handed quark fields are represented as a set of three $\mathrm{SU}(2)_L$ doublets of $\mathrm{SU}(3)_C$ triplet quarks and can be written as,

$$Q_L^i = \begin{bmatrix} u_L^i \\ d_L^i \end{bmatrix} \in \left\{ \begin{bmatrix} u_L \\ d_L \end{bmatrix}, \begin{bmatrix} c_L \\ s_L \end{bmatrix}, \begin{bmatrix} t_L \\ b_L \end{bmatrix} \right\}, \tag{2.6}$$

the set of three right handed singlets of up-type and down-type quarks are then,

$$u_R^i \in \{u_R, c_R, t_R\}, \tag{2.7}$$

$$d_R^i \in \{d_R, s_R, b_R\}. \tag{2.8}$$

Analogously the lepton fields can be written as a set of three $\mathrm{SU}(2)_L$ left-handed doublets and right-handed singlets,

$$L_L^i = \begin{bmatrix} \nu_L^i \\ l_L^i \end{bmatrix} \in \left\{ \begin{bmatrix} \nu_L^e \\ e_L \end{bmatrix}, \begin{bmatrix} \nu_L^\mu \\ \mu_L \end{bmatrix}, \begin{bmatrix} \nu_L^\tau \\ \tau_L \end{bmatrix} \right\}, \tag{2.9}$$

$$l_R^i \in \{e_R, \mu_R, \tau_R\}. \tag{2.10}$$

Right-handed neutrinos are not included in the SM[2].

Only quarks feel the strong force and are charged with the colour quantum numbers. All Fermions feel the weak force as they all have a weak isospin and the Fermions with an electric charge (quarks, e, μ and τ) also feel the electromagnetic force. All these particles also have antiparticle states with their electric charge and colour quantum numbers flipped. The three generation of particles are identical and only differ by their mass.

The local gauge invariance requirement of $\mathrm{SU}(3)_C$, $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ predict massless gauge vector Bosons which are spin-1 particles. These are the eight gluons G_μ^a ($a \in [1,8]$) for $\mathrm{SU}(3)_C$, three W_μ^a ($a \in [1,3]$) for $\mathrm{SU}(2)_L$ and one B_μ for $\mathrm{U}(1)_Y$. The electroweak symmetry breaking mixes W_μ^a and B_μ to create the particles $W^\pm$, Z^0 and γ. Only γ stays massless and belongs to the unbroken group $\mathrm{U}(1)_{EM}$.

[2]Within the SM, neutrinos stay exactly massless. However, experimental observation show that neutrinos have a small mass. Introducing right-handed neutrinos to the SM is one solution to this problem and is out of scope for the present discussion.

2.1.1 The SM Lagrangian

The SM Lagrangian can be split into four terms and written as

$$\mathcal{L}_{SM} = \mathcal{L}_{Fermion} + \mathcal{L}_{gauge} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}. \tag{2.11}$$

The first term is the Fermion term and for a Fermion field ψ, it contains terms of the form[3]

$$\mathcal{L}_{Fermion} \sim \overline{\psi} i D_\mu \gamma^\mu \psi, \tag{2.12}$$

where D_μ is the covariant derivative and is defined as,

$$D_\mu = \partial_\mu - i g_1 Y B_\mu \qquad \text{for lepton singlets } l_R^i, \tag{2.13}$$

$$D_\mu = \partial_\mu - i g_1 Y B_\mu - i g_2 \frac{\sigma^a}{2} W_\mu^a \qquad \text{for lepton doublets } L_L^i, \tag{2.14}$$

$$D_\mu = \partial_\mu - i g_1 Y B_\mu - i g_s \frac{\lambda^a}{2} G_\mu^a \qquad \text{for quark singlets } u_R^i \text{ and } d_R^i, \tag{2.15}$$

$$D_\mu = \partial_\mu - i g_1 Y B_\mu - i g_2 \frac{\sigma^a}{2} W_\mu^a - i g_s \frac{\lambda^a}{2} G_\mu^a \qquad \text{for quark doublets } Q_L^i, \tag{2.16}$$

$$\tag{2.17}$$

where g_1, g_2 and g_s are the couplings for $\mathrm{U}(1)_Y$, $\mathrm{SU}(2)_L$ and $\mathrm{SU}(3)_C$ respectively. The notations σ^a and λ^a are the well known Pauli and Gell-Mann matrices respectively. The covariant derivatives are the same for all three generations.

The second term in Eq. 2.11 is the kinetic term for the three gauge Bosons and contains terms of the form

$$\mathcal{L}_{gauge} \sim -\frac{1}{4} F_{\mu\nu}^a F^{\mu\nu,a}, \qquad \text{for } F_{\mu\nu}^a \in \{G_{\mu\nu}^a, W_{\mu\nu}^a, B_{\mu,\nu}^a\} \tag{2.18}$$

where the $F_{\mu\nu}^a$ are the gauge's field strength tensor and given by,

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c, \tag{2.19}$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_2 \epsilon^{abc} W_\mu^b W_\nu^c, \tag{2.20}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{2.21}$$

$$\tag{2.22}$$

f^{abc} and ϵ^{abc} are the group structure constants of $\mathrm{SU}(3)_C$ and $\mathrm{SU}(2)_L$ respectively.

2.1.2 The Higgs mechanism

Mass terms of the form $-m\overline{\psi}\psi = -m(\overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L)$ for Fermions are not allowed in the SM. This is because they couple left-handed and right-handed fields that transform differently under

[3] Space-time contraction is implied.

$SU(2)_L$ leaving the Lagrangian not invariant and therefore not allowed in the SM. Mass terms for gauge Bosons $M^2 A^a_\mu A^{\mu,a}$ are also not allowed as they break gauge symmetry. However, Fermions and Bosons have been observed to have mass. The Higgs mechanism provides a solution to this by introducing the Higgs boson and spontaneously breaking the symmetry of $SU(2)_L \times U(1)_Y \to U(1)_{EM}$.

The kinetic term of the Higgs with its couplings to the gauge Bosons and the Higgs potential $V(H)$, is contained within the term $\mathcal{L}_{Higgs}$ in Eq. 2.11 and is given by,

$$\mathcal{L}_{Higgs} = \left(D_\mu H\right)^\dagger \left(D^\mu H\right) - V(H), \tag{2.23}$$

here the covariant derivative D_μ is the same as for left-handed leptons given in Eq. 2.14.

The Higgs field is formed of two complex scalar fields $H = (H_+, H_0)^T$ which is a colour neutral $SU(2)_L$ doublet. The Higgs potential of the SM is,

$$V(H) = \lambda (H^\dagger H)^2 - \mu^2 H^\dagger H \tag{2.24}$$

$$= \lambda \left(H^\dagger H - \frac{\mu^2}{2\lambda} \right)^2 - \frac{\mu^4}{4\lambda^2}, \tag{2.25}$$

and if $\mu^2 > 0$, has a non zero minimum for[4]

$$H^\dagger H = \frac{\mu^2}{2\lambda}. \tag{2.26}$$

The vacuum expectation value (vev) of this field is then,

$$\frac{v}{\sqrt{2}} \equiv \sqrt{\frac{\mu^2}{2\lambda}}. \tag{2.27}$$

An example of this potential is shown in Figure 2.1. The full circle of degenerate minima at the trough is the vev and a global gauge choice corresponds to choosing a point on this circle as the ground state. By choosing a ground state, the gauge is fixed and the $SU(2)_L$ symmetry is spontaneously broken for perturbations around this ground state. Without loss of generality, the vacuum state of the field can be written as,

$$\langle 0|H|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.28}$$

Expanding the field around the chosen ground state yields,

$$H = \begin{pmatrix} h_+ + i\chi_+ \\ \frac{1}{\sqrt{2}}(v + h_0 + i\chi_0) \end{pmatrix}, \tag{2.29}$$

[4] Another convention for the potential $V(H)$ commonly seen in literature is to have both signs positive in Eq. 2.25, in this case a non zero minimum exist for $\mu^2 < 0$. The convention chosen in the argument above directly relates μ^2 to the squared Higgs mass m_h^2 as shown later.

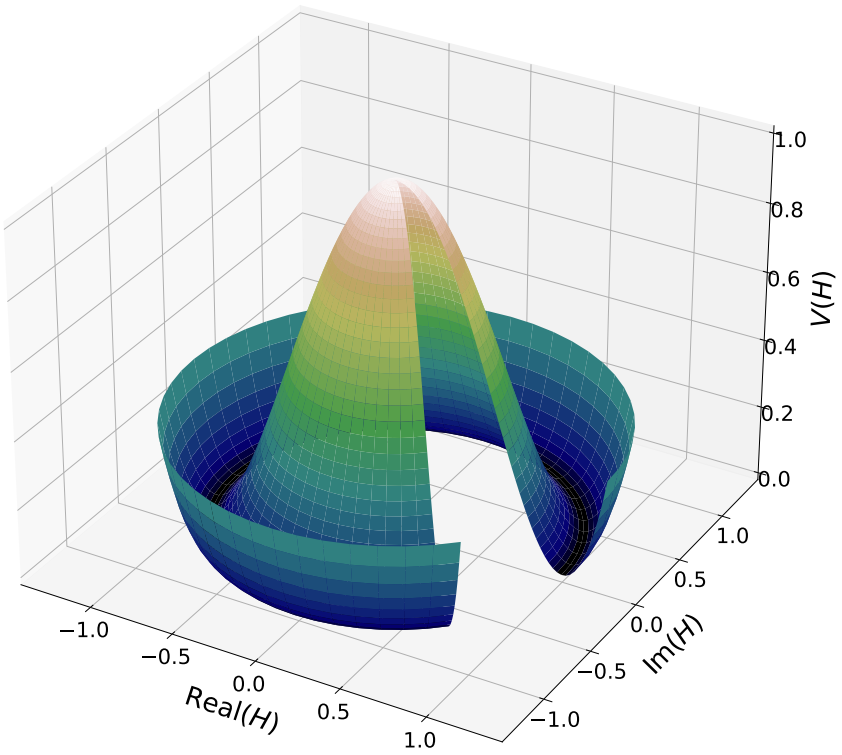

Figure 2.1: An illustration of the shape of the Higgs potential. The vacuum expectation value is the degenerate minima along the trough of this potential. The fact the ground state of the Higgs field has a non-zero vev breaks the electroweak symmetry.

where $h_{0,+}$, $\chi_{0,+}$ are real scalar fields. It can be shown that inserting this into Eq. 2.23 gives only one massive term and is proportional to h_0^2

$$2\mu^2 h_0^2 \implies m_h = \sqrt{2}\mu = \sqrt{2\lambda}v. \tag{2.30}$$

This is the mass of the Higgs, h_+, $\chi_{0,+}$ correspond to the three Goldstone Bosons and are massless. By choosing the 'unitary gauge' this form simplifies to[5]

$$H = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + h_0). \end{pmatrix}, \tag{2.31}$$

2.1.3 Gauge boson mixing and the origin of their mass

The Higgs mechanism provides a way to introduce mass terms into the SM Lagrangian. The gauge Bosons gain mass through their interaction terms with the Higgs field. The covariant derivative of the Higgs field is

$$D_\mu = \partial_\mu - ig_1 Y B_\mu - ig_2 \frac{\sigma^a}{2} W_\mu^a = \partial_\mu - i \begin{pmatrix} g_1 Y B_\mu + \frac{g_2}{2} W_\mu^3 & \frac{g_2}{2}(W_\mu^1 - W_\mu^2) \\ \frac{g_2}{2}(W_\mu^1 + W_\mu^2) & g_1 Y B_\mu - \frac{g_2}{2} W_\mu^3 \end{pmatrix}, \tag{2.32}$$

[5]Other gauge choices such as the Coulomb or Lorentz gauge will also lead to the same results. The unitary gauge is chosen as it makes the physics visible more readily.

where we have used the Pauli matrices σ^a ($a = 1,2,3$) as they are the generators of $SU(2)_L$. The hypercharge assignment for the Higgs field is $Y = \frac{1}{2}$.

Substituting this into $(D^\mu H)^\dagger (D_\mu H)$ gives the terms

$$\begin{aligned}(D^\mu H)^\dagger (D_\mu H) = &\frac{v^2}{8}\Big[g_2^2 (W_\mu^1 + iW_\mu^2)(W^{\mu,1} - iW^{\mu,2}) \\ &+ (g_2 W^{\mu,3} - g_1 B^\mu)(g_2 W_\mu^3 - g_2 B_\mu)\Big] + \text{other terms}\end{aligned} \tag{2.33}$$

In this equation the real $W_\mu^{1,2}$ can be replaced with $W_\mu^\pm$ through,

$$W_\mu^+ = [W_\mu^-]^\dagger = \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2) \tag{2.34}$$

This yields a mass term for m_W,

$$\frac{v^2 g_2^2}{4} W_\mu^- W^{\mu,+}, \qquad \text{with } m_W = \frac{v g_2}{2}. \tag{2.35}$$

The second term of the RHS in Eq.2.33 also looks like a mass term however it indicates at mixing of W_μ^2 and B^μ. By introducing the Weinberg mixing angle θ_w an orthogonal transformation of $(B_\mu, W_\mu^3) \to (A_\mu, Z_\mu)$ can be performed,

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} cos\theta_w & sin\theta_w \\ -sin\theta_w & cos\theta_w \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} = \begin{pmatrix} cos\theta_w B_\mu + sin\theta_w W_\mu^3 \\ -sin\theta_w B_\mu + cos\theta_w W_\mu^3 \end{pmatrix}, \tag{2.36}$$

By relating the couplings g_1, g_2 to the mixing angle,

$$g_1 = sin\theta_w \sqrt{g_1^2 + g_2^2}, \tag{2.37}$$

$$g_2 = cos\theta_w \sqrt{g_1^2 + g_2^2}, \tag{2.38}$$

Eq. 2.36 becomes,

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \frac{1}{\sqrt{g_1^2 + g_2^2}} \begin{pmatrix} g_1 W_\mu^3 + g_2 B_\mu \\ g_2 W_\mu^3 - g_1 B_\mu \end{pmatrix}. \tag{2.39}$$

Using this relation, the second term in Eq. 2.33 can be rewritten as,

$$\frac{1}{2}\frac{v^2(g_1^2 + g_2^2)}{4} Z_\mu Z^\mu \qquad \text{with } m_Z = \frac{v^2(g_1^2 + g_2^2)}{4} = \frac{m_W}{cos\theta_w}, \tag{2.40}$$

giving the mass term for a neutral Z^0 boson. Finally A_μ is the massless photon described in QED.

2.1.4 The origin of the flavour structure in the SM

The mass terms of the Fermions and the flavour structure in the SM Lagrangian arises from the interaction of the Fermions with the Higgs field. This is encoded in the $\mathcal{L}_{Yukawa}$ of Eq. 2.11, it can further be split for the quarks and leptons and is given by

$$\mathcal{L}^{q}_{Yukawa} = -y^{d}_{ij}\overline{Q^{i}_{L}}Hd^{j}_{R} - y^{u}_{ij}\overline{Q^{i}_{L}}H^{*}C^{j} + h.c., \tag{2.41}$$

and

$$\mathcal{L}^{l}_{Yukawa} = -y^{l}_{ij}\overline{L^{i}_{L}}Hl^{j}_{R} + h.c., \tag{2.42}$$

where Q_L, d_R,u_R, L_L and l_R are as defined in Eq. 2.6 to 2.10 and y are the couplings with the indices i, j spanning the different generations. The Hermitian Conjugate is denoted by $h.c.$ and $H^{*} = i\sigma^{2}H$. No such term for the neutrinos exist in the SM.

The ground state interaction of the Higgs field gives rise to quark Yukawa terms given by,

$$\begin{aligned}\mathcal{L}^{q}_{Yukawa} =& -y^{d}_{ij}\begin{bmatrix}\overline{u^{i}_{L}} & \overline{d^{i}}_{L}\end{bmatrix}\begin{bmatrix}0\\ \frac{v+h_0}{\sqrt{2}}\end{bmatrix}d^{j}_{R} - y^{d}_{ij}\overline{d^{j}_{R}}\begin{bmatrix}0 & \frac{v+h_0}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}u^{i}_{L}\\ d^{i}_{L}\end{bmatrix}\\ & -y^{u}_{ij}\begin{bmatrix}\overline{u^{i}_{L}} & \overline{d^{i}}_{L}\end{bmatrix}\begin{bmatrix}\frac{v+h_0}{\sqrt{2}}\\ 0\end{bmatrix}u^{j}_{R} - y^{u}_{ij}\overline{u^{j}_{R}}\begin{bmatrix}\frac{v+h_0}{\sqrt{2}} & 0\end{bmatrix}\begin{bmatrix}u^{i}_{L}\\ d^{i}_{L}\end{bmatrix} \qquad (2.43)\\ =& -\frac{y^{d}_{ij}v}{\sqrt{2}}\overline{d^{i}_{L}}d^{j}_{R} - \frac{y^{d}_{ij}v}{\sqrt{2}}\overline{d^{i}_{R}}d^{j}_{L} - \frac{y^{u}_{ij}v}{\sqrt{2}}\overline{u^{i}_{L}}u^{j}_{R} - \frac{y^{u}_{ij}v}{\sqrt{2}}\overline{u^{i}_{R}}u^{j}_{L}\\ & + h_0 \text{ terms.} \qquad (2.44)\end{aligned}$$

Note here the two indices (i, j) on the couplings (y) and quarks imply summation over the three generations. Therefore rewriting this in matrix notation along with the similarly derived lepton Yukawa term yields

$$\mathcal{L}^{q}_{Yukawa} = -\frac{v}{\sqrt{2}}\left(\overline{\mathbf{d}_{L}}\mathbf{Y}^{d}\mathbf{d}_{R} + \overline{\mathbf{d}_{R}}\mathbf{Y}^{d}\mathbf{d}_{L} + \overline{\mathbf{u}_{L}}\mathbf{Y}^{u}\mathbf{u}_{R} + \overline{\mathbf{u}_{R}}\mathbf{Y}^{u}\mathbf{u}_{L}\right) + h_0 \text{ terms}, \tag{2.45}$$

$$\mathcal{L}^{l}_{Yukawa} = -\frac{v}{\sqrt{2}}\left(\overline{\mathbf{l}_{L}}\mathbf{Y}^{l}\mathbf{l}_{R} + \overline{\mathbf{l}_{R}}\mathbf{Y}^{l}\mathbf{l}_{L}\right) + h_0 \text{ terms}, \tag{2.46}$$

after defining

$$\mathbf{d}_{L} \equiv \begin{bmatrix}d_L\\ s_L\\ b_L\end{bmatrix}; \quad \mathbf{d}_{R} \equiv \begin{bmatrix}d_R\\ s_L\\ b_R\end{bmatrix}; \quad \mathbf{Y}^{d} \equiv \begin{bmatrix}y^{d}_{11} & y^{d}_{12} & y^{d}_{13}\\ y^{d}_{21} & y^{d}_{22} & y^{d}_{23}\\ y^{d}_{31} & y^{d}_{32} & y^{d}_{33}\end{bmatrix}; \tag{2.47}$$

$$\mathbf{u}_{L} \equiv \begin{bmatrix}u_L\\ c_L\\ t_L\end{bmatrix}; \quad \mathbf{u}_{R} \equiv \begin{bmatrix}u_R\\ c_R\\ t_R\end{bmatrix}; \quad \mathbf{Y}^{u} \equiv \begin{bmatrix}y^{u}_{11} & y^{u}_{12} & y^{u}_{13}\\ y^{u}_{21} & y^{u}_{22} & y^{u}_{23}\\ y^{u}_{31} & y^{u}_{32} & y^{u}_{33}\end{bmatrix}; \tag{2.48}$$

$$\mathbf{l}_L \equiv \begin{bmatrix} e_L \\ \mu_L \\ \tau_L \end{bmatrix}; \quad \mathbf{l}_R \equiv \begin{bmatrix} e_R \\ \mu_R \\ \tau_R \end{bmatrix}; \quad \mathbf{Y}^u \equiv \begin{bmatrix} y^l_{11} & y^l_{12} & y^l_{13} \\ y^l_{21} & y^l_{22} & y^l_{23} \\ y^l_{31} & y^l_{32} & y^l_{33} \end{bmatrix}. \tag{2.49}$$

Using biunitary transformations of the form[6]:

$$\begin{aligned} &\mathbf{L}_L \to U^l_L \mathbf{L}_L; \quad \mathbf{Q}_L \to U^u_L \mathbf{Q}_L; \\ &\mathbf{l}_R \to U^l_R \mathbf{l}_R; \quad \mathbf{d}_R \to U^d_R \mathbf{d}_R; \quad \mathbf{d}_R \to U^u_R \mathbf{u}_R, \end{aligned} \tag{2.50}$$

the Yukawa matrices can be diagonalised,

$$\mathbf{Y}^d_{diag.} = U^{d\dagger}_L \mathbf{Y}^d U^d_R = diag(y_d, y_s, y_b), \tag{2.51}$$

$$\mathbf{Y}^u_{diag.} = U^{u\dagger}_L \mathbf{Y}^u U^u_R = diag(y_u, y_c, y_t), \tag{2.52}$$

$$\mathbf{Y}^l_{diag.} = U^{l\dagger}_L \mathbf{Y}^l U^l_R = diag(y_e, y_\mu, y_\tau), \tag{2.53}$$

where $U^{l,u,d}_{L,R}$ are five 3×3 unitary matrices which leave the Fermion interaction Lagrangian (Eq. 2.12) invariant.

Notice here that since left-handed quark and lepton fields are embedded in doublets $\mathbf{Q}_L$ and $\mathbf{L}_L$, they need to be transformed with the same transformation matrix. Here, this is chosen to be U^u_L and U^l_L. This means that U^d_L, required to diagonalize $\mathbf{Y}^d$ in Eq. 2.51, is missing. Alternatively, choosing U^d_L to transform $\mathbf{Q}_L$ would mean $\mathbf{Y}^u$ cannot be diagonalized. It can thus be concluded that both $\mathbf{Y}^d$ and $\mathbf{Y}^u$ cannot be simultaneously diagonalized with such transformations. In the absence of right-handed neutrinos in the SM this problem is not present for the leptons.

After performing these transformations, the Yukawa Lagrangian can be written as

$$\mathcal{L}^q_{Yukawa} = -\frac{v}{\sqrt{2}} \left(\overline{\mathbf{d}_L} U^{u\dagger}_L U^d_L Y^d_{diag} \mathbf{d}_R + \overline{\mathbf{u}}_L \mathbf{Y}^u_{diag} \mathbf{u}_R \right) + h_0 \text{ terms } + h.c., \tag{2.54}$$

$$\mathcal{L}^l_{Yukawa} = -\frac{v}{\sqrt{2}} \left(\bar{\mathbf{l}}_L \mathbf{Y}^l_{diag} \mathbf{l}_R \right) + h_0 \text{ terms } + h.c., \tag{2.55}$$

where the first term in $\mathcal{L}^q_{Yukawa}$ is non-diagonal. An additional transformation of the down quarks from the flavour eigenstates is necessary to diagonalize it,

$$\mathbf{d}_L = U^{u\dagger}_L U^d_L \mathbf{d'}_L. \tag{2.56}$$

Here $\mathbf{d'}_L$ are the mass eigenstates. This transformation shows the mismatch between the mass and weak eigenstates. It also modifies the Fermion interaction Lagrangian and is responsible for flavour transitions in the SM. The matrix $U^{u\dagger}_L U^d_L$ is the CKM matrix V_{CKM} named after Cabbibo, Kobayashi and Maskawa [17, 18]. The strength of flavour transitions is determined by the elements of V_{CKM} and it is unitary by construction. The relation between the weak and mass eigenstates in the most common notation can be written as

[6]Defining $\mathbf{L}_L = [L^1_L, L^2_L, L^3_L]$ and $\mathbf{Q}_L = [Q^1_L, Q^2_L, Q^3_L]$ from Eq. 2.6 and 2.9.

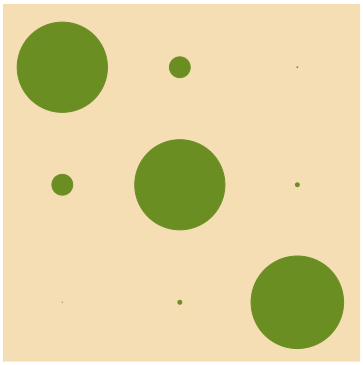

Figure 2.2: A visualization of the CKM matrix parameters. The radius of the circles are proportional to the magnitude of their respective CKM matrix elements. The hierarchical structure caused by the near diagonal matrix elements can be seen.

$$\mathbf{d}_L = V_{CKM}\mathbf{d'}_L = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d'_L \\ s'_L \\ b'_L \end{pmatrix} \tag{2.57}$$

Inserting the vev into the diagonalized Yukawa coupling matrices yields the mass matrices:

$$\begin{aligned} diag(m_d, m_s, m_b) &= \frac{v}{\sqrt{2}} diag(y_d, y_s, y_b) \\ diag(m_u, m_c, m_t) &= \frac{v}{\sqrt{2}} diag(y_u, y_c, y_t) \\ diag(m_e, m_\mu, m_\tau) &= \frac{v}{\sqrt{2}} diag(y_e, y_\mu, y_\tau) \end{aligned} \tag{2.58}$$

The CKM matrix is a complex 3×3 unitary matrix and therefore has 9 degrees of freedom. Of this, 5 of the phases can be absorbed by the quark fields leaving one phase and three angles as free parameters. The structure of the CKM matrix is visualized in Figure 2.2. A common parameterization of this matrix is the Wolfenstein parameterization [19] and is expanded around the parameter $\lambda \approx 0.22$,

$$\begin{bmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{bmatrix} + \mathcal{O}(\lambda^4), \tag{2.59}$$

with parameters A, ρ, η having values between 0 and 1. The complex elements are the source of the CP asymmetry in the SM. A huge experimental effort is dedicated to measuring the values of the CKM matrix and the latest results are summarized in [20, 21].

2.1.4.1 The Flavour puzzle

This mechanism of mass generation is validated by the observation of the Higgs Boson [22] and measurement of the Yukawa couplings. However, the mass of the Fermions and the

4 independent parameters of the CKM matrix are still free. This leaves some unanswered questions such as: why do the mass of the Fermions span a wide range (a few MeV/c^2 to $\sim 100\ \text{GeV}/c^2$); what is the cause for hierarchical structure of the CKM matrix? why only three generations exists? Any observed inconsistencies between precision measurements and SM predictions would point towards BSM physics and help answer these questions.

2.1.5 Interaction of gauge Bosons with Fermions in the SM

The couplings of quarks and leptons to the gauge Bosons arise from the covariant derivatives in the Fermion interaction Lagrangian $\mathcal{L}_{Fermion}$. After spontaneous symmetry breaking and the introduction of $W^{\pm}$, A_μ and Z_μ the covariant derivatives for quark doublets (Eq. 2.16) can be rewritten as

$$\begin{aligned} D_\mu \mathbf{Q}_L = \Bigg[& \partial_\mu + i\frac{g_2}{\sqrt{2}}\left(W_\mu^+ \frac{\sigma^+}{2} + W_\mu^- \frac{\sigma^-}{2}\right) && \text{(charged current interactions)} \\ & + i\sqrt{g_1^2 + g_2^2}\left(\frac{\sigma^3}{\sqrt{2}} - \sin^2\theta_w \mathbb{Q}\right) Z_\mu + ie\mathbb{Q}A_\mu && \text{(neutral current interactions)} \\ & + ig_s G_\mu^a \frac{\lambda^a}{2}\Bigg] \mathbf{Q}_L, && \text{(strong interactions)} \end{aligned} \tag{2.60}$$

The covariant derivative for leptons $\mathbf{L}_L$ follows the same pattern but without the strong interaction term. The following definitions have also been introduced:

$$Y = -1/2 \qquad \text{(hypercharge assignment for } \mathbf{L}_\text{L}) \tag{2.61}$$

$$Y = 1/6 \qquad \text{(hypercharge assignment for } \mathbf{Q}_\text{L}) \tag{2.62}$$

$$\mathbb{Q} = T^3 + Y = \frac{\sigma^3}{2} + Y = \begin{bmatrix} \frac{1}{2} + Y & 0 \\ 0 & Y - \frac{1}{2} \end{bmatrix}, \qquad \text{(the electric charge generator)} \tag{2.63}$$

$$e \equiv g_2 sin\theta_w, \qquad \text{(the electric coupling strength)} \tag{2.64}$$

$$\sigma^\pm = \sigma^1 \pm i\sigma^2 = \begin{bmatrix} 0 & 1 \pm 1 \\ 1 \mp 1 & 0 \end{bmatrix}. \tag{2.65}$$

Expanding the interaction term $i\mathbf{Q}_L \gamma^\mu D_\mu \mathbf{Q}_L$ and focusing on the coupling to the charged Bosons gives[7]

$$i\overline{\mathbf{Q}_L}\gamma^\mu \frac{ig_2}{\sqrt{2}} W_\mu^+ \frac{\sigma^+}{2} \mathbf{Q}_L + h.c. \tag{2.66}$$

$$= \frac{-g_2}{\sqrt{2}} W_\mu^+ \begin{bmatrix} \overline{u_L^i} & \overline{d_L^i} \end{bmatrix} \begin{bmatrix} 0 & \gamma^\mu \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_L^j \\ d_L^j \end{bmatrix} + h.c. \tag{2.67}$$

$$= \frac{-g_2}{\sqrt{2}} W_\mu^+ \begin{bmatrix} \overline{u_L^i} & \overline{d_L^i} \end{bmatrix} \begin{bmatrix} \gamma^\mu d_L^j \\ 0 \end{bmatrix} + h.c., \tag{2.68}$$

[7] The charged current interaction in the lepton sector is ignored as it is not of relevance to this work.

with $\mathbf{Q}_L$ defined in the weak eigenbasis and written in element notation. Imposing the transformations Eq. 2.50 and using the CKM matrix to convert to the mass eigenbasis implies:

$$Q_L^i = \begin{bmatrix} u_L^i \\ d_L^i \end{bmatrix} = \begin{bmatrix} (U_L^u)^{im} u'_{Lm} \\ (U_L^u)_{im}(V_{CKM})_{mk} d'^k_L \end{bmatrix} = (U_L^u)^{im} \begin{bmatrix} u'_{Lm} \\ (V_{CKM})_{mk} d'^k_L \end{bmatrix}, \tag{2.69}$$

where the unitary condition of U_L^u was imposed. Eq. 2.68 then becomes,

$$\frac{-g_2}{\sqrt{2}} W_\mu^+ \begin{bmatrix} \overline{u'_{Lm}} & \overline{(V_{CKM})_{mk} d'^k_L} \end{bmatrix} (U_L^{u\dagger})^{im} (U_L^u)_{jp} \begin{bmatrix} \gamma^\mu (V_{CKM})_{pq} d'^q_L \\ 0 \end{bmatrix} + h.c. \tag{2.70}$$

$$= \frac{-g_2}{\sqrt{2}} W_\mu^+ \overline{u'_{Lm}} \gamma^\mu (V_{CKM})_{mq} d'^q_L, \tag{2.71}$$

where the unitary condition of U_L^u is again imposed. This describes flavour transition of a d-type quark to a u-type quark mediated by the charged W^+ boson. The strength of this transition is determined by the corresponding element in the V_{CKM} matrix. It is straightforward to show that the absence of W_μ term in the covariant derivative of right-handed quarks implies that $W^\pm$ couples only to the left-handed quarks. This means that parity is maximally violated in charge current interactions.

Next $i\mathbf{Q}_L \gamma^\mu D_\mu \mathbf{Q}_L$ can be expanded with a focus on the neutral current interactions:

$$i\mathbf{Q}_L \gamma^\mu i\sqrt{g_1^2 + g_2^2} \left(\frac{\sigma^3}{\sqrt{2}} - \sin^2\theta_w \mathbb{Q} \right) Z_\mu \mathbf{Q}_L + i\mathbf{Q}_L \gamma^\mu ie\mathbb{Q} A_\mu \mathbf{Q}_L \tag{2.72}$$

$$= -\sqrt{g_1^2 + g_2^2} Z_\mu \begin{bmatrix} \overline{u_L^i} & \overline{d_L^i} \end{bmatrix} \begin{bmatrix} (\frac{1}{2} - \frac{2}{3}\sin^2\theta_w)\gamma^\mu & 0 \\ 0 & (-\frac{1}{2} + \frac{1}{3}\sin^2\theta_w)\gamma^\mu \end{bmatrix} \begin{bmatrix} u_L^j \\ d_L^j \end{bmatrix} \tag{2.73}$$

$$- eA_\mu \begin{bmatrix} \overline{u_L^i} & \overline{d_L^i} \end{bmatrix} \begin{bmatrix} \frac{2}{3}\gamma_\mu & 0 \\ 0 & -\frac{1}{3}\gamma_\mu \end{bmatrix} \begin{bmatrix} u_L^j \\ d_L^j \end{bmatrix} \tag{2.74}$$

Again converting to the mass eigenbasis and simplifying yields:

$$= -\sqrt{g_1^2 + g_2^2} Z_\mu \left\{ \left(\frac{1}{2} - \frac{2}{3}\sin^2\theta_w \right) \overline{u'_{Lm}} \gamma^\mu u'^m_L + \left(-\frac{1}{2} + \frac{1}{3}\sin^2\theta_w \right) \overline{d'^k_L} \gamma^\mu \overline{(V_{CKM})_{mk}} (V_{CKM})_{mq} d'^q_L \right\}$$
$$- eA_\mu \frac{2}{3} \overline{u'_{Lm}} \gamma^\mu u'^m_L - \frac{1}{3} \overline{d'^k_L} \gamma^\mu \overline{(V_{CKM})_{mk}} (V_{CKM})_{mq} d'^q_L \tag{2.75}$$

$$= -\sqrt{g_1^2 + g_2^2} Z_\mu \left\{ \left(\frac{1}{2} - \frac{2}{3}\sin^2\theta_w \right) \overline{u'_{Lm}} \gamma^\mu u'^m_L + \left(-\frac{1}{2} + \frac{1}{3}\sin^2\theta_w \right) \overline{d'^q_L} \gamma^\mu d'^q_L \right\}$$
$$- eA_\mu \frac{2}{3} \overline{u'_{Lm}} \gamma^\mu u'^m_L - \frac{1}{3} \overline{d'^q_L} \gamma^\mu d'^q_L, \tag{2.76}$$

where the unitary property of the CKM matrix $\overline{(V_{CKM})_{mk}}(V_{CKM})_{mq} = \delta_{kq}$ was used.

This result implies that neutral currents interactions do not mix different quark flavours and that Flavour Changing Neutral Current (FCNC) transitions do not occur at the tree-level in the SM. This consequence arising from the unitarity of the CKM matrix is known as the Glashow-Iliopolus-Maiani (GIM) mechanism [23]. Through the exchange of a virtual W boson,

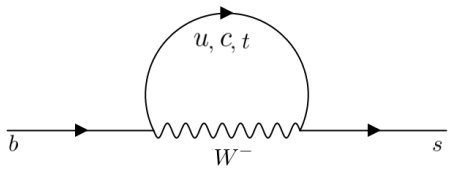

Figure 2.3: A picture of the $b \to s$ process in the SM.

FCNCs can however proceed at the loop level. The transition of $b \to s$ is illustrated in Figure 2.3. All three up-quark types appear and the amplitude is of the form:

$$\mathcal{A} = V_{ub}V_{us}^* f(\hat{m}_u^2) + V_{cb}V_{cs}^* f(\hat{m}_c^2) + V_{tb}V_{ts}^* f(\hat{m}_t^2) \quad \text{with } \hat{m}_q \equiv m_q/m_W; \tag{2.77}$$

where $f(\hat{m}_q)$ is an unspecified loop function. Due to the presence of an additional vertex compared to tree-level transitions, loop processes are suppressed by the factor $(g_2/4\pi)^2$. In addition, using the CKM unitarity relation:

$$V_{ub}V_{us}^* + V_{cb}V_{cs}^* + V_{tb}V_{ts}^* = 0, \tag{2.78}$$

the amplitude can be rewritten as,

$$\mathcal{A} = V_{tb}V_{ts}^* \left(f(\hat{m}_t^2) - f(\hat{m}_u^2)\right) + V_{ub}V_{us}^* \left(f(\hat{m}_c^2) - f(\hat{m}_u^2)\right). \tag{2.79}$$

The amplitude is suppressed by the small off-diagonal elements of the CKM matrix. Furthermore, in the presence of degenerate quark masses, this amplitude would be zero. Decays which are suppressed as such are known to be GIM suppressed. Since $m_t > m_W \gg m_{u,c}$, the contribution of the top quark dominate such transitions. Combining all of this together along with the small suppression from $V_{tb}V_{ts}^* \approx \lambda^2$ implies that $b \to s$ $\{l^+l^-, \gamma\}$ transitions in the SM are rare. Their SM branching fractions range from $\mathcal{O}(10^{-4})$ to $\mathcal{O}(10^{-9})$. This rare nature of $b \to s$ transitions makes them a good place to search for BSM physics as new interactions are not required to have the same structure. New Physics FCNC could happen at the tree-level through virtual BSM particles causing visible effects or interference with the SM contribution. If precision measurement of their observables show deviation from predictions it would be a clear sign of BSM physics.

2.1.6 Quantum Chromodynamics and its asymptotic freedom

In the discussion thus far, focus was given to weak interactions which proceed via the weak force. Even though QCD interactions are flavour blind they play a significant role in studying $b \to s\ell^+\ell^-$ transitions. This is because the quarks hadronise into quark-antiquark bound states called mesons (e.g. B^+: $[\bar{b}u]$) via the strong interaction.

The property of QCD of most relevance here is its asymptotic freedom [24, 25]. It means that strong interaction between particles gets asymptotically weaker in large energy scales

where the momentum transfer is high. This is equivalent to stating that the strong coupling is small when the distance between the particles is short. Physically what is happening is that the colour charge of a quark polarizes the vacuum creating virtual quark-antiquark pairs. These virtual particles align themselves such as to cancel out the effect of the quark field at increasing finite distances, this is analogous to the screening effect in QED. However in QCD virtual gluons loops are also created which do not screen the quark field but rather augment its color. This has the effect of making interactions weaker at short distances and is called anti-screening. The anti-screening effect is larger than the screening effect resulting in the asymptotic freedom of QCD.

The procedure of renormalization to counteract the ultraviolet divergence introduces energy scale dependent *running* coupling constant. For a given process with a momentum transfer of μ^2, the strong coupling $g_s^2 \equiv \frac{\alpha_s}{4\pi}$ is written as,

$$\alpha_s(\mu^2) = \frac{12\pi}{(11N - 2F)\log\left(\frac{\mu^2}{\Lambda^2_{QCD}}\right)}, \tag{2.80}$$

where N is the number of colour charges (3 in QCD) and F is number of quark flavours (6 from the three generations). The factor $\Lambda_{QCD} \approx 200\,\text{MeV}$ formally defines the scale at which the coupling diverges.

The other implication of the asymptotic freedom is that strong interaction effects can be calculated using perturbation theory at short-distances when α_s is small. However, the use of perturbation theory at long-distance becomes very tricky and alternative strategies are usually employed. These calculations are the leading source of theoretical uncertainties in the study of $b \to s\ell^+\ell^-$ transitions.

2.2 An Effective Field Theory to describe $b \to s\ell^+\ell^-$ transitions.

An EFT allows one to search for BSM physics in a model independent way. The use of EFTs are very common in particle physics and they work by exploiting the different energy scales involved in a process to describe the dynamics of a system in terms of a set of local operators $\mathcal{O}_i$. The couplings of these operators are determined by the scalar quantities C_i known as Wilson coefficients. This separation of energy scale is especially useful when perturbative QCD calculations are difficult.

For $b \to s\ell^+\ell^-$ transitions Heavy Quark Effective Theory (HQET) [26] is used where the expansion scale is chosen to be $\mu = m_b \approx 5\,\text{GeV}/c^2$. It describes the dynamics of hadrons with a single heavy b quark. The low energy effects are absorbed by the local operators which have a generic Lorentz structure. This is a good description of low energy physics which includes: u,d,s,c quarks; photons; and the non-perturbative QCD effects happening at the Λ_{QCD} energy scale.

The interesting high energy weak decay which happens at the m_W scale is split from the low energy effects [27, 28]. The heavy physics are then explicitly integrated out and are absorbed by the Wilson coefficients which describe the strength of the operators at a given scale μ. Within the SM this would include the interactions from the $W^{\pm}$, Z Bosons and the t quark.

The thus constructed effective Hamiltonian is a sum of these local operators along with their Wilson coefficient. Using this, the amplitude of a generic decay process, $i \to f$ can be written as,

$$\mathcal{A}(i \to f) = \langle f|\mathcal{H}_{eff}|i\rangle = \frac{G_F}{\sqrt{2}} \sum_j V^j_{CKM} C_j(\mu)\ \langle f|\mathcal{O}_j(\mu)|i\rangle, \tag{2.81}$$

where G_F is the Fermi constant and V^j_{CKM} is the relevant CKM matrix factor and can be different for different operators. This Hamiltonian is then systematically expanded at some energy scale through a procedure known as Operator Product Expansion (OPE). More intuitively it can be visualized as the effective operators $\mathcal{O}_j$ describing the physics of effective vertices with coupling constants C_i . The high energy short-distance physics is thus encoded in C_i , while $\mathcal{O}_j$ contains the low energy long-distance effects. An example of this is the Fermi theory of beta decay before the discovery of $W^{\pm}$ Bosons [29]. Here the process is described using a four point vertex operator whose coupling constant is $G_F/\sqrt{2}$, this constant describes the now known high energy physics of the $W^{\pm}$ Bosons.

At this stage, the EFT does not have any predictive power, and to gain it, the a priori unknown Wilson coefficients and the hadronic matrix elements corresponding to the operators need to be calculated. This is done through a procedure called 'matching' by comparing the effective theory with the full SM theory in its perturbative regime [30]. In the presence of heavy NP, the observed decay process will be modulated. By treating the Wilson coefficients as generalized couplings, this can be studied by comparing their SM and experimentally measured value. The measured Wilson coefficients will include the SM value and be sensitive to a broad set of NP particles in a model independent way, $C_j^{exp} = C_j^{SM} + C_j^{NP}$. As an example, the presence of a heavy scalar NP particle will change the Wilson coefficient corresponding to the operator which describes a generic scalar interaction. Observation of deviations to the SM value will allow the inference of NP.

2.2.1 The effective Hamiltonian

The effective Hamiltonian for $b \to s\ell^+\ell^-$ transitions in the SM is given by [31–34],

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb} V^*_{ts} \sum_i C_i(\mu)\mathcal{O}_i(\mu) + h.c.. \tag{2.82}$$

The dimension-6 operators of most relevance to the current discussion are,

$$\mathcal{O}_7 = \frac{e^2}{16\pi^2}\overline{m_b}(\bar{s}\sigma_{\mu\nu}P_R b)F^{\mu\nu}; \quad \mathcal{O}_9 = \frac{e^2}{16\pi^2}\overline{m_b}(\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma^\mu l); \quad \mathcal{O}_{10} = \frac{e^2}{16\pi^2}\overline{m_b}(\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma^\mu\gamma^5 l); \tag{2.83}$$

Here $P_{L,R}$ are the chiral projection operators and $\overline{m_b}$ is the running mass of the b-quark[8] and e is the electromagnetic coupling constant. The operator $\mathcal{O}_7$ is called the tensor current operator while $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are the vector and axial vector current operators respectively.

The gluonic-QCD penguin operators, $\mathcal{O}_{3..6}$ and $\mathcal{O}_8$, describe the electroweak radiative gluon process, they are explicitly given in Ref. [35]. Along with the current-current operators $\mathcal{O}_{1q,2q}$ with $q = u, c$ these operators describe processes where intermediate $q\overline{q}$ pairs are created. They are responsible for the non-local effects [36] and will be discussed in Section 2.3.2. For reference $\mathcal{O}_{1q,2q}$ are given below and the process schematically shown in Figure 2.4 as they have the largest contribution:

$$\mathcal{O}_{1q} = \frac{V_{qb}V_{qs}^*}{V_{tb}V_{ts}^*}(\bar{s}\gamma_\mu T^a P_L q)(\bar{q}\gamma^\mu T^a P_L b); \quad \mathcal{O}_{2q} = \frac{V_{qb}V_{qs}^*}{V_{tb}V_{ts}^*}(\bar{s}\gamma_\mu P_L q)(\bar{q}\gamma^\mu P_L b) \quad \text{for } q \in \{u, c\}, \quad (2.84)$$

where T^a is the generator of $SU(3)_C$.

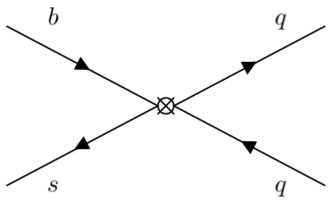

Figure 2.4: An illustration of the process described by the operators $\mathcal{O}_{1q,2q}$ where q is a u- or c-quark.

The effective theory is matched with the full theory at scale $\mu = m_W$ and evolved down to scale $\mu = m_b$. Through this evolution the high energy short-distance physics is transferred from the hadronic matrix elements of the operators to the Wilson coefficients [37]. This has the effect of mixing different operators and it is therefore convenient to reassemble Wilson coefficients. A detailed account of this procedure can be found in Ref. [35]. As a result of the mixing, effective Wilson coefficients can be written and appear in physical amplitudes as [37, 38],

$$C_7^{\text{eff}} = C_7 - \frac{1}{3}\left[C_3 + \frac{4}{3}C_4 + 20C_5 + \frac{80}{3}C_6\right] + O(\alpha_s); \quad (2.85)$$

$$C_9^{\text{eff}} = C_9 + Y(q^2); \quad (2.86)$$

where $Y(q^2)$ is a function which depends on the momentum transfer between the b-hadron and s-hadron in $B \to K$ decays. It accounts for all the non-local long distance effects of a quark loop generated by $\mathcal{O}_{1,2..6,8}$ which couple to a lepton pair via a virtual photon. In other words, it describes the intermediate resonance states like $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \to K^+\psi(2S)(\mu^+\mu^-)$.

[8]The running mass is a feature of renormalizable theories and encodes the energy scale dependence of the mass, the reader is referred to Section 4.3.4 in Ref.[13] for details on this.

Wilson coefficient	SM Value
C_7^{eff}	-0.304
C_9	4.211
C_{10}	-4.103

Table 2.1: The SM values of the Wilson coefficients calculated up to next-next leading-logarithm accuracy at $\mu = m_b = 4.8\,\text{GeV}/c^2$ from Ref. [39]. The chirality flipped counterparts $C'_{7,9,10}$ are negligible in the SM.

Figure 2.5 shows Feynman diagrams of such intermediate states along with an example of a direct $B^+ \to K^+\mu^+\mu^-$ decay. Accounting for the interference of such effects with $B^+ \to K^+\mu^+\mu^-$ decays is at the core of this analysis, a more detailed discussion on this topic is deferred to Section 2.3.2.

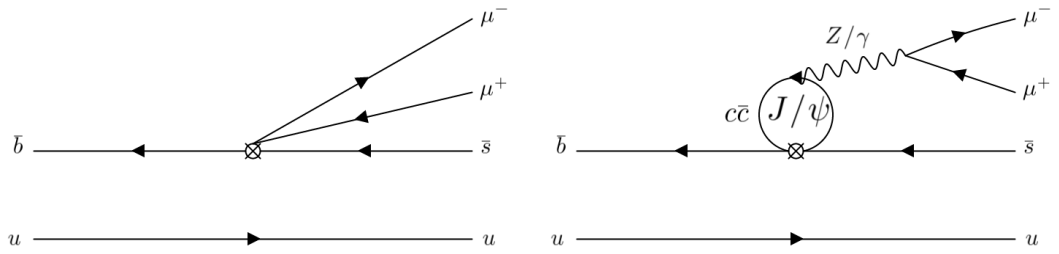

Figure 2.5: An exemplary Feynman diagrams of $B^+ \to K^+\mu^+\mu^-$ decay (left) described by $\mathcal{O}_{7,9,10}$ and $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ decay (right) described largely by $\mathcal{O}_{1,2}$ in the Effective Field Theory.

The SM value for C_7^{eff} and $C_{9,10}$ up to next-next leading-logarithm accuracy was calculated in Ref. [39] and is given in Table 2.1. Chirality flipped operators denoted by the prime $\mathcal{O}'_{7,9,10}$ with Wilson coefficients $C'_{7,9,10}$ can also be written through the exchange of $P_L \leftrightarrow P_R$ to the unprimed operators. Within the SM, the primed operators vanish or are highly suppressed.

The measurement presented in this thesis is sensitive to $C_9 + C'_9$ and $C_{10} + C'_{10}$ and the results are quoted as C_9 and C_{10} respectively [34]. Any deviations to the calculated SM value indicates at influence from NP process of either chirality. Possible interpretations of NP to either chiral partners is beyond the scope of this thesis. In addition, Wilson coefficients are assumed to be real valued. The existence of complex Wilson coefficients implies Charge Parity (CP) violation and this measurement is insensitive to them as it is performed using a CP averaged data set. An extension to study any CP violating phases which could be present is left for a future work.

In principle, BSM physics contributing to $b \to s\ell^+\ell^-$ transitions could in principle also be of (pseudo-)scalar and (pseudo-)tensor nature. The most general operators which account for this are $\mathcal{O}_{S,P,T,T5}$ and their chiral counterparts, their precise definitions are given in Ref. [34]. In

this analysis these operators are safe to be neglected. This is because the global fits discussed in Section 2.6 do not favour these contributions and indicates any NPis likely to be in $C_{9,10}^{(\prime)}$. Moreover, branching fraction measurements of $B_s^0 \to \mu^+\mu^-$ put stringent constraints on the scalar operators. The tensor operators can also be neglected since they are not generated at the dimension-6 level [40, 41].

2.3 The Hadronic Contributions

The b and s quarks exist in a bound state with the spectator u quark, as such this induces hadronic contributions from gluon exchanges. The energy scale of these interactions are at the $\Lambda_{QCD} \ll m_b$ scale and form part of the local long-distance effects. Such contributions are parameterized through so-called form factors and are discussed in Section 2.3.1. The other type of long-distance effect comes from the production of intermediate quark/meson loops which scatter to a lepton pair. These are the non-local contributions, the analysis presented in this thesis determines these contributions through the data-driven method discussed in Section 2.3.2.

2.3.1 The local hadronic form factors

Once the short distance physics has been transferred to the Wilson coefficients, the leftover hadronic matrix elements encode the low-energy long-distance physics. The calculation of these local matrix elements is complicated since the quarks exist in a bound mesonic state. This means that they would need to account for gluon exchanges as shown by the exemplary Feynman diagram in Figure 2.6. Such hadronic quantities involve processes which are typically at scales of the order Λ_{QCD} and are difficult to calculate in perturbative QCD[9] [30].

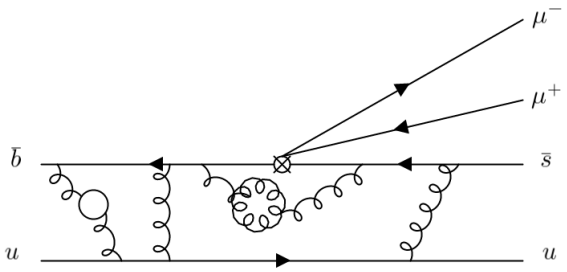

Figure 2.6: An exemplary Feynman diagram $B^+ \to K^+\mu^+\mu^-$ decay with the non-perturbative QCD effects.

For $B \to K$ decays the dominant matrix elements arise from $\mathcal{O}_{7,9,10}$ at lowest order [42]. The matrix elements of the tensor and vector current operators can be written in terms of three

[9] This follows from the fact that QCD is asymptotically free only at short-distances ($\mu \gtrsim m_b$).

	$i=0$	$i=1$	$i=2$
b_i^0	0.466 ± 0.014	-0.885 ± 0.128	-0.213 ± 0.548
b_i^0	0.292 ± 0.010	0.281 ± 0.125	0.150 ± 0.441
b_i^T	0.460 ± 0.019	-1.089 ± 0.236	-1.114 ± 0.971

Table 2.2: The coefficients of the z-expansion parameterization used to describe the form factors $f_+(q^2)$, $f_0(q^2)$ and $f_T(q^2)$ calculated using LatticeQCD techniques in Ref. [44].

non-vanishing form factors $f_{T,+,0}(q^2)$ as given in [43, 44] and first described in [45],

$$\langle K(k)|\bar{s}\sigma_{\mu\nu}b|B(p)\rangle = i\frac{f_T(q^2)}{M_B+M_K}\left[(p+k)_\mu q_\nu - q_\mu(p+k)_\nu\right], \tag{2.87}$$

$$\langle K(k)|\bar{s}\gamma_\mu b|B(p)\rangle = f_+(q^2)(p+k)_\mu + \left[f_0(q^2)-f_+(q^2)\right]\frac{M_B^2-M_K^2}{q^2}q_\mu, \tag{2.88}$$

where p and k are the four momenta of the mesons with μ,ν space-time indices and $M_{B,K}$ are the mass of their respective mesons. In the study of $b \to s\ell^+\ell^-$ transitions, it is common convention to use the kinematic variable q^2 which is defined as $q^2 \equiv (p-k)^2 \equiv m_{ll}^2$, where m_{ll}^2 is the square of the invariant mass of the dilepton system.

Techniques such as LatticeQCD [44] and Light Cone Sum Rule (LCSR) [46] are used to calculate the form factors. LatticeQCD calculations are most precise in the low hadronic recoil regime ($q^2 \gtrsim 17\,\mathrm{GeV}^2/c^4$) and are performed there. On the other hand, LCSR calculations are performed in the unphysical negative q^2 region where it is most precise. The model independent z-expansion method is then used to extrapolate to the full physical range of $4m_l^2 \le q^2 \le (m_B - m_K)^2$ [47]. In this work, the form factors are parameterized as [48]

$$f_{+,T}(q^2) = \frac{1}{1-q^2/M_{B_s^*}}\sum_{i=0}^{N-1} b_i^{+,T}\left[z^i - (-1)^{i-N}\left(\frac{i}{N}\right)z^N\right], \tag{2.89}$$

$$f_0(q^2) = \frac{1}{1-q^2/M_{B_{s0}}^2}\sum_{i=0}^{N-1} b_i^0 z^i \tag{2.90}$$

with the z variable defined as

$$z(q^2) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}}, \tag{2.91}$$

with $t_\pm \equiv (M_B \pm M_K)^2$ and $t_0 \equiv (M_B + M_K)(\sqrt{M_B} - \sqrt{M_K})^2$. The mass $M_{B_s^*} = 5415.4\,\mathrm{MeV}/c^2$ and $M_{B_{s0}} = 5711\,\mathrm{MeV}/c^2$ are the masses of the lowest lying excited B_s^0 meson states. The quantities extracted from Ref. [44] are the coefficients $b_i^{0,+,T}$ listed in Table 2.2 where $N = 3$ was chosen as the order of expansion. The correlations between the different coefficients are also provided which are used in this analysis. Figure 2.7 shows the form factors as a function of q^2 compared with other LatticeQCD and LCSR calculations.

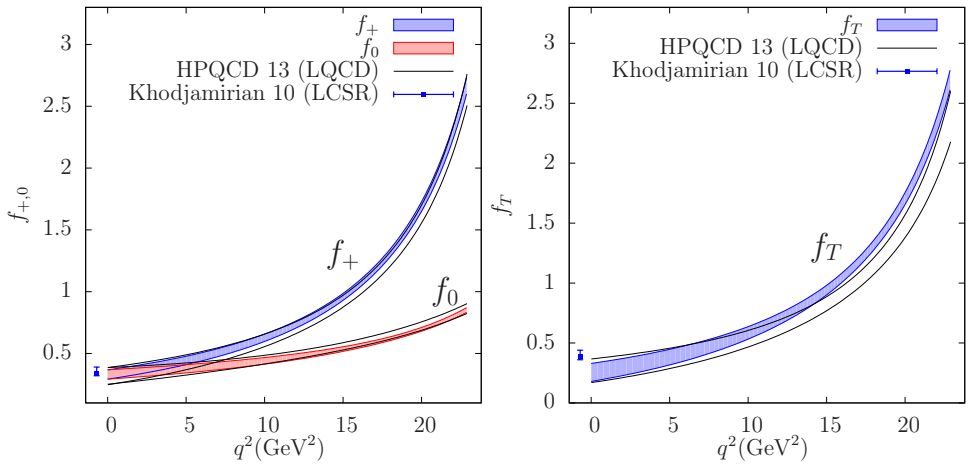

Figure 2.7: The form factors $f_{+,0,T}$ as calculated by Ref. [44] is given in the solid colour bands. The HPQCD and LCSR result are from Ref. [49] and Ref. [46] respectively. Plots taken from Ref. [44].

2.3.2 Non-local hadronic contributions

The non-local hadronic contributions are the intermediate states such as $B^+ \to K^+ J/\psi(\mu^+\mu^-)$. They arise from intermediate quark loops generated by the 4-quark operators $\mathcal{O}_{1,2}$ and to a lesser extent by the QCD penguin operators $\mathcal{O}_{3..6,8}$. The electromagnetic interaction of these quark loops produce the lepton pair in the final state. This process is known as non-local as the lepton pair production is detached from the B meson interaction.

The $J^{PC} = 1^{--}$ states are the amplitudes which interfere with the vector part of the rare decay. Therefore the only light valance quark loops which need to be included are $\rho(770)$, $\omega(782)$ and $\phi(1020)$. In addition the charmonium vector states ψ with $m_\psi < (M_B - M_K)$ which contribute includes the resonances J/ψ, $\psi(2S)$, $\psi(3770)$, $\psi(4040)$, $\psi(4160)$ and $\psi(4415)$. Here the J/ψ and $\psi(2S)$ resonances dominant. The quark loop hadronic contributions are referred to as one-particle (1P) intermediate states. A continuum of charmed hadron states (DD, D^*D^*, DD^*) can also contribute to the $B^+ \to K^+\ell^+\ell^-$ spectrum. They are referred to as intermediate two-particle (2P) states. The 2P contribution from light quarks are neglected as they are heavily loop- or CKM suppressed[10].

Within the EFT the leading order Feynman diagram for the 1P and 2P states are shown in Figure 2.8. The next-to-leading order corrections from hard-gluon exchanges between the loop and the other on-shell quarks and soft gluon emissions between the loop and vacuum as shown in Figure 2.9 are also expected. It is these corrections which make calculations of such loops difficult [36, 46].

By comparing the amplitude of these processes to the vector current operator $\mathcal{O}_9$ it is

[10] In principle there is also a contribution from three or more particles of both charmonium and light states but they are heavily phase-space suppressed and therefore neglected.

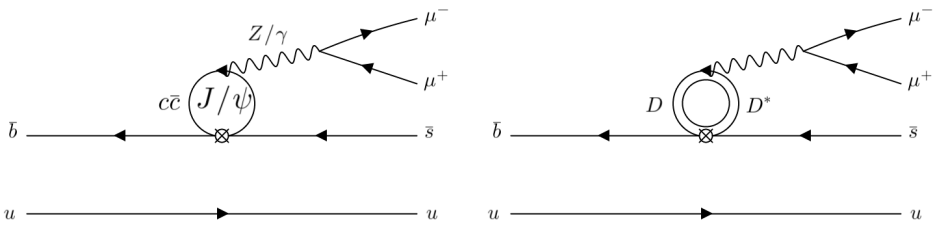

Figure 2.8: The one-particle (left) and two-particle (right) non-local contributions to the $B^+ \to K^+\mu^+\mu^-$ amplitude. Here the J/ψ and DD^* are used as an example.

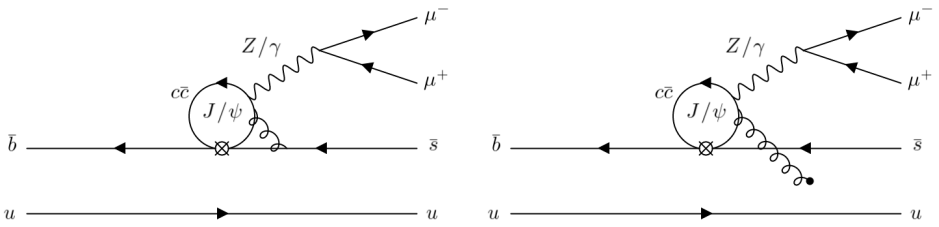

Figure 2.9: The hard (left) and soft (right) gluon emission hadronic effects affecting the exemplary intermediate charmonium state J/ψ. The gluon exchanges between the b and u quarks illustrate the local hadronic effects.

convenient to redefine these contributions as q^2 dependent corrections to the Wilson coefficient $C_{9.}$. Following Refs. [50], C_9^{eff} can be defined as,

$$C_9^{\text{eff}}(q^2) = C_9 + Y_{\text{light}}(q^2) + Y_{c\bar{c}}(q^2), \tag{2.92}$$

where the $Y_{\text{light}}(q^2)$ and $Y_{c\bar{c}}(q^2)$ encodes the non-perturbative hadronic contributions and are related to the hadronic transition matrix element $\mathcal{H}_I^{B\to K}$ (I=light, $c\bar{c}$) through,

$$Y_I(q^2) = \frac{16\pi^2}{f_+(q^2)}\mathcal{H}_I^{B\to K}(q^2). \tag{2.93}$$

Here $f_+(q^2)$ is the form factor introduced in the previous section. $\mathcal{H}_I^{B\to K}$ accounts for both the hard- and soft- gluon emission from the loops.

In the scattering amplitude of $q\overline{q} \to \mu^+\mu^-$, the imaginary parts describe the propagation of on-shell intermediate states. The requirement of causality gives rise to an analytic structure which allows one to relate the real and imaginary parts of $q\overline{q} \to \mu^+\mu^-$ using dispersion relations. The non-perturbative functions Y_I are written using hadronic dispersion relations to describe their q^2 dependence. In the following sections the charmonium and light states are individually considered.

2.3.2.1 The charmonium states

The subtracted form of the dispersion relation is used to describe the charmonium states. The $Y_{c\bar{c}}(q^2)$ contribution is subtracted at $q^2 = q_0^2$ and decomposed into the 1P and 2P states [50].

$$Y_{c\bar{c}}(q^2) = Y_{c\bar{c}}^{q_0^2} + \Delta Y_{c\bar{c}}^{1P}(q^2) + \Delta Y_{c\bar{c}}^{2P}(q^2), \tag{2.94}$$

with $\Delta Y_{c\bar{c}}^{1P}(q_0^2) = \Delta Y_{c\bar{c}}^{2P}(q_0^2) = 0$. The subtraction is done in order to ensure the convergence of the dispersion relation at large q^2. The subtracted dispersion relations can be written as

$$\Delta Y_{c\bar{c}}(q^2) = \frac{(q^2 - q_0^2)}{\pi} \int_{q^2_{min}}^{\infty} ds \frac{s \cdot Im\left[Y_{c\bar{c}}(s)\right]}{(s - q_0^2)(s - q^2 - i\epsilon)} \tag{2.95}$$

$$\equiv \frac{(q^2 - q_0^2)}{\pi} \int_{q^2_{min}}^{\infty} ds \frac{\rho_{c\bar{c}}(s)}{(s - q_0^2)(s - q^2 - i\epsilon)}, \tag{2.96}$$

where $\rho_{c\bar{c}}(s) = \rho_{c\bar{c}}^{1P}(s) + \rho_{c\bar{c}}^{2P}(s)$ is called the spectral density function describing the 1P and 2P hadronic states characterized by s, their invariant mass squared. They are the intermediate states $B \to Kc\bar{c} \to K\mu^+\mu^-$ and $B \to KM_{j1}M_{j2} \to K\mu^+\mu^-$ which enter the $B^+ \to K^+\mu^+\mu^-$ amplitude. The value for q^2_{min} does not matter as long as it is below J/ψ, the first resonance. The 1P spectral density is given by,

$$\rho_{c\bar{c}}^{1P}(s) \propto \sum_{j = J/\psi \ldots \psi(4415)} \mathcal{A}(B^+ \to K^+ \psi_j)\delta(s - m_j) \tag{2.97}$$

where $\mathcal{A}$ is the amplitude, m_j is the mass of the 1P states which are treated as infinitely narrow. A Breit-Wigner function can be used to describe the finite widths which results in

$$\Delta Y_{c\bar{c}}^{1P}(q^2) = \sum_{j = J/\psi \ldots \psi(4415)} \eta_j e^{i\delta_j} \frac{(q^2 - q_0^2)}{(m_j^2 - q_0^2)} A_j^{res}(q^2), \quad A_j^{res}(q^2) = \frac{m_j \Gamma_{0j}}{(m_j^2 - q^2) - i m_j \Gamma_j(q^2)}, \tag{2.98}$$

here η_j is the magnitude of the resonant contributions and δ_j is the phase difference with the electroweak $B^+ \to K^+\mu^+\mu^-$ decays. Determining these quantities from data is one of the main results of this thesis. The variables m_j and Γ_{0j} are the resonance's physical mass and width respectively. It is worth highlighting here that the factor $\frac{(q^2 - q_0^2)}{(m_j^2 - q_0^2)}$ arises due to the subtraction of the dispersion relation. The factor $\Gamma_j(q^2)$ is the running width of the resonances given by,

$$\Gamma_j(q^2) = \frac{p}{p_{0j}} \frac{m_j}{\sqrt{q^2}} \Gamma_{0j} \qquad \text{for j} \neq \psi(3770), \tag{2.99}$$

$$\Gamma_j(q^2) = \frac{p}{p_{0j}} \frac{m_j}{\sqrt{q^2}} \Gamma_{1\psi(3770)} + \frac{q_D}{q_{0D}} \frac{m_j}{\sqrt{q^2}} \Gamma_{1\psi(3770)} \qquad \text{for j} = \psi(3770), \tag{2.100}$$

where p is the momentum of the muons in the rest frame of the dimuon system evaluated at $\sqrt{q^2}$ and p_{0j} is the momentum evaluated at m_j. The factor $q_D = \sqrt{q^2}\sqrt{1 - \frac{4m_D^2}{q^2}}$ and q_{0D} is the momentum of outgoing D^0 mesons with mass m_D evaluated at m_j. The $\psi(3770)$ resonance is treated differently using a Flatté function [51] to account for the open charm threshold at $2m_D$

and since $\psi(3770)$ is OZI suppressed below this threshold [52–54]. The two widths $\Gamma_{1\psi(3770)}$ and $\Gamma_{2\psi(3770)}$ correspond to the partial widths of $\psi(3770)$ to states below and above the open charm threshold and are taken from Ref. [55].

Moving to the 2P charmonium contribution, the spectral density arising from the rescattering of $B\to KM_{j1}M_{j2}\to K\mu^+\mu^-$ can be written as:

$$\rho^{2P}_{c\bar{c}}(s)\propto\sum_{j=DD,D^*D^*,DD^*}\int dp_j^2\delta(s-p_j^2)\int\frac{d^3\vec{p}_{j1}d^3\vec{p}_{j2}}{16\pi^2E_{j1}E_{j2}}\mathcal{A}(B\to KM_{j1}M_{j2})\delta^{(4)}(p_j-p_{j1}-p_{j2}) \quad (2.101)$$

where $p_{j1,2}$ are four-momenta of the intermediate state and p_j is the total four-momentum of the 2P system. The contribution to C_9 can then be written as

$$\Delta Y^{2P}_{c\bar{c}}(q^2)=\sum_{j=DD,D^*D^*,DD^*}\eta_j e^{i\delta_j}A_j^{2P}(q^2),\quad A_j^{2P}(q^2)=\frac{(q^2-q_0^2)}{\pi}\int_{q^2_{min}}^{\infty}ds\frac{\hat{\rho}_j(s)}{(s-q_0^2)(s-q^2-i\epsilon)}, \quad (2.102)$$

where the normalized spectral densities of the 2P states are given by $\hat{\rho}_j$. The lower threshold of the integral q^2_{min} does not matter as long as it is lower than the square of the mass of the least heaviest pair of charmonium states. As before η_j and δ_j describe the magnitude and phase of these states.

The spectral densities are written using the two-body phasespace Källén function λ around their thresholds. Generally for a pair of states $j1, j2$, they can be written as,

$$\hat{\rho}_{j1j2}(s)\propto\sqrt{\frac{\lambda(s,m_{j1}^2,m_{j2}^2)}{s}}^{2L+1}, \quad (2.103)$$

where m_{j1} and m_{j2} are the masses of the first and second particle in the pair and L is the relative orbital angular momentum between the two particles. From angular momentum conservation, the DD and D^*D^* pairs are in the P-Wave configuration with $L=1$ while the DD^* pair is in the S-Wave configuration with $L=0$. Using this, the spectral density functions of the two-particle (2P) states are,

$$\hat{\rho}_{DD}(s)\equiv\hat{\rho}_D(s)=\left(1-\frac{4m_D^2}{s}\right)^{3/2},\qquad \hat{\rho}_{D^*D^*}(s)\equiv\hat{\rho}_{D^*}(s)=\left(1-\frac{4m_{D^*}^2}{s}\right)^{3/2} \quad (2.104)$$

$$\hat{\rho}_{DD^*}(s)\equiv\hat{\rho}_{\bar{D}}(s)=\left(1-\frac{4m_{\bar{D}}^2}{s}\right)^{1/2} \quad (2.105)$$

with the definition $m_{\bar{D}}\equiv(m_D+\mathrm{m}_{D^*})/2$ used as an excellent approximation for the mass of the intermediate DD^* state (note the redefinition of symbols for the sake of brevity).

The approximate solution to the integrals in Eq. 2.102 of the 2P contribution to C_9 is calculated in Ref. [50] to be:

$$\Delta Y^{2P}_{c\bar{c}}(q^2)=\eta_{\bar{D}}e^{i\delta_{\bar{D}}}h_S\left(m_{\bar{D}},q^2\right)+\sum_{j=D,D^*}\eta_j e^{i\delta_j}h_P\left(\mathrm{m}_j,q^2\right), \quad (2.106)$$

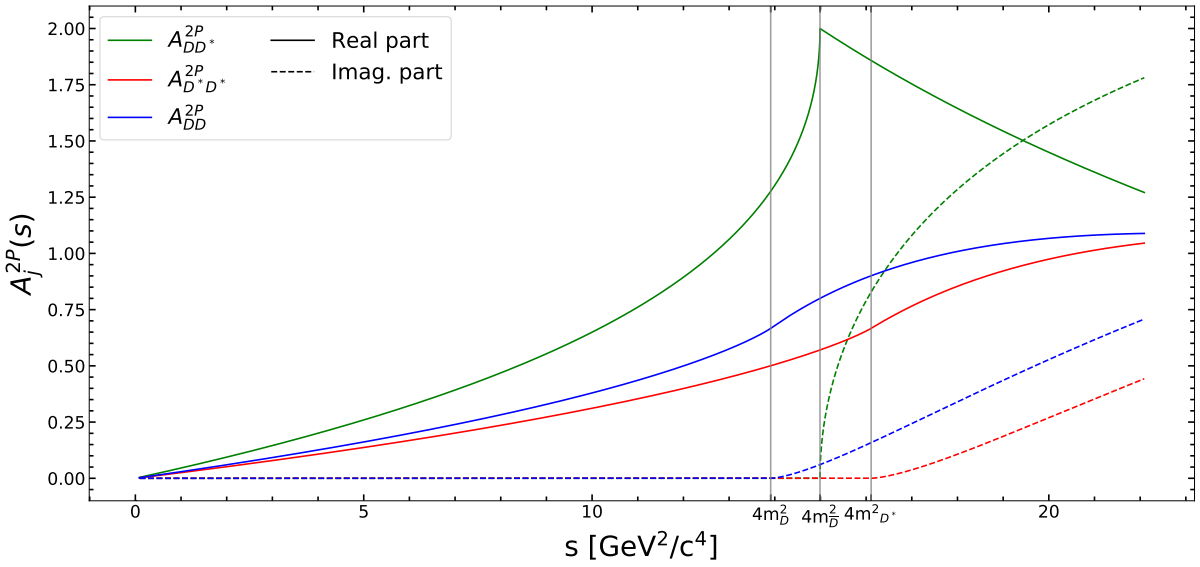

Figure 2.10: The real (solid) and imaginary (dashed) part of the 2P state amplitudes given in Eq. 2.102 and approximated using Eq. 2.106. For reference, the mass of the 2P states: m_D, m_{D^*} and $m_{\bar{D}} \equiv (m_D + m_{D^*})/2$ are marked by the vertical grey lines.

with

$$h_S(m,q^2) = 2 - G\left(1 - \frac{4m^2}{q^2}\right), \qquad h_P(m,q^2) = \frac{2}{3} + \left(1 - \frac{4m^2}{q^2}\right) h_S(m,q^2), \tag{2.107}$$

$$G(y) = \sqrt{|y|}\left\{\Theta(y)\left[\ln\left(\frac{1+\sqrt{y}}{1-\sqrt{y}}\right) - i\pi\right] + 2\,\Theta(-y)\arctan\left(\frac{1}{\sqrt{-y}}\right)\right\}. \tag{2.108}$$

Here $h_S(m,q^2)$ and $h_P(m,q^2)$ are the S- and P- Wave amplitudes of the 2P states. The S-Wave configuration of the DD^* pair produces a singular *cusp*-like behaviour at the threshold. This is shown in Figure 2.10 where the real and imaginary part of the amplitude for the three states are shown.

Having described the 1P and 2P contribution, what is left in $Y_{c\bar{c}}(q^2)$ is the subtracted term $Y_{c\bar{c}}^{q_0^2}$ in Eq. 2.94. This was introduced to ensure the convergence of the dispersion relations for the 2P states. In principle the subtraction can be at any point, for this analysis it is chosen to be $q_0^2 \equiv 0$. This value has the benefit of decoupling the determination of the q^2 spectrum with the value of the Wilson coefficient C_9 when fitting to data. The value of $Y_{c\bar{c}}^0$ containing the charm-loop effect at the subtraction point (q^2 =0) is expressed as a correction to C_9. Following Ref. [50], the value estimated in Ref. [36] is used:

$$Y_{c\bar{c}}^0 \approx -0.10 \pm 0.05. \tag{2.109}$$

2.3.2.2 The light hadronic states

The light hadronic states refer to the one-particle resonances $\rho(770)$, $\omega(782)$ and $\phi(1020)$, other states are neglected since their leptonic decay rates are very small. As stated before only 1P states are considered as the production of light quarks is loop- and CKM- suppressed. In Ref. [50] it was found that there was no clear advantage in using the subtracted vs unsubtracted dispersion relation for the light states. Therefore the unsubtracted version was used as default, the same approach is also followed in this analysis. Otherwise proceeding similarly to the determination of $\Delta Y^{1P}_{c\bar{c}}$ for the charmonium resonance, the light quark contribution is given by[11],

$$Y_{light}(q^2) = Y^{1P}_{light}(q^2) = \sum_{j=\rho,\omega,\phi} \eta_j e^{i\delta_j} A^{res}_j(q^2), \tag{2.110}$$

with A^{res}_j given exactly as in Eq. 2.98 for $j \neq \psi(3770)$.

2.3.2.3 Isobar Model

An alternative empirically motivated method to account for the intermediate states is the so-called isobar approach. Here, the 1P $c\bar{c}$ states and light hadronic states are described using sums of relativistic Breit-Wigner amplitudes. The resulting contributions to the Wilson coefficient C_9 are exactly as given in Eq. 2.110 for all the resonances and any contribution from 2P states are neglected. In contrast to the use of dispersion relations, the isobar approach does not guarantee the convergence of the model to agree with the theory calculations from LCSR at low q^2.

It is worth highlighting that the previous iteration of this measurement by the LHCb collaboration uses the isobar approach. Therefore, not only does this analysis supersede the previous result in terms of improved statistics, the physics interpretations of the result will be more accurate.

2.4 Building the Differential Decay Rate of $B^+ \to K^+\mu^+\mu^-$ Decays

In the previous two sections an EFT was built to describe $b \to s\ell^+\ell^-$ transitions introducing effective operators and Wilson coefficients. This theory was matched to the SM at the m_W scale and evolved down to the m_b scale. The remaining hadronic matrix elements are then encoded in the form factors which are estimated using techniques such as LatticeQCD [44]. The non-local long distance effects from 1P and 2P intermediate states are modelled using dispersion relations which allows a data-driven method to estimate these effects. This contribution is constructed such that they enter as corrections to the Wilson coefficient C_9.

[11] Notice the absence of the factor $\frac{q^2}{m_j}$

Consolidating this and following Refs. [34] and [30], the amplitude for $B^+ \to K^+\ell^+\ell^-$ decays is written as[12]:

$$i\mathcal{M}[B^+ \to K^+\ell^+\ell^-] = i\frac{G_F\alpha_e}{\sqrt{2}\pi}V_{tb}V_{ts}^* f_+(q^2)\left(F_A p_B^\mu[\ell^+\gamma_\mu\gamma_5\ell^-] + F_P[\ell^+\gamma_5\ell^-] + F_V p_B^\mu[\ell^+\gamma_\mu\ell^-]\right), \tag{2.111}$$

where p_B is the four momentum of the B meson and α_e is the electromagnetic structure constant. The functions $F_i \equiv F_i(q^2)$ are given by,

$$F_A = C_{10}; \tag{2.112}$$

$$F_P = m_l C_{10}\left[\frac{M_B^2 - M_K^2}{q^2}\left(\frac{f_0(q^2)}{f_+(q^2)} - 1\right) - 1\right]; \tag{2.113}$$

$$F_V = C_9 + Y(q^2) + 2C_7^{\text{eff}}\frac{m_b + m_s}{M_B + M_K}\frac{f_T(q^2)}{f_+(q^2)}; \tag{2.114}$$

with $M_{B,K,b,s}$ being the mass of their respective mesons and quarks. All other notations are exactly as given in the previous section.

The $B^+ \to K^+\ell^+\ell^-$ decay can be described using only two kinematic variables q^2 and $cos\theta_\ell$ with q^2 being the momentum carried away by the lepton pair as defined earlier. The angle θ_ℓ is defined as the angle between the track of the negatively charged lepton and B^+ meson in the rest frame of the dimuon system[13]. The double differential decay rate for $\ell = e, \mu$ with respect to q^2 and $cos\theta_\ell$ is given as,

$$\frac{d^2\Gamma_\ell}{dq^2 dcos\theta_\ell} = C(q^2)\left[a_\ell(q^2) + b_\ell(q^2)cos\theta_\ell + c_\ell(q^2)cos^2\theta_\ell\right], \tag{2.115}$$

using the following definitions,

$$C(q^2) = \frac{G_F^2\alpha_e^2|V_{tb}V_{ts}^*|^2}{512\pi^5 M_B^3}\beta(q^2)\sqrt{\lambda(q^2)}f_+^2(q^2), \tag{2.116}$$

$$a_\ell(q^2) = \frac{\lambda(q^2)}{4}\left(|F_A|^2 + |F_V|^2\right), \tag{2.117}$$

$$b_\ell(q^2) = 0, \tag{2.118}$$

$$c_\ell(q^2) = -\frac{\lambda(q^2)}{4}\beta^2(q^2)\left(|F_A|^2 + |F_V|^2\right), \tag{2.119}$$

and

$$\beta(q^2) = \sqrt{1 - \frac{4m_\ell^2}{q^2}}, \tag{2.120}$$

[12] In Refs. [34] and [30] the amplitude includes BSM operators which we can safely neglect here as their contributions are expected to be negligible. Inclusion of such contributions is left for a future work.

[13] For the CP conjugate decay of $B^- \to K^-\ell^-\ell^+$, θ_ℓ is the angle between the positively charged lepton and B^- meson.

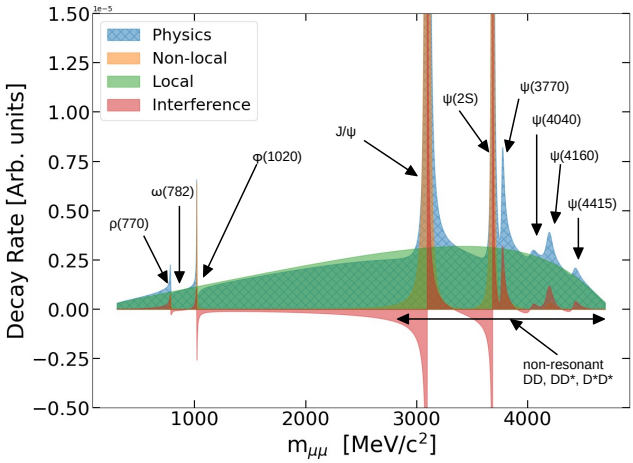

Figure 2.11: An illustration of the total physics model used to describe $B^+ \to K^+\mu^+\mu^-$ decays. The individual contributions from the local and non-local effects and from the interference between them are given by the different colour bands. Labelled are the various resonant 1P and non-resonant 2P states.

$$\lambda(q^2) = M_B^4 + M_K^4 + q^4 - 2M_B^2 M_K^2 - 2M_B^2 q^2 - 2M_K^2 q^2. \tag{2.121}$$

This analysis is performed in q^2 with the decay rate integrated in $cos\theta_\ell$ between $[-1,1]$. The SM description of the CP-averaged differential decay rate for $B^+ \to K^+\ell^+\ell^-$ decays is then given by,

$$\begin{aligned}\frac{d\Gamma_\ell}{dq^2} = &\frac{G_F^2\alpha^2|V_{tb}V_{ts}^*|^2}{2^7\pi^5}|\mathbf{k}|\beta_+\left\{\frac{2}{3}|\mathbf{k}|^2\beta_+^2\left|C_{10}f_+(q^2)\right|^2 + \frac{m_\ell^2(M_B^2-M_K^2)^2}{q^2M_B^2}\left|C_{10}f_0(q^2)\right|^2\right.\\ &\left. + |\mathbf{k}|^2\left[1-\frac{1}{3}\beta_+^2\right]\left|C_9^{\text{eff}}f_+(q^2) + 2C_7^{\text{eff}}\frac{m_b+m_s}{M_B+M_K}f_T(q^2)\right|^2\right\},\end{aligned} \tag{2.122}$$

where $\mathbf{k} = \sqrt{E_K^2 - M_K^2}$ is the kaon momentum in the rest frame of the B-meson and is related to the λ function by,

$$4M_B^2|\mathbf{k}| = \lambda(q^2). \tag{2.123}$$

Using the definition $q^2 \equiv m_{\mu\mu}^2$, the total physics model for the dimuon mass spectrum of $B^+ \to K^+\mu^+\mu^-$ decays can be determined and is shown in Figure 2.11. The peaks corresponding to the non-local 1P states which contribute to the amplitude are labeled accordingly. The non-resonant 2P states have an amplitude structure which spans the entire dimuon range, the region where they are most dominant is labelled. The total non-local contribution is highlighted in yellow. The contribution highlighted in green is the local penguin contribution whose amplitude is parameterized by $C_{9,10}$ and C_7^{eff} and the form factors. The red bands which have both positive and negative parts describes the interference between the local and non-local contributions.

2.4.1 Branching fraction measurement

By integrating only the non-local contributions in Eq. 2.122, their branching fractions can be calculated.

$$\mathcal{B}(B^+ \to K^+ X_j) \times \mathcal{B}(X_j \to \ell^+\ell^-) = \frac{\tau_B}{\hbar} \frac{G_F^2 \alpha^2 |V_{tb} V_{ts}^*|^2}{128\pi^5} \int_{4m_\ell^2}^{(M_B - M_K)^2} |\mathbf{k}|^2 \left[\beta_+^2 - \frac{1}{3}\beta_+^3\right] \left|f_+(q^2)\right|^2 \left|\eta_j\right|^2 \left|S_j(q^2) A_j^{res}(q^2)\right|^2 dq^2, \tag{2.124}$$

where τ_B is the lifetime of a B^+ meson and $\hbar$ is the reduced Planck's constant. The factor $S_j(q^2) = \frac{q^2}{m_j^2}$ for the charmonium resonances $j = J/\psi ... \psi(4415)$ and $S_j(q^2) = 1$; for the light resonances $j = \rho, \omega, \phi$. This factor is a direct consequence of the subtracted dispersion relation used for the charmonium states. The branching fractions of intermediate 2P states can be calculated by making the substitution $A_j^{res} \to A_j^{2P}$ and using $S_j(q^2) = 1$.

2.5 Searching for Lepton Universality Violating τ couplings

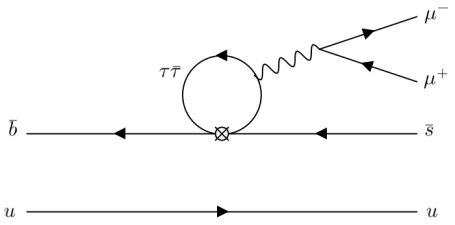

Figure 2.12: The production of intermediate tau loops which could have a sizable contribution to the $B^+ \to K^+\mu^+\mu^-$ amplitude in the presence of NP.

Another amplitude which has the potential to contribute to the $B^+ \to K^+\mu^+\mu^-$ spectrum is through the scattering process of $B^+ \to K^+\tau^+\tau^- \to K^+\mu^+\mu^-$. Here an intermediate τ-loop is produced which electromagnetically scatters to a muon pair. The dominant Feynman diagram for this process is shown in Figure 2.12. The τ-loop contribution can be calculated perturbatively as a shift to C_9^{eff} [50],

$$C_9^{\text{eff}} \to C_9^{\text{eff}} + Y_{\tau\bar{\tau}}(q^2), \tag{2.125}$$

with

$$Y_{\tau\bar{\tau}}(q^2) = -\frac{\alpha^2}{2\pi} C_9^\tau \left[h_S(m_\tau, q^2) - \frac{1}{3} h_P(m_\tau, q^2)\right], \tag{2.126}$$

where $h_S(m_\tau, q^2)$ and $h_P(m_\tau, q^2)$ are as defined in Eq. 2.107 and C_9^τ is the vector coupling to taus.

The contribution of $\tau^+\tau^- \to \mu^+\mu^-$ in $B^+ \to K^+\mu^+\mu^-$ transitions allows one to probe the short distance $b \to s\tau^+\tau^-$ amplitude and thereby lepton universality violating contributions to the branching fraction of $B^+ \to K^+\tau^+\tau^-$ decays. The current upper bounds on this branching fraction is given by the BaBar experiment $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) < 2.25 \times 10^{-3}$ at 90% CL [56] while $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)_{SM} \approx 1.5 \times 10^{-7}$ [57]. Only assuming contributions from $\mathcal{O}_{9,10}$ the BaBar result implies $C_9^\tau \leq 510(910)$ for $C_{10}^\tau = -C_9^\tau (C_{10}^\tau = 0)$ [50]. This is two orders of magnitude larger than the SM value of $C_9^{\tau,SM} \approx 4.2$. At this upper bound a clear modulation of the dimuon spectrum and a "cusp" at $2m_\tau$ is expected. In Ref. [50] it is shown that with the total number $B^+ \to K^+\mu^+\mu^-$ candidate events collected at LHCb enables one to set limits on C_9^τ with precision commensurate to current limits. Given the difficulty of reconstructing taus at LHCb, measuring C_9^τ through the dimuon spectrum of $B^+ \to K^+\mu^+\mu^-$ decays offers a novel way of accessing NP in τ couplings.

2.6 Hints of Beyond Standard Model Physics in $b \to s\ell^+\ell^-$ transitions

The chapter has so far discussed the theoretical description of $b \to s\ell^+\ell^-$ transitions and describes the measurement performed for this thesis. This work has been motivated by the anomalies currently seen in B-meson decays when compared to SM predictions in multiple channels sensitive to different observables. These include branching fraction measurements, angular observables and Lepton Flavour Universality (LFU) tests. Combined together this points to the existence of physics beyond the SM. In this section, a summary of the various $b \to s\ell^+\ell^-$ measurements are shown along with global fits which help characterize the NP contributions.

2.6.1 The anomalous measurements

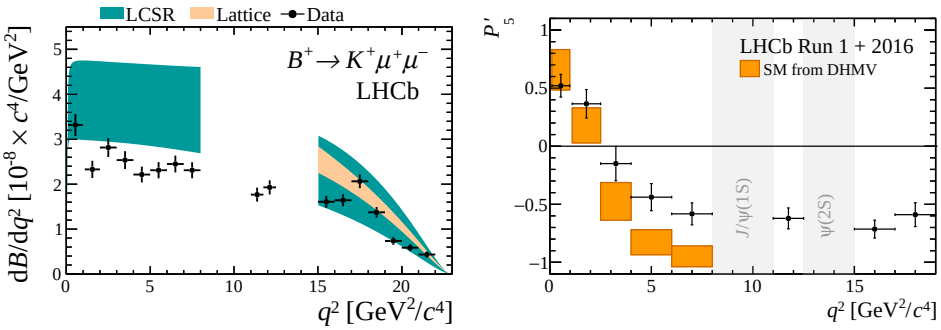

Figure 2.13: (Left) The q^2 dependent decay rate measurement of $B^+ \to K^+\mu^+\mu^-$ decays [4]; (Right) The q^2 dependent distribution of the 'clean' angular observable P'_5 measured using $B^0 \to K^{*0}\mu^+\mu^-$ decays [7]. In both plots the solid colour bands are from theoretical estimates.

Comparing the measured q^2 dependent decay rate with the SM prediction will highlight NP effects which influence b-hadron decays. Multiple decay channels such as $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^{*0}\mu^+\mu^-$, $\Lambda_b \to \Lambda\mu^+\mu^-$, etc. [4, 58–60] have been studied and the results for $B^+ \to K^+\mu^+\mu^-$ are shown in Figure 2.13. In this plot the large colour bands indicate the large uncertainty in theoretical estimates of the form factor from LCSR or LatticeQCD calculations. Nevertheless, the measured value has been observed to be below these estimates and with similar observations for the other decay channels with differing spectator quarks. This indicates a genuine NP artefact rather than a systematic effect.

Another point to note here is the removal of data from the regions occupied by the intermediate resonances $\phi(1020)$, J/ψ and $\psi(2S)$. In these mass regions, the amplitude is dominated by these narrow resonances and have a large theoretical uncertainty. However, this contribution has the potential to mimic NP and understanding its effect is crucial to understand the current anomalies.

It is a common strategy to construct observables which are ratios or differences where the form factors cancel at least to leading order. Referred to as 'theoretically clean' since the pollution from the form factors is kept to a minimum level, this includes observables of the form

$$R_X = \frac{\int_{q^2_{min}}^{q^2_{max}} \mathcal{B}(B \to X\mu^+\mu^-) dq^2}{\int_{q^2_{min}}^{q^2_{max}} \mathcal{B}(B \to Xe^+e^-) dq^2} \tag{2.127}$$

where $X = K, K^{*0}$ with $K^{*0} \to K^+\pi^-$ and the integral is performed over the same q^2 range for electrons and muons. In the SM, this ratio is expected to be $R_X = 1 + \mathcal{O}(m^2_\mu/m^2_b)$ and has a much smaller hadronic uncertainty $\mathcal{O}(10^{-4})$ [34, 61]. The QED corrections are also at $\mathcal{O}(1\%)$ level [62]. Observation of any significant deviations from 1 is a clear sign of Beyond Standard Model physics. The latest measurement of R_K done by LHCb using the full available data set shows evidence of LFU violation at 3.1σ [1]. Similar patterns can be seen in $R_{K^{*0}}$ [2] and the baryonic decay ratio [3]

$$R^{-1}_{pK} = \frac{\int_{q^2_{min}}^{q^2_{max}} \mathcal{B}(\Lambda^0_b \to pK^-e^+e^-) dq^2}{\int_{q^2_{min}}^{q^2_{max}} \mathcal{B}(\Lambda^0_B \to pK^-\mu^+\mu^-) dq^2}, \tag{2.128}$$

with tensions up to 2.5 σ. The experimental uncertainty on these ratios is currently statistically dominated and expected to reduce with more data. The measured ratios $R_{K^{*0}}$ and R_{pK} do not use the full LHCb data set currently, the update of which is expected to be released soon. However the current results are consistent with branching fraction measurements and R_K which indicate a deficit of muons w.r.t. electron. This shows signs of BSM physics which has lepton non-universal behaviour and a hierarchical structure.

In addition, the branching fraction of the doubly helicity suppressed $B^0_s \to \mu^+\mu^-$ decay was recently measured at the LHC by ATLAS, CMS and LHCb [63–66]. The combined result

is compatible with the SM within 2 σ. This and the current best limits on $B_s^0 \to e^+e^-$ [67] and $B_s^0 \to \tau^+\tau^-$ [68] sets stringent constraints on possible NP models using $\mathcal{O}_{S,P}$ [40].

Another set of observables available at our arsenal is to use the angular distribution of $b \to s\ell^+\ell^-$ transitions. With this, angular observables can be constructed which exploit the available information better and offers complimentary sensitivity towards NP. Although specific conventions vary, the differential decay rate of $B \to X\mu^+\mu^-$ decays can be described by 3 angles[14] $(\theta_k, \theta_l, \phi)$ and q^2. Similar to decay rate measurements, theoretically clean observables can also be constructed using angular observables which are at leading order form factor independent. One such variable is P_5', Figure 2.13 shows the measured value for this observable using $B^0 \to K^{*0}\mu^+\mu^-$ decays where deviations from predictions can be observed. The most interesting aspect of this is that a Wilson fit of all 8 CP averaged observables with LHCb data has a $\sim 3\sigma$ tension with the SM value for $B^+ \to K^{*+}\mu^+\mu^-$ [6] and $B^0 \to K^{*0}\mu^+\mu^-$ [7] decays.

2.6.2 Global fits and significance of New Physics

Multiple phenomenology groups have been working to interpret the anomalies. This procedure involves using all the available experimental results of $b \to s\ell^+\ell^-$ transitions and studying them using the EFT framework. Within this framework, a general hypothesis of short-distance NP which contribute to $b \to s\ell^+\ell^-$ transitions are studied. As discussed is Section 2.2 short distance effects are encoded in Wilson coefficients, the couplings to the operators.

The global fits combine the experimental observables and identify combinations which are anomalous with the SM prediction. This exercise primarily helps distinguish between NP scenarios and to identify more likely scenarios in a model-independent way. This is done in Ref. [69] and later updated with the latest evidence of NP in R_K in Ref. [9]. Figure 2.14 shows the hypothesis where NP only contributes to C_9, C_{10} of the muon mode with a global fit value 6.8σ from the SM value. In Ref. [9] various other scenarios are also explored which over all indicate some form of NP. Models which describe a LeptoQuark (LQ) or a heavier neutral current Z' have been demonstrated to be good candidates to explain these anomalies [70].

While such studies inform towards well motivated NP models, they do not give a global unbiased significance of the existence of NP. This question is addressed in Ref. [8] where an ensemble of pseudo-experiments are generated at the SM values with the experimental precision. The ensemble is used to calculate the probability to observe the coherent deviations seen in data. Allowing all possible deviations in measurements and Wilson coefficients, Ref. [8] calculates a global significance of 4.3σ for the $b \to s\ell^+\ell^-$ anomalies. This result signifies that existence of physics beyond the SM and shows that precision measurements of heavy flavour physics are the best place to look for it. The current precision on most experimental results are statistically dominated. Therefore, gaining more information from nature on heavy flavour physics especially $b \to s\ell^+\ell^-$ transitions are important now more than ever.

[14] In the case of $B^+ \to K^+\mu^+\mu^-$ decays where the final state has only three particles, θ_k and ϕ are not defined.

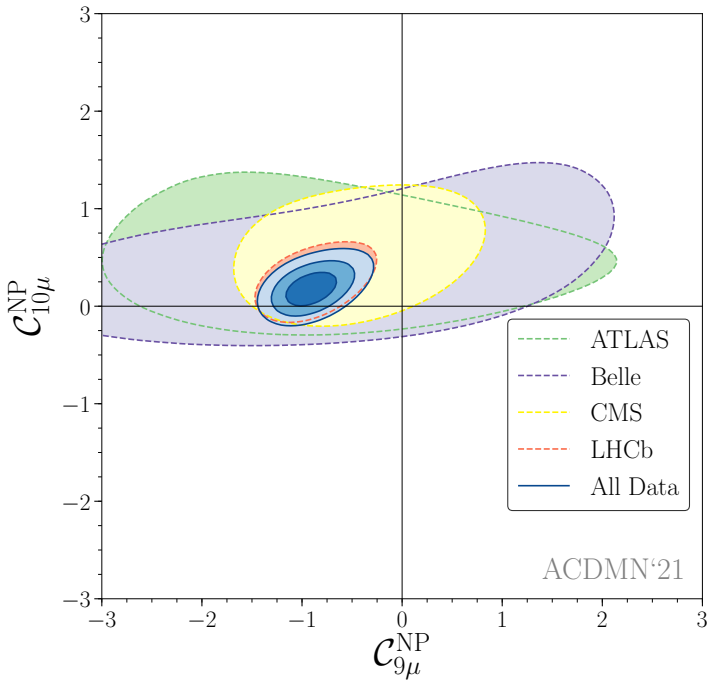

Figure 2.14: In this figure $C_{9\mu}^{NP}$ and $C_{10\mu}^{NP}$ are the contributions to Wilson coefficients C_9 and C_{10} in the muon mode of $b \to s\ell^+\ell^-$ transitions. This lepton flavour universality violating scenario is away from the SM with a significance of 6.8σ [9].

2.6.3 Anomalous results in $b \to c\ell\overline{\nu}_\ell$ transitions

The discussion in this thesis has solely been focused on $b \to s\ell^+\ell^-$ transitions, however, the tree-level decay of a b-quark to a c-quark and a lepton-neutrino pair is also a good place to look for NP. Similar to the R-ratios discussed above, theoretically clean observables of $R_{D^{(*)}}$ using the decay $B \to D^{(*)}\ell\overline{\nu}_\ell$ with $\ell = e, \mu$ in the denominator and τ in the numerator can be constructed. Recent combinations of these ratios also show a disagreement with the SM value at the $\sim 3\sigma$ level [5]. This result implies at a hierarchical BSM effect, with the largest contribution to the τ modes. Such a NP effect could also affect $b \to s\ell^+\ell^-$ transitions by significantly enhancing $b \to s\tau^+\tau^-$ decays such as $B^+ \to K^+\tau^+\tau^-$ by 3 orders of magnitude w.r.t. the SM value [10, 71, 72]. The current experimental limits allows such a value and observation of this decay could be within reach.

2.7 Motivation for a full q^2 dependent $B^+ \to K^+\mu^+\mu^-$ decay rate measurement

When studying $b \to s\ell^+\ell^-$ transitions it is common practice to avoid q^2 regions dominated by non-local effects caused by intermediate resonances. For the decay rate and angular measure-

ments discussed earlier, the data in q^2 regions from $\phi(1020)$, J/ψ and $\psi(2S)$ are removed. The q^2 regions in LFU tests also are chosen to avoid the resonances. This is done to avoid the large theoretical uncertainties in estimating the non-local effects.

This conservative choice ignores the fact that if the non-local effects are sufficiently large they have the potential to mimic NP effects. While global fits account for non-local contributions from theory calculations, given the difficulty in theoretically calculating the non-local effects it is important to use data to measure these intractable contributions. Figure 2.15 shows the differential decay rate from Eq. 2.122 evaluated with only the local contribution and non-local J/ψ and $\psi(2S)$ resonances. The effect of the resonances with different phase differences and thereby at different interferences levels has a significant effect on the dimuon mass spectrum. Accurate knowledge of the phases is therefore necessary to be able to unequivocally claim the existence of BSM physics.

The analysis presented in this thesis tries to answer this question for $B^+ \to K^+\mu^+\mu^-$ decays. To this result, the full q^2 spectrum of $B^+ \to K^+\mu^+\mu^-$ decays is fitted with the non-local effects explicitly accounted for. Through this method any residual deviation to C_9 and C_{10} would purely be from a NP contribution. This result will add to the global fit results and further guide NP model building and resolve the debate as to the extent of the effect of non-local contributions. The measurement was performed earlier by LHCb using Run1 data of $3\,\text{fb}^{-1}$ with the isobar approach; the effect of the interference was found to be minimal. The analysis presented here will supersede the previous measurement using both Run1+Run2 data with a ~ 3.5 fold increase in the number of $B^+ \to K^+\mu^+\mu^-$ decays and an improved description of the non-local effects.

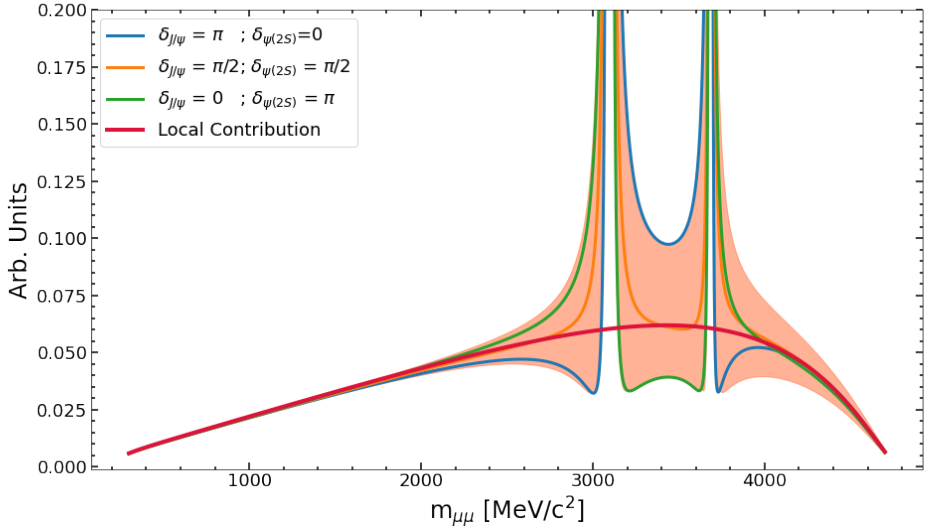

Figure 2.15: An illustration of the differential decay rate from Eq.2.122 with only the J/ψ and $\psi(2S)$ non-local contributions. Different phases w.r.t. the local contribution are shown by the orange band with selected phases marked by the curves. In order to highlight the extent of the effect of different phases, the local contribution is marked by the red curve.

CHAPTER

3 The LHCb Experiment

This chapter broadly follows [73] unless otherwise stated.

The LHCb experiment is one of the four large experiments operating at the Large Hadron Collider (LHC) at CERN, Geneva. The LHC is a particle collider designed to collide protons and ions. It uses the tunnel built for the Large Electron Positron (LEP) collider which is a 27 km ring about 100 m underground along the Franco-Swiss border.

This dissertation only discusses data collected in pp collision with collision frequency of 40 MHz. In addition, this dissertation uses data taken in two run periods of the LHC commonly known as LHC Run1 (2010-2012) and LHC Run2 (2015-2019). Run1 data collected in 2011 and 2012 is used here when the LHC was operated at reduced collision centre-of-mass energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV respectively. The integrated luminosity collected in 2011 and 2012 correspond to 1 fb^{-1} and 2 fb^{-1} respectively. Run2 data collected between 2016-2018 is used when the LHC was operated at centre-of-mass collision energies of $\sqrt{s} = 13$ TeV. The integrated luminosity in Run2 corresponds to about 2 fb^{-1} for each year.

The LHCb experiment is located at interaction Point 8 near Ferney-Voltaire, France just adjacent to Geneva airport and uses the cavern used by the DELPHI experiment. Designed to study the properties of b hadrons it is a single-arm forward spectrometer covering the azimuthal beam angle range of $10 < \theta < 250$ mrad or equivalently the pseudorapidity[1] range $2 < \eta < 5$. The choice of using this angular acceptance for the detector was driven by the enhanced $b\bar{b}$ production at small angles (Figure 3.1). The LHC optics were modified to displace the interaction point by 11.25 m from the centre, permitting the maximum usage of the cavern. In addition, the instantaneous luminosity for the LHCb experiment can be changed independently of other experiments. This enables the experiment to run at lower instantaneous luminosities which has the advantage of having only a single pp interaction per bunch crossing. This is

[1] Pseudorapidity is defined as $\eta = -ln(\tan(\theta/2))$

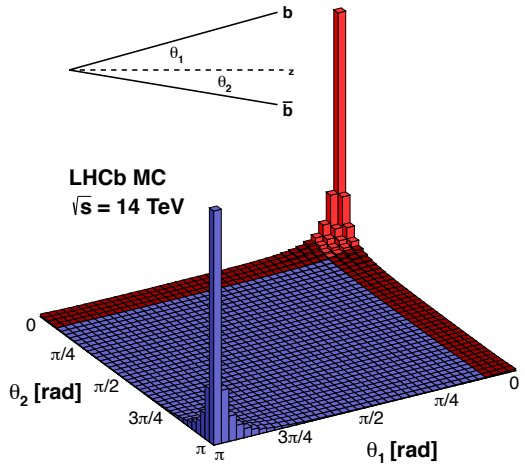

Figure 3.1: The simulated $b\overline{b}$ quark pair production as a function of opening angle θ_1 and θ_2. The red coloured section shows the acceptance of the LHCb detector [73].

a crucial requirement for the experiment since the number of overlapping interactions with multiple vertices is significantly reduced, allowing for excellent track and vertex reconstruction of short-lived particles.

A cross-sectional diagram of the detector is shown in Figure 3.2. The pp interaction point is at the origin of the cartesian coordinates on the left. After a bunch crossing, forward-going particles move from left to right along the detector in this diagram. In literature usually 'upstream' is used refer to the part closer to the interaction point and 'downstream' to the part away from it. The detector comprises a warm dipole magnet providing an integrated field of 4 Tm and several sub-detectors used for tracking, Particle Identification (PID), and calorimetry.

The particles produced in the collision first traverse through the VErtex LOcator (VELO) which provides high resolution tracking and vertexing information. Then they encounter the first of two Ring Imaging Cherenkov (RICH) detectors which contributes to the PID information followed by the upstream tracking station, Tracker Turicensis (TT). Next is the dipole magnet and the downstream tracking station T1-T3 followed by the second Ring Imaging Cherenkov (RICH). After this is the first muon station M1 and the calorimetry system: the Scinitillating Pad Detector (SPD), Preshower Detector (PS), Electromagnetic Calorimeter (ECAL), Hadronic Calorimeter (HCAL).

The calorimetry system is used for the triggers and it identifies electrons, photons and hadrons as well as measure their energies and positions. Finally there are the rest of the muon stations M2-M5 used to identify and measure the kinematics of muons. The rest of this chapter details the working mechanism and the performance of these sub-detectors.

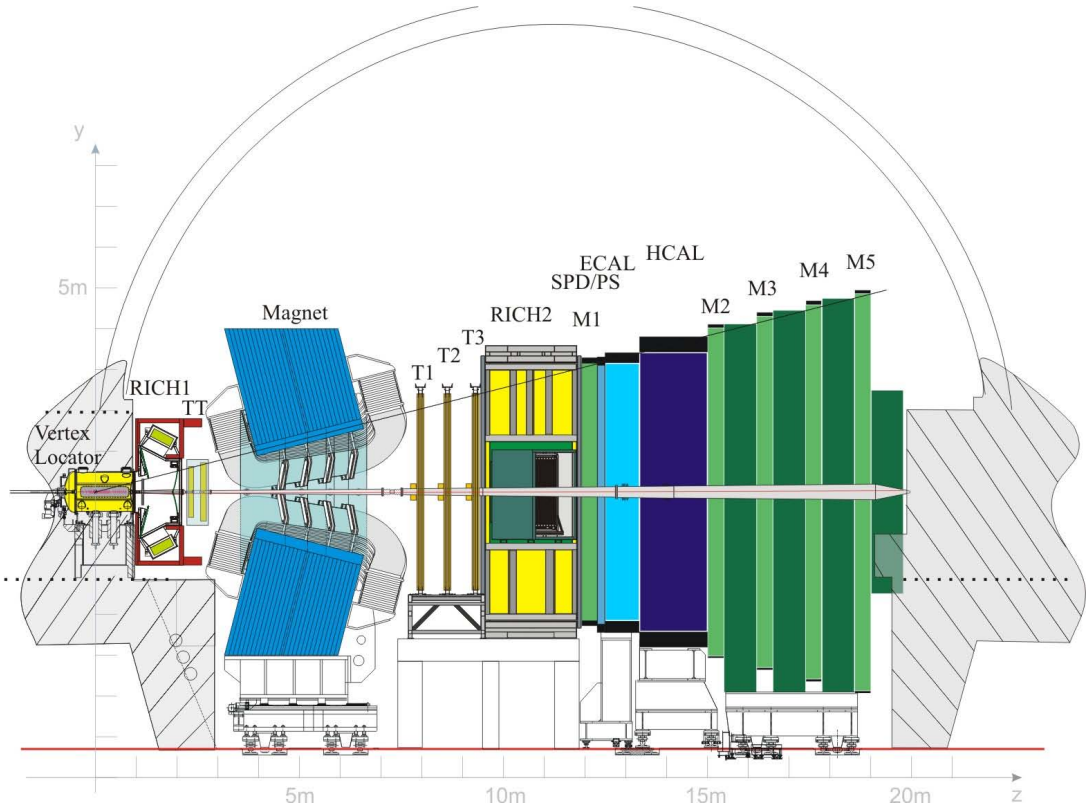

Figure 3.2: A side view of the LHCb detector [73].

In this thesis $B^+ \to K^+\mu^+\mu^-$ decays are studied. A B hadron produced in the pp collision travels only a few millimeters and decays within the VELO. The charged kaon and muons produced in this decay then travel through the VELO and the tracking stations where they deposit energy onto detector pads which record the "hits". Particle tracks are reconstructed by performing a fit to the hits in each event. Each track is then associated to vertices and a PID or mass hypothesis is assigned to them using information from the RICH and muon detectors.

If a set of three tracks associated with a K^+, μ^+ and μ^- is reconstructed to originate from the same vertex it is deemed as a candidate. This could be from a true $B^+ \to K^+\mu^+\mu^-$ decay or from a background process. It should be noted that an event of a pp bunch crossing could have also have multiple candidates primarily due to a different PID hypothesis or combination of tracks from other B hadron decays in the event. Every interesting event is initially selected to be stored on tape based on simple tracking and vertex reconstruction algorithm, this process is known as triggering and is discussed in Section 3.5. The stored events are then analysed "offline" using more complex algorithms which take more time to filter out background events from interesting $B^+ \to K^+\mu^+\mu^-$ candidates.

3.1 Tracking Charged Particles

The tracking of charged particles is done by the VELO, the TT upstream of the magnet and the tracking stations T1-T3 downstream of the magnet. The VELO and TT are silicon microstrip detectors while T1-T3 uses silicon microstrip close to the beam pipe (the Inner Tracker (IT)) and straw-tube in the outer regions of the detector (Outer Tracker (OT)). The VELO is a crucial detector which gives detailed information regarding the decay of the b hadron and is therefore discussed in detail in Section 3.1.1. The other tracking stations and their combined performance is briefly discussed in Section 3.1.2.

3.1.1 VELO

The VELO gives precise track coordinates close to the pp interaction point. Since the b hadron decays within a few millimeters at LHC collision energies, the information from the VELO is vital to distinguish the Primary Vertex (PV) and Secondary Vertex (SV). Figure 3.3 shows an illustration of a decay, the PV is the pp collision point producing the b-hadron along with other particles. The b-hadron travels for the Flight Distance (FD) before decaying into the child particles at the SV.

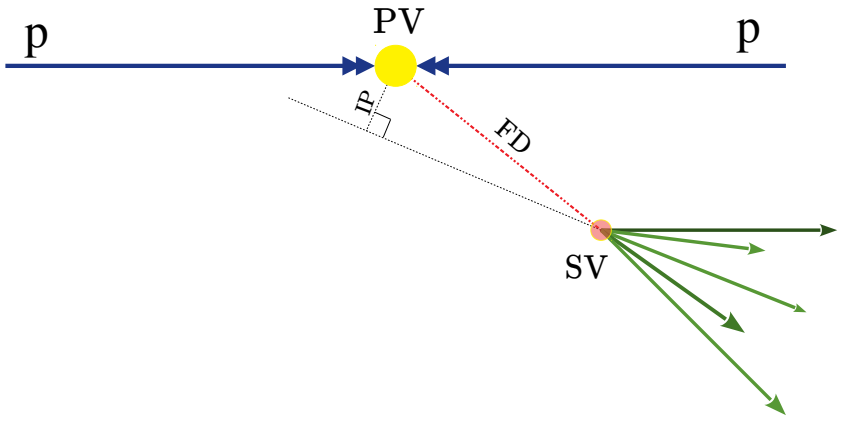

Figure 3.3: An illustration of a B hadron decay within the VELO. The b-hadron produced in the pp collision at the Primary Vertex (PV) travels the length of the Flight Distance (FD) typically a few cm before decaying into the child particles at the Secondary Vertex (SV). The Impact Parameter (IP) is defined as the line perpendicular to the line of flight of the reconstructed tracks and the PV.

The VELO is a fine grained tracker using a series of silicon modules. Each module has two set of strips arranged either in radial or angular geometry for measuring the radius (R) and azimuthal angle (ϕ) respectively. Figure 3.4 shows an illustration of one such module and the left half of the detector. There are in total 42 modules with 21 on each half of the VELO. An additional 4 modules are placed in front of the VELO sensors to improve the PV reconstruction and reduce multiple interaction events. The two arms are designed to be retracted during beam

injection[2] at the LHC and are brought in to 7mm from the beam during stable pp collision conditions. For PV emerging in the range $|z| < 10.6\,\text{cm}$ the VELO covers the pseudorapidity range of $1.6 < \eta < 4.9$ covering the angular acceptance of all downstream detectors.

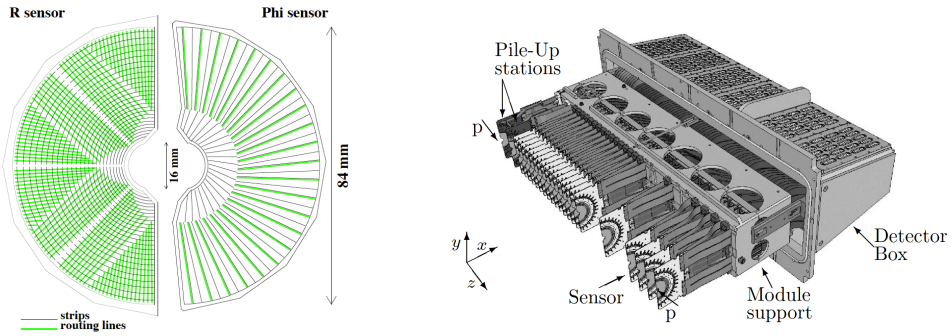

Figure 3.4: (left) The Rϕ sensor module of the VELO [74]. (right) An overview of the left half of the VELO module along with different components highlighted. [75].

The choice of using cylindrical geometry (Rϕ) was done to allow fast reconstruction of the Impact Parameter (IP) with sufficient resolution to select b-hadrons. The IP is the minimum distance from the line of flight of a track to the PV (see Figure 3.3). The decay vertex of the B mesons studied are generally displaced from the PV and therefore a good resolution of the IP is important to select candidates. The IP resolution depends on the component of momentum transverse to the beam line known as transverse momentum (p_{T}) and is measured to be $(15 + 29\,\text{GeV}^{-1}\,c/p_{\text{T}})\,\mu\text{m}$. This means a higher p_{T} results in a better IP resolution saturating at 15 μm.

The tracking efficiency of the VELO for long tracks is 98% or higher as measured from data. The hit resolution of the silicon devices depends on the inter-strip pitch and the charge shared between strips (which depends on the position of the strip in the VELO). This resolution is measured to be between 4 and 30 μm with the worst resolution being at large strip pitches. The PV resolution depends on N, the number of tracks used to reconstruct the PV. With the minimum required tracks of 5, the resolution is about 35 μm in the x and y direction and follows the relations $\sigma_{xV} = -1 + 97.7/N^{0.61}$ and $\sigma_{yV} = -1.5 + 92.7/N^{0.59}$. In the z direction the PV resolution is about 300 μm for 5 tracks and follows the relation $\sigma_{zV} = -16 + 923/N^{0.69}$ [74].

3.1.2 Tracking stations and magnet performance

The Tracker Turicensis (TT) upstream of the magnet and the Inner Tracker (IT) for each T1-T3 downstream of the magnet are made of silicon microstrip. The single hit resolution of these silicon strips ranges between 50 to 55 μm with a hit efficiency above 99% [76]. The outer parts of T1-T3 are made of tubes filled with gas. A charged particle traversing the tube ionises the

[2]Beam injection is when protons are injected to LHC from upstream accelerators and the LHC increases the energy of the colliding particles to the desired value

gas along its trajectory. The electrons then drift towards the wire which is present at the center of each tube. This drift-time is used to calculate the distance of the particle trajectory to the wire. The single hit resolution is about 200 μm with hit efficiency about 99% within a radius of 1.25 mm from the wire [76]. The magnet has a bending power of 4 Tm and in combination with tracking stations upstream and downstream the momentum of the tracks can be measured. An illustration of the different tracking stations is shown in Figure 3.5. The relative momentum resolution, $\delta p/p$ for the momentum of particles used in this measurement is about 0.5% [76]. The polarity of the magnet is changed periodically during each year to be able to perform CP violation measurements and assess any polarity bias.

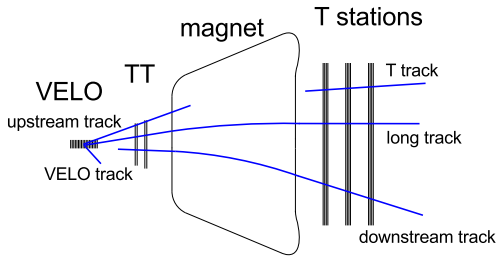

Figure 3.5: A illustration of the layout of tracking stations and the different track types reconstructed by the track finding algorithms. [77]

3.2 Calorimeters

LHCb employs a set of 4 sub-detectors in its calorimeter system, the Scinitillating Pad Detector (SPD), Preshower Detector (PS), Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL). The SPD and PS are upstream of the ECAL and the HCAL is downstream of the ECAL. The layout of the calorimeter system is shown in Figure 3.6. The four sub-detectors are each segmented into sections in the x-y plane with its size growing from the inner to outer region, this is to accommodate for the high particle densities closer to the beam pipe. The SPD, PS and ECAL are used in the trigger and offline analysis to identify electrons, photons and neutral pions. The HCAL is used in the trigger systems (Section 3.5) by measuring the hadronic shower transverse energy and to improve electron/hadron separation.

The SPD and PS have two plastic scintillator layers separated by a 15 mm thick lead plate. Electrons and photons radiate in the lead plate which is sampled by the scintillator downstream, the light from the scintillator is sent to the Photomultiplier Tubes (PMT) using wavelength-shifter (WLS) fibres. The ECAL has modules made of 4 mm scintillating tiles and 2 mm lead tile, stacked one behind the other. A total of 66 layers of such modules are used with WLS to guide the light to the PMT in each ECAL cell. The HCAL is made of iron absorbers

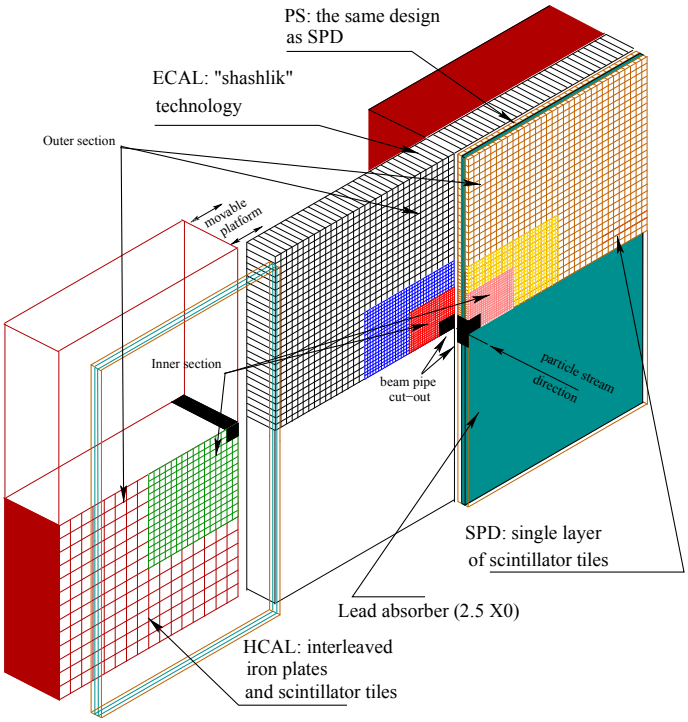

Figure 3.6: The layout of the calorimeter system.

and scintillating tiles with the scintillating tiles placed parallel to the beam axis. Again WLS fibres are used to transport the light to the PMT [78].

The information from the calorimeter system is only of limited use to select $B^+ \to K^+\mu^+\mu^-$ candidates. Their primary use is providing information for particle identification to identify kaons and discriminate against pions, electrons and protons. The energy resolution $\sigma(E/p)$ of the ECAL as a function of energy as determined using electrons from converted photons is shown in Figure 3.7 [78].

3.3 Particle Identification

The particle identification is primarily performed by the RICH systems with complimentary information of the track momentum from the tracking stations. This section first discusses the RICH detector followed by its performance and how it is used to identify particles.

3.3.1 The RICH Detectors

The RICH system at LHCb is made of two sub-detectors, RICH1 and RICH2. A large number of electrons, muons, pions, kaons and protons are produced in heavy flavour decays such

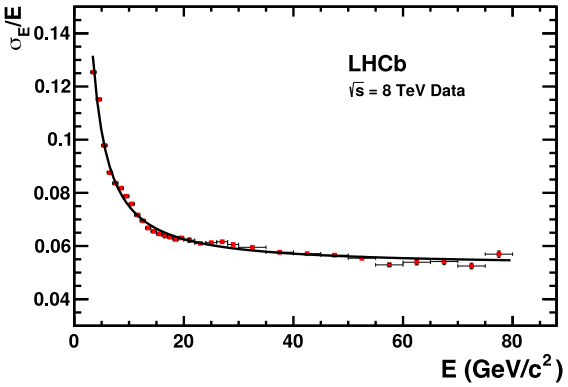

Figure 3.7: The energy resolution $\sigma(E/p)$ of the ECAL as a function of energy as determined using 2012 data of electrons from converted photons [78].

as decays involving the B meson. For this case it is important to be able to identify and select between them when reconstructing the invariant mass of the decaying particle This significantly helps reduce the background contributions. The RICH systems is vital to identify the kaon in $B^+ \to K^+\mu^+\mu^-$ decays used in this measurement.

The two RICH sub-detectors are designed to identify particles in a different momentum range. RICH1 works in the low and intermediate momentum range of 2 - 40 GeV/c and covers the angular acceptance of 25-300 mrad. RICH2 covers momentum ranges of 15-100 GeV/c with angular acceptance of 15-120 mrad. The choice of two detectors is due to the strong correlation between momentum and polar angle of the particles as higher momentum particles are produced at lower polar angles. RICH1 is the second sub-detectors in the chain, immediately downstream of the VELO. RICH2 is placed downstream of the magnet just after the tracking stations (see Figure 3.2).

When high energy particles move faster than the speed of light in a medium they produce a cone of light through a phenomenon called Cherenkov radiation [79]. The opening angle of the light cone depends on the refractive index of the medium and the velocity of the particle. In order to identify particles across a large range of momenta two RICH detectors are used with radiators of different refractive index. For RICH1, C_4F_{10} and an aerogel are used due to their sensitivity to low momentum tracks. In Figure 3.8 the Cherenkov angle as seen by RICH1 in the radiator C_4F_{10} is shown. RICH2 uses CF_4 as the radiating medium since it is better able to distinguish high momentum tracks. The refractive index of C_4F_{10} and CF_4 is 1.0014 and 1.0005 at 0°C, 101.325 kPa and 400 nm.

In order to reduce the amount of material in the acceptance region of the detector, both RICH detectors use a combination of flat and spherical mirrors to focus the ring images onto photo sensor planes. Hybrid Photo Detectors (HPD) which are sensitive to Cherenkov light in the wavelength range 200-600 nm are placed in regions of low magnetic field and low

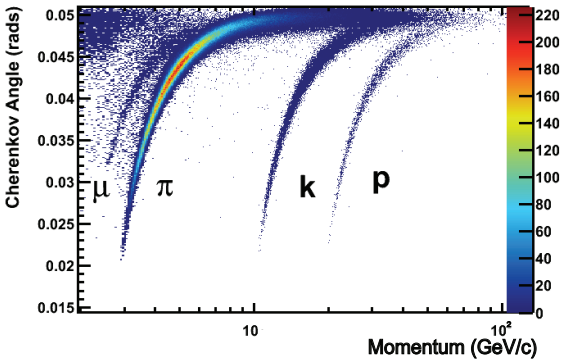

Figure 3.8: The dependence of the Cherenkov angle on momentum and identity of different particles for C_4F_{10} radiator [80].

radiation. Quartz windows are used to separate the HPD from the gas volumes.

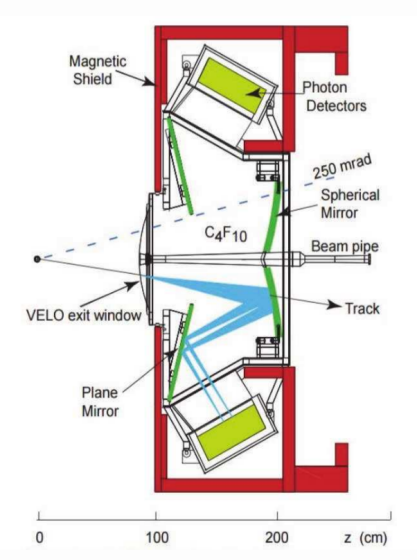

Figure 3.9: The layout of the RICH1 sub-detectors as viewed from the side [73].

The layout of RICH1 is shown in Figure 3.9. The silica aerogel [81] medium with refractive index 1.03 make up the 50 mm thick wall placed at the entrance to the sub-detector inside the gas enclosure. The gas enclosure is filled with C_4F_{10} gas and operated at room temperature and pressure. Four Carbon Fibre Reinforced Polymer (CFRP) spherical mirrors are used to deflect the light onto two sets of 8 flat glass mirrors, above and below the beam pipe which then focus the rings onto the photo sensor plane. CFRP was used to reduce the amount of material since the spherical mirrors are within the LHCb detector acceptance region. The kaons are required to have a minimum momentum of 9.3 GeV/c in order to produce Cherenkov light. Below this momentum, particles with no Cherenkov light would be identified as kaons instead of pions

(kaon veto mode). This refers to the configuration of identifying the track as a kaon track due to the lack of any associated Cherenkov light. The use of the aerogel helps to identify low momentum kaons better [80]. For Run2 data collection period, the aerogel was removed as its ability to provide kaon identification is compromised by the total number of photons in a high occupancy environment [82].

RICH2 in contrast to RICH1 is arranged horizontally with the HPD to the left and right of the beam pipe. It is filled with CF_4 gas with 5% CO_2 (increased to 10% in Run2 [82]) to reduce scintillation light and operated at room temperature and pressure. Since the material budget is higher because it is downstream, glass mirror substrates were used for both flat and spherical mirrors in RICH2 even though the spherical mirrors are within the LHCb detector acceptance.

3.3.2 Cherenkov angle resolution

The hits in the HPD are processed and data with HPD occupancy above 20% (average ~ 1%) are rejected to exclude excessively large events. The hits are reconstructed on the HPD plane to get the photon hit position correcting for HPD alignment, electrostatic focusing parameters and magnetic field. The information given by the tracking systems of LHCb gives the reconstructed trajectory of the particle within the sub-detector. The midpoint of this trajectory in the radiator is assumed to be the photon candidate emission point[3]. Using the combined information of the photon hit position on the HPD plane and the photon emission point, the analytic solution of the optical systems is used to calculate the Cherenkov angle θ_C.

The Cherenkov angle resolution for RICH1 and RICH2 gas radiators was found to be 1.618 $\pm$ 0.002 mrad and 0.68 $\pm$ 0.02 mrad respectively and in good agreement with simulation [80]. This value was not found to change for Run2 with the aerogel removed [82]. The aerogel resolution in Run1 was found to have an average value 5.6 mrad which is about 1.8 times worse than simulation. The discrepancy can partially be explained by adsorption of C_4F_{10} by the aerogel at its contact plane.

3.3.3 Particle Identification procedure and performance

This section describes the algorithm developed in [83] and described in [80]

Since the RICH detectors work in a high occupancy environment as seen in Figure 3.10, all the tracks in both detectors are considered simultaneously. This allows optimal treatment of tracks when Cherenkov cones overlaps. A log-likelihood algorithm is executed to distinguish between electron, muon, kaon, pion and proton hypotheses. This algorithm works by first assuming all tracks are pions since they are the most abundant particle produced in pp collisions at the LHC. An overall likelihood under this hypothesis is assigned to the event using information from the distribution of photon hits, the associated tracks and their errors.

[3] In reality, this could be anywhere along the track of the particle but works as a good enough approximation.

Next, in turn for each track, the likelihood is recomputed changing the mass hypothesis to e, μ, π, K and p in turn while the hypotheses for other tracks are kept unchanged. The mass hypothesis change for the track which gives the largest increase in the event likelihood amongst all tracks is identified and that mass hypothesis is set for that track. This is repeated until the four momenta of all tracks have been set to their optimal hypothesis and no further improvement is seen in the event likelihood.

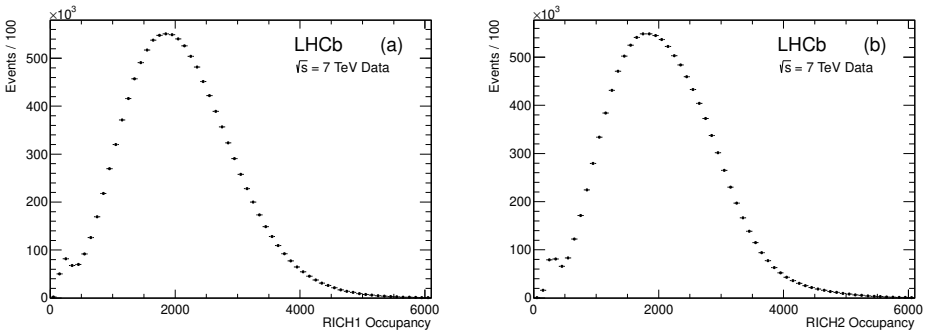

Figure 3.10: The distribution of number of pixels hit in (a) RICH1 and (b) RICH2 [80].

This procedure is clearly very CPU intensive and exponentially grows with the number of tracks in the event. Therefore some optimisation techniques are used in order to reduce the number of combinations without compromising the global solution. Firstly, when finding the track which gives the largest change in the event likelihood, all tracks are sorted according to largest increase in likelihood value. If the first track on that list shows a change in event likelihood above a certain threshold, the search is stopped and a mass hypothesis is assigned to that track. Secondly, if a track clearly shows a preference for a particular mass hypothesis then it is set to that hypothesis and removed from the next iteration. The background contribution to the likelihood is calculated by comparing the expected signal yield in the HPD and the observed signal. This information is included in the likelihood calculation.

This algorithm can be run several times until all the tracks are set to its optimal hypothesis and the overall event likelihood reaches convergence. Typically convergence is reached within two iterations of the algorithm. Finally for each track the difference between the log-likelihoods ($\Delta\log\mathcal{L}$) between the pion hypothesis and the electron, muon, kaon and proton hypotheses is computed. This information is combined with information from the muon systems to improve muon identification and is discussed in Section 3.4.

3.3.3.1 Performance of the particle identification

Calibration sample decays with π, K and p in the final state with defined topologies were chosen to measure the performance of the sub-detectors. Decays containing these particles in the final state were chosen such that they can be identified purely based on kinematical

selections, independent of the RICH. This include decays such as $K^0_S \to \pi^+\pi^-$, $\Lambda \to p\pi^-$ and $D^{*+} \to D^0(K^-\pi^+)\pi^+$. The selection of these samples have high purity and were chosen to closely match the momentum and pseudorapidity of physics events of interest to the experiment.

The kaon identification efficiencies and pion misidentification rate versus the particle momentum for Run1 are shown in Figure 3.11. The requirement of the log-likelihood of the kaon hypothesis to be greater than the pion hypotheses for each track ($\Delta\log\mathcal{L}(K-\pi) > 0$) is shown. Averaging over the momentum range 2-100 GeV/c, the kaon efficiency and pion misidentification fraction were found to be ~ 95% and ~ 10% respectively. This changes to 85% and 3% respectively when imposing a tougher requirement of $\Delta\log\mathcal{L}(K-\pi) > 5$.

The same set of curves for Run2 is shown in Figure 3.11. The calibration sample's range of pseudorapidity and momentum was increased to better match the interesting physics events [82]. It can be seen that pion and kaon separation performance has improved especially in the momentum range 2-20 GeV/c. After correcting for the change in beam energy and number of tracks per event differences between Run1 and Run2, no degradation in the performance was observed.

In addition to the likelihood method a more advanced classifier using a Neural Network (NN) is trained and the classification probabilities are stored in the offline variables `ProbNNx` (where `x = K,pi,`μ) The NN combines the information from the likelihood ratios and information with the kinematic variables from the tracking systems. It also uses information from the PID systems not used in the calculation of the likelihood ratio such as the geometrical acceptance of the RICH radiators. The full list of variables is discussed in Ref. [84]. The performance of the `ProbNNK` variable estimated using the 2018 data set is shown in Figure 3.12 [85]. A similar efficiency has also been observed for other run periods.

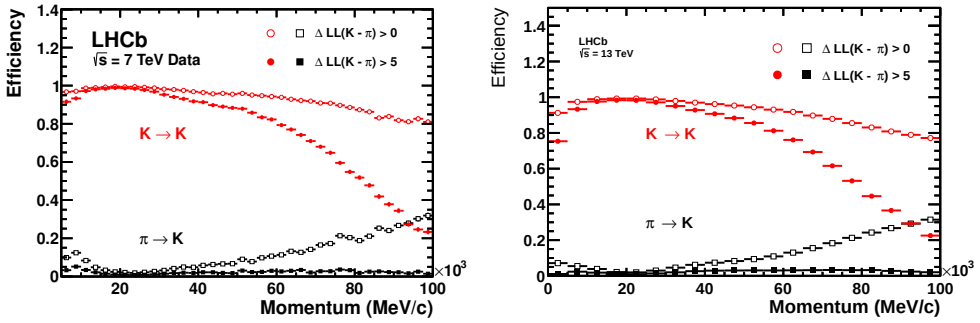

Figure 3.11: The kaon identification efficiency and pion mis-identification rate measured using calibration samples for Run1 on the left [80] and Run2 on the right [82].

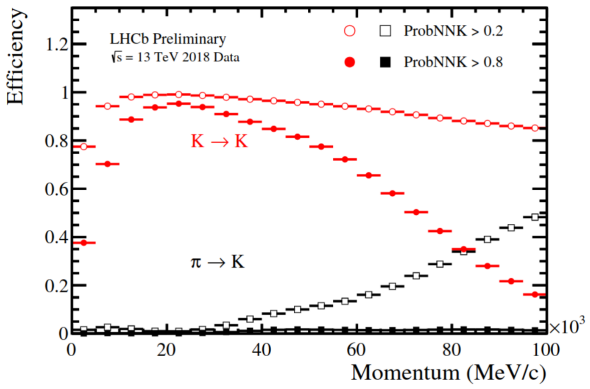

Figure 3.12: The performance of the NN based PID variable `ProbNNK` estimated using 2018 data. [85].

3.4 Muon System

The muon system is vital as it helps towards muon identification. In addition it also provides fast information for high-pT triggers for event selection and offline analysis. A sketch of the muon system is shown in Figure 3.13. It is made of 5 stations, M1 is upstream of the calorimeters and M2-M5 are downstream of it. Stations M2-M5 are interleaved with 80 cm thick iron absorbers to select penetrating muons. Muons with momentum above 6 GeV/c typically cross all 5 stations.

Each station is divided into four regions (R1-R4) with size increasing radially from the beam axis as shown in Figure. 3.13. The particle flux and channel occupancy are roughly the same in the four regions. Triple-GEM (Gas Electron Multiplier) detectors are used in the inner most region of the M1 station (M1R1) and Multi-wire propotional chambers (MWPC) are used for the rest of the system. Stations M2-M5 each have four active layers working in a logical OR configuration to improve the average signal efficiency. Station M1 only has two layers to minimize the material upstream of the calorimeters.

Each station is partitioned into rectangular logical pads whose dimension define the x,y resolution. Stations M1-M3 provide high spatial resolution in the x coordinates along the bending plane of the magnet. They define the track direction and help calculate the p_{T} of the muon candidate. Stations M4 and M5 are used mainly to identify penetrating particles. Each station is designed to perform with an efficiency of 99% in a 20 ns time window with a noise rate below 1 kHz.

3.4.1 Muon identification performance

A track is identified as a muon in three stages based on the association of hits of the extrapolated trajectory from the tracking stations in the muon system [86].

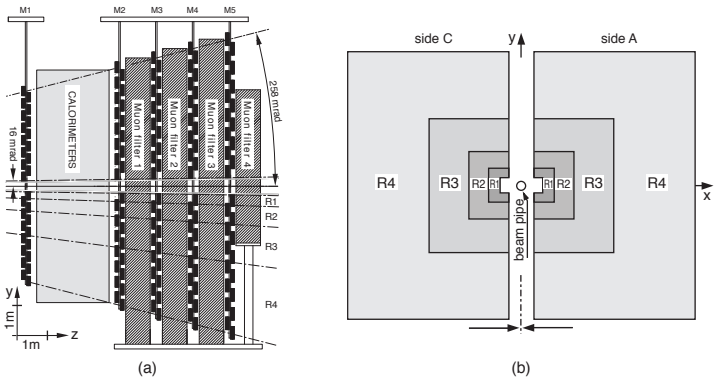

Figure 3.13: Illustrations of the muon stations as seen from the side (left) and along the beam direction (right) [86].

- A binary selection as a function of momentum based on the penetration of muons through the muon stations. The hits in the muon system are associated to extrapolated tracks from the tracking stations. This information is stored in the variable: `IsMuon`

- A likelihood is computed for muon and non-muon hypotheses based on the distribution of hits in the muon system around the track extrapolated from the tracking stations. The difference between log of the likelihood from the two hypotheses is used as the discriminating variable.

- A combined likelihood using information from the calorimeter and RICH is computed for different particle hypotheses. The difference between the logarithm of the muon and pion hypotheses is used as the discriminating variable.

The performance of muon identification is optimized to maximize the efficiency and minimize misidentification probabilities. The most common particles misidentified as a muon are protons, pions and kaons. Protons are incorrectly assigned as muons either due to a combination of random hits aligned with a proton track or a true muon which points in the same direction of a proton. The second scenario could be because the muon was produced very close to the interaction point or in the calorimeter shower. The main reason kaons and pions are misidentified are due to decays in flight.

3.4.1.1 Binary selection

The binary selection is passed if the field of interest around an extrapolated track contains a hit. The combination of stations required to have a hit to identify a track as a muon for different momentum ranges is shown in Table 3.1. Muon identification efficiency above 95% with misidentification kept at the level of 1-2% is achieved for the trigger and offline reconstruction

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

Table 3.1: The muon stations required to trigger the `IsMuon` decision for different momentum ranges.

using the combined response of all five stations [86]. This binary information is stored in the variable `IsMuon` for offline analysis.

3.4.1.2 Performance of likelihood discrimination

For each track, the average squared distance significance D^2 of the all the hits in all chambers within the extrapolated track's field of interest is calculated. A normalized distribution of D^2 using calibration samples of muons and non-muons is also calculated. The precise definition of D^2 is given in Ref. [86]. For muons, this distribution is expected to be narrower than of other particles.

An integral of the calibration distribution of the muon between 0 and the measured average squared distance significance for the track (D_0^2) is defined as the likelihood for muon hypothesis. For the non-muon hypothesis (pions, kaons and protons), the integral of the calibration distribution of the proton between 0 and D_0^2 is used. This is done as such because for the other charged hadrons the D^2 has a component identical to protons and another component similar to muons due to decays in flight. The difference of the log of the likelihoods between the two hypotheses then gives the discriminating variable `muDLL`. The D^2 and `muDLL` distribution for different hypotheses is shown in Figure 3.14.

The performance of muon identification can be improved by summing the logarithms of the likelihoods from the RICH systems and calorimeters with the `muDLL` calculated above. This gives a combined log-likelihood for each track for different mass hypotheses of electron, muon, pion, kaon and proton. The difference between the combined log-likelihoods for the muon and pion hypotheses is used to identify muons and is given by the variable $\mathrm{DLL}_{\mu\pi}$. Similar variables for the electron, kaon and proton mass hypotheses ($\mathrm{DLL}_{e\pi}$, $\mathrm{DLL}_{K\pi}$ and $\mathrm{DLL}_{p\pi}$) are also assigned[4]. Average muon misidentification rates of 0.65% and 0.38% with kaons and pions can be achieved using a cut[5] with an average decrease in efficiency of 5% [86]. The efficiency as a function of misidentification rates is shown in Figure 3.15.

[4]The non-muon likelihood from the muon systems is summed with the electron, pion, kaon and proton likelihoods from the other sub-detectors.

[5]'Cut' is a terminology commonly used in particle physics when a conditional requirement is placed on variables.

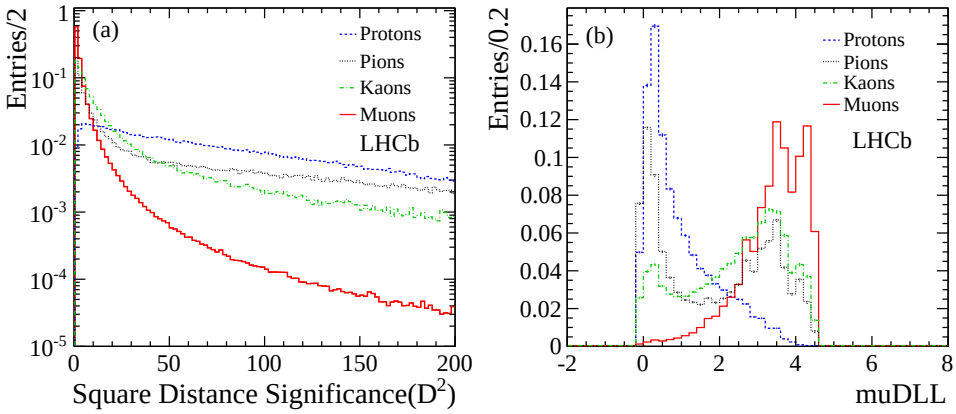

Figure 3.14: The average D^2 distribution (left) and the corresponding `muDLL` variable calculated from muon and a hadron hypothesis. For the charged hadrons, the calibration distribution of the proton is used to calculate the non-muon likelihood and `muDLL`. The other hadron distributions are shown here only for illustration. Figure taken from Ref. [86].

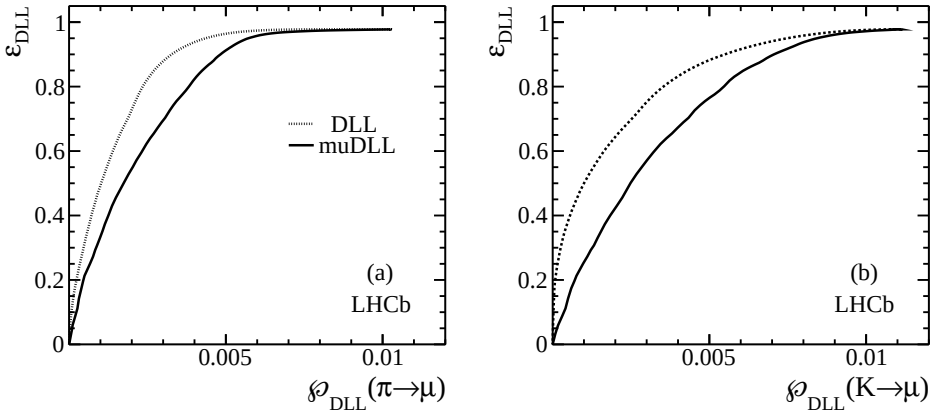

Figure 3.15: Average efficiency of $DLL_{\mu\pi}$ as a function of misidentification probabilities for pions (left) and kaons (right) along with `muDLL` for particles in the momentum range p > 3 MeV/c [86].

3.5 The Trigger System

At the LHC, during normal operations pp bunch crossings happens at the rate of 40 MHz. The LHC optics can be modified to control the rate of visible pp collisions for each bunch crossing. This rate was set to $\mu \approx 1.4$ [87], 1.6 [88] and 1.1 [89] in 2011, 2012 and Run2 respectively. The rate is controlled in order to not overwhelm the sub-detectors.

Only about 1% of the visible pp collision at LHC energies is expected to produce $b\bar{b}$ events. Of these events, about 15% is expected to produce a b-hadron whose decay products pass through the full acceptance of the detector. The LHCb trigger system is therefore designed to be able to select the interesting events efficiently while discriminating against background

events.

Most of the *b*- and *c*- hadrons studied at LHCb, including the B^+ meson which is of interest to this measurement have long lifetimes. This means that they have a long Flight Distance (FD) and their decay products leave a measurable Impact Parameter (IP) in the VELO as shown in Figure 3.3. They also have a large mass which leads to the final state particles having a significant transverse momentum p_{T}. These properties can be measured using information from all the sub-detectors to select events of interest.

The major constraints on the trigger system are the availability of computing/data storage resources and the need to reduce the data rate from 40 MHz to around 5-12.5 kHz. For this reason the trigger decision needs to be very quick and efficient in selecting the interesting events. In order to achieve this, the trigger system is split into two stages. The first is the hardware stage referred to as the Level-0 (L0) trigger which reduces the data rate to about 1 MHz (Section 3.5.1). And the second is the software stage referred to as the High Level Trigger (HLT) which is more flexible and further reduces the data rate to a manageable level (Section 3.5.2). The software stage of the trigger system was modified for Run2 to be able to better select *c*- and *s*- hadrons by enabling full reconstruction of tracks in the trigger system. The differences between the Run1 and Run2 trigger system are discussed in Section 3.5.3.

3.5.1 The hardware trigger

The hardware trigger, also known as the Level-0 (L0) trigger, reduces the data rate from 40 MHz to 1 MHz such that the whole detector can be read out. To achieve this, a decision must be made by the trigger in under 4 μs. The trigger uses information from the calorimeters, the muon systems and the pile-up sensors[6] housed in the VELO. The three components referred to as `L0Calorimeter`, `L0Muon` and `L0PileUp` are implemented using Field-Programmable Gate Arrays (FPGA).

Only the `L0Muon` trigger component is used to select the events used in the measurement made in this thesis. Multiplicity in the SPD is a good indicator of large occupancy in the OT and IT which leads to the use of disproportionately large fractions of the computing resources in the software stage of the trigger. For this reason, events with multiplicity above a threshold value are rejected.

The `L0Muon` trigger uses the muon stations M1-M5 and splits them into 4 quadrants with each quadrant connected to a L0 muon processor. Each processor looks for hits defining a straight line through the five stations. The 4 processors do not exchange information with each other so particles traversing the boundaries of quadrants are unable to be selected by the `L0Muon` trigger. The track position in the first two stations determines the p_{T}, a resolution of 25% relative to offline reconstruction is achieved by the trigger. Each processor then passes the two tracks with the largest p_{T} to the `L0Muon` trigger [87].

[6]This refers to the 4 additional sensors discussed in Section 3.1.1 placed in the front of the VELO

Two different algorithms are then employed by the trigger to select the event: `L0Muon` and `L0DiMuon`. Of the eight tracks, `L0Muon` selects the track with the highest p_{T}, if this value passes a threshold the event is selected for the next stage of the trigger. The value of the thresholds along with the SPD multiplicity thresholds can be found in [88] for Run1 and [89] for Run2 The thresholds are increased between run periods taking into account the increase in the pp collision rate.

The alternative algorithm, `L0DiMuon` requires the product of the two highest transverse momenta to be above a set threshold. Figure 3.16 shows the efficiency of the two trigger algorithms. The combination of the two algorithms offers only a marginal increase in efficiency as there is a significant overlap of events between them. Therefore only the `L0Muon` algorithm is used in this measurement.

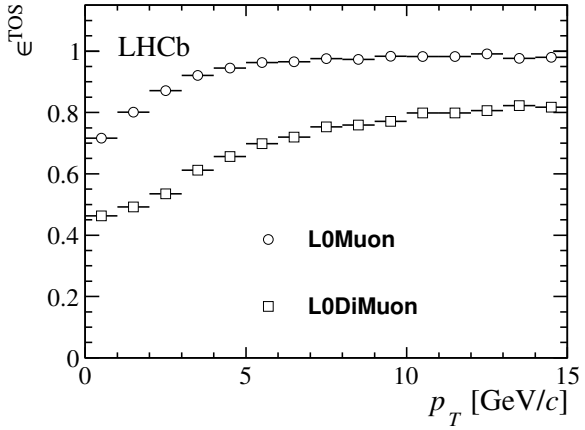

Figure 3.16: The selection efficiency of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ events as a function of p_{T} for `L0Muon` and `L0DiMuon` trigger catagories from Ref. [87].

3.5.2 The software trigger

The software based High Level Trigger (HLT) is further split into HLT1 and HLT2. HLT1 reduces the event rate of ~ 1 MHz output from the L0 to 40 kHz (80 kHz) in 2011 (2012) by performing a partial event reconstruction and inclusive selection of signal candidates. HLT2 then performs a full event reconstruction on this output and uses inclusive and exclusive selections to reduce the event rate to 3 kHz and 5 kHz in 2011 and 2012 respectively [88]. A particular sequency of event reconstruction algorithms and selections is commonly referred to as a Trigger Line. A trigger line returns a boolean accept or reject decision in order to select the event. Events which pass any of the HLT1 trigger lines and any of the HLT2 trigger lines are stored for offline analysis.

The software used to perform the event reconstruction is based on the same one used for LHCb data processing and simulation [90]. The full reconstruction done in the offline

processing takes about 2 s per event, this is too slow for the trigger. The methods employed to speed up the reconstruction for the trigger lines used to select the decay studied in this thesis along with their selection requirements and efficiencies are discussed below.

3.5.2.1 High Level Trigger 1 (HLT1)

The computing resources available allows the full event reconstruction in the VELO but not for the full detector. If a vertex within the VELO has at least 5 tracks originating from it and is within a radius of 300 μm of the PV_{xy}^{mean} it is considered to be a Primary Vertex (PV). The PV_{xy}^{mean} is determined from the VELO tracks at the start of each LHC fill and is measured to be stable up to a few μm.

To identify muon candidate tracks, for each track in the VELO, a search window is opened in the muon station M3. This search window is selected according to the bending angle of a track with momentum 6 GeV/c as this is the minimum required momentum of a muon candidate. Hits within this search window are combined with the VELO track to form candidate tracks and additional search windows are opened for hits in muon stations M2, M4 and M5. If at least one of these stations contain a hit, the track is provisionally accepted. Finally a linear χ^2 fit is performed to all the VELO and muon stations hits for the candidate tracks until the first track with $\chi^2/\text{ndf} < 25$ is identified. At this point the candidate track is labelled as a muon candidate and the algorithm stops. This track is passed onwards to the next stage of the reconstruction. Alternatively, since not all trigger lines require muon candidates, track segments with a large IP are also passed to the next stage.

Track segments in the OT and IT stations are reconstructed for the muon candidate tracks and tracks with large IP. A procedure known as forward tracking [91] is used for this. Imposing minimum momentum thresholds in the forward tracking procedure reduces the search windows needed and thereby reduces the processing time required for the reconstruction. A Kalman filter [92], [93] based fit is performed to the reconstructed tracks to obtain its χ^2 and covariance matrix. A simpler geometry and fewer iterations are used compared to the offline reconstruction. The loss due to this to the invariant mass resolution of $J/\psi \to \mu^+\mu^-$ events was found to be only 3% [87]. In addition, the muon identification algorithm [94] as in the offline analysis is used on the muon sample to improve its purity. This allows the use of selective cuts in IP, mass and momentum.

HLT1 lines used in the selection of $B^+ \to K^+\mu^+\mu^-$ decays are `Hlt1TrackMuon`, `Hlt1DiMuonHighMass`, `Hlt1DiMuonLowMass` and `Hlt1TrackAllL0`. The latter selects tracks based on a minimum IP > 0.1 mm and $p_T > 1.6$ GeV/c requirement and the former three require their tracks to be muon candidates. The `Hlt1TrackMuon` line selects muon tracks similar to `Hlt1TrackAllL0` and has the requirements IP > 0.1 mm and momentum $p_T >$ 1 GeV/c. The `DiMuonHighMass` line selects tracks based on an invariant mass requirement of $m_{\mu\mu} > 2.5$ GeV/c^2 without any IP requirements and compensates for the low efficiency of the

`Hlt1TrackMuon` line at low lifetimes of the decaying particle. The `Hlt1DiMuonLowMass` line selects tracks with a lower mass requirement of $m_{\mu\mu} > 1\,\text{GeV}/c^2$. A full list of requirements can be found in Ref. [87]. Efficiencies of the lines as a function of p_T and B^+ lifetime calculated from $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ decays in 2011 is shown in Figure 3.17.

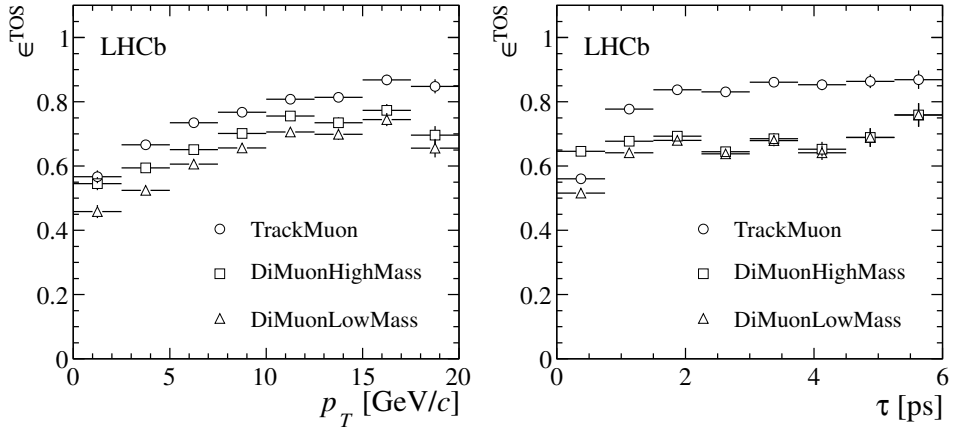

Figure 3.17: The selection efficiency of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ events as a function of p_T (left) and B^+ lifetime τ (right) for various HLT1 trigger lines in 2011 from Ref. [87].

3.5.2.2 High Level Trigger 2 (HLT2)

For all the tracks with $p_\text{T} > 500\,\text{MeV}/c$ (lowered to $300\,\text{MeV}/c$ for 2012) the full reconstruction is performed and muon identification is performed on all tracks. While two complementary reconstruction algorithms are used in the offline reconstruction, only one is used in High Level Trigger 2 (HLT2) to limit the processing time required. This lowers the efficiency by 1-2% per track. HLT2 has two types of trigger lines implemented, the topological lines and exclusive lines.

The topological lines use a combination of 2, 3 or 4 special tracks referred to as a Topo-Track and a multivariate analysis (MVA) classifier to select events. These lines are relatively more sophisticated and select a large share of the 3 kHz and 5 kHz output rate for 2011 and 2012 respectively. A Topo-Track is a HLT2 track with additional requirements on track fit quality (χ^2/ndf), IP and muon and electron identification. Two Topo-Track are combined to form a 2-body object with the requirement that their distance of closest approach (DOCA) is less than 0.2 mm. A 2-body track is combined with another Topo-Track to create a 3-body track with the same DOCA < 0.2 mm. And similarly a 4- body track is created with a 3-body track and a Topo-Track.

The topological lines do not require all the final state particles of the b-hadron to be reconstructed in the n-body track. In order to remain inclusive and account for the particles

not reconstructed, a variable called the corrected mass m_{corr} is introduced [87].

$$m_{corr} = \sqrt{m^2 + |p'_{Tmiss}|^2} + |p'_{Tmiss}|, \tag{3.1}$$

where $|p'_{Tmiss}|$ is the missing momentum transverse to the FD. To select events, the invariant mass m and the corrected mass m_{corr} of the n-body candidate are used. In addition, the sum of p_{T} of all the tracks and the minimum p_{T} along with the IP, DOCA and the FD significance are used.

It was found that using a MVA classifier in addition to simple selection requirements better rejects background. This can be achieved without compromising the signal efficiency using a boosted decision tree (BDT) [95, 96]. A BDT works by classifying regions in the n-dimensional multivariate space using labeled training data. The signal regions in this space can be selected using the classification variable returned by the BDT.

In order to be fast enough for the trigger, a modified version of this method referred to as the bonsai boosted decision tree (BBDT) is used at LHCb [97]. Here the variables are discretized and the BDT is trained in this discrete space. Thresholds are placed on the response of the BBDT to select candidates. The 2-body algorithm (`Hlt2Topo2BodyBBDT`) is evaluated first followed by the 3- and 4- body algorithms by adding one Topo-Track at a time. The choice of the discrete intervals used and the training strategy of the BBDT is described in Ref. [87].

A total of four topological algorithms for the 2- and 3-body topologies are used to select $B^+ \to K^+\mu^+\mu^-$ candidates. These are `Hlt2TopoNBodyBBDT, Hlt2TopoMuNBodyBBDT` with `N=2,3` where the latter refers to the n-body topologies with one or more muon candidates. The efficiencies of the topological lines as a function of p_{T} and B^+ lifetimes are shown in Figure 3.18.

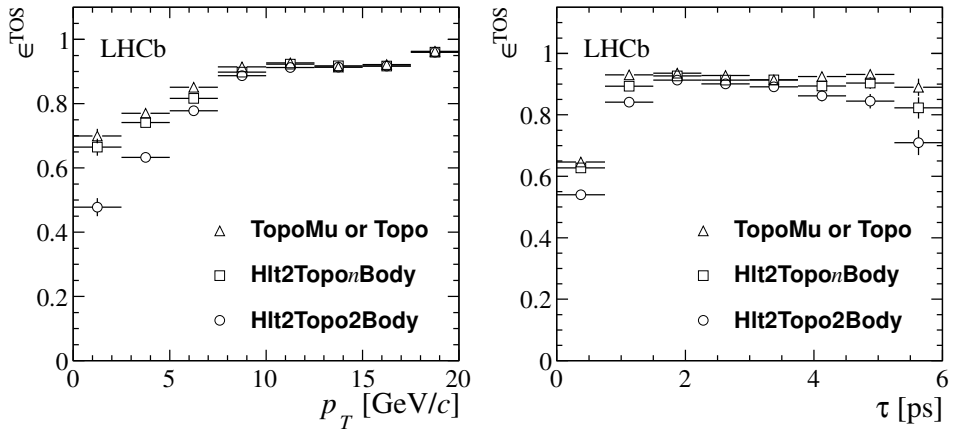

Figure 3.18: The selection efficiency of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ events as a function of p_{T} (left) and B^+ lifetime τ (right) for various HLT2 trigger lines in Run1 from Ref. [87]. The efficiency is measured relative to events which have been triggered using `Hlt1TrackAllL0` or `Hlt1TrackMuon`.

The other type of HLT2 trigger lines which exist are the exclusive lines. The one used in this thesis are `Hlt2SingleMuon`, `Hlt2DiMuonDetached` and `Hlt2DiMuonDetachedHeavy` These lines select events based on one or two identified muons in the final state. The single muon line selects tracks with minimum requirements on the IP, IP χ^2, p_T and an upper limit on the χ^2/ndf. The rate of this line is scaled down by a factor of 2 to reduce the rate, events were selected at the rate of 480 Hz in 2011 by this line. The other two exclusive lines are based on two identified muons. These lines use the separation of the dimuon vertex and the PV as the main discriminant. `Hlt2DiMuonDetached` selects low mass muon pairs while `Hlt2DiMuonDetachedHeavy` selects high mass muons with a relaxed FD criteria.

Stable beam conditions occurs only 30% of the time at the LHC. For this reason ~ 20% of the L0 selected events are stored temporarily and processed during beam downtimes. This method increases the effective computing power available and allows improved track reconstruction and get better efficiencies [98].

3.5.3 The changes to the software trigger implemented for Run2

This section follows Ref. [89] unless stated otherwise.

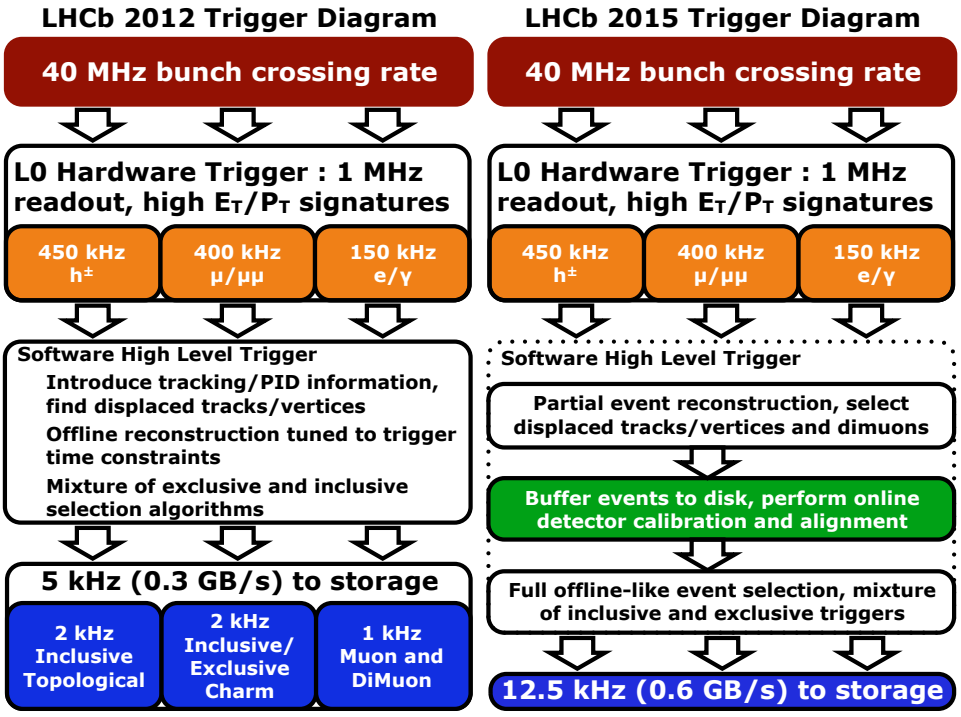

Figure 3.19: The dataflow diagrams of the Run1 (left) and Run2 (right) trigger systems [99]

For Run2 data taking period major improvements where implemented to enable offline quality reconstruction in real time, at the trigger level. This allows for fully optimised Particle

Identification (PID) to select interesting physics events. The flow diagram of Run1 and Run2 trigger systems is shown in Figure 3.19. The hardware-based L0 trigger largely remained unchanged except for the thresholds, optimizing it each year to the pp collision rate. The software trigger system was modified by the introduction of a 10 PB buffer between HLT1 and HLT2. At the HLT1 rate of 110 kHz and average event size of 55 kB, the buffer allows the storage of two consecutive weeks of data taking before HLT2 needs to be executed. The computing resources available are also increased in Run2. With this, the full alignment and calibration of the detector could be performed before HLT2 enabling the complete offline-quality reconstruction at the trigger level.

The software framework used for the Real Time Analysis (RTA) project is discussed in more detail in Ref. [100]. This forms part of the study to move to a fully software based trigger for the next run period (Run3) [101, 102].

3.5.3.1 High Level Trigger 1 (HLT1) in Run2

In Run2, the track reconstruction is performed in three main steps. The tracks are reconstructed within the VELO first by fitting hits with a straight line loosely pointing towards the beam line using a simplified Kalman filter [103]. This track is then extrapolated to the TT where at least three hits are required within a small region around the extrapolated line to form 'upstream tracks' [104]. Finally search windows are opened in the IT and OT depending on the deflection produced by particles with $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$. Search windows are restricted using a charge estimate[7] of the upstream track. This charge estimate allows lowering the p_{T} requirement from 1200 MeV/c to 500 MeV/c between Run1 and Run2. As in Run1, clusters of hits within the search window are combined with the upstream track to define the final long track [91]. All the tracks are the fitted with a Kalman filter [92], [93] to obtain estimates of the track. Both online and offline algorithms are setup identically in Run2.

The PV position is important to select displaced particles. In Run2, the PV vertex reconstruction is done using VELO tracks only. When compared to finding the PV with VELO and long tracks, it was found that the resolution does not degrade. This also has the additional advantage of reducing systematic uncertainties by using a consistent PV throughout the analysis chain. The simpler Kalman filter used also has a more accurate PV covariance matrix compared to Run1 offline analysis.

While the computing resources available does not permit full particle identification, muon identification is performed by HLT1. The algorithm to identify muons is the same used in HLT1, HLT2 and offline as discussed in Section 3.4. This muon identification procedure is only performed on fully fitted tracks with $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$. To identify muons with a lower transverse momentum a complementary algorithm similar to the Run1 HLT1 muon

[7] The charge can be estimated since the TT is in the fringe field of the dipole magnet. This additionally also allows the rejection of low p_{T} upstream tracks.

identification algorithm is used. Upstream tracks are extrapolated to the muon stations and matched with hits within search windows. This extends the muon identification down to tracks with $p_T > 80$ MeV/c and improvement in the efficiency is found in Run2 compared to Run1. The efficiency of the HLT1 muon trigger lines as a function of p_T and B^+ lifetime calculated using $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ data is shown in Figure 3.20. The lines used in this analysis are the logical or between `Hlt1TrackMuon`, `Hlt1TrackMVA`, `Hlt1DiMuonLowMass` and `Hlt1DiMuonHighMass` which follow the same definitions from the Run1 trigger system.

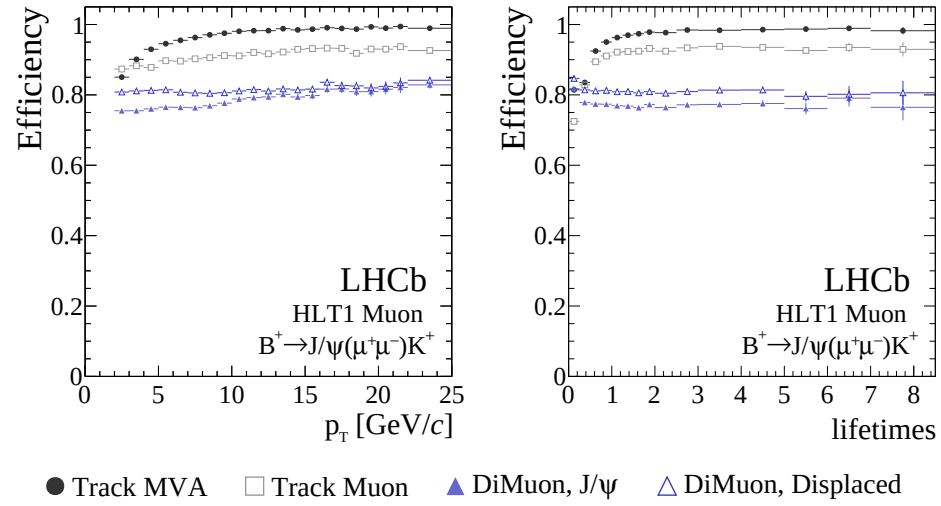

Figure 3.20: The selection efficiency of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ events as a function of p_T (left) and B^+ lifetime (right) for the different HLT1 trigger lines in Run2 from Ref. [89]. The lifetime plot is binned in the x-axis in units of the B^+ lifetime in its rest frame.

3.5.3.2 High Level Trigger 2 (HLT2) in Run2

The Run2 HLT2 lines benefits from the in-situ alignment and calibration of the detector which allows better track reconstruction and improved PID. The major tasks for this are the real time alignment of the VELO modules, tracking stations, the mirrors in the RICH and muon stations. In addition: the global time of the OT, gas refractive index and the photo detectors of the RICH systems and ECAL are calibrated. Each of these tasks have a dedicated HLT1 trigger line to get the relevant data. More information on the alignment and calibration procedure can be found in Ref. [105].

The track reconstruction algorithms were also improved which results in a two times faster execution time and better performance than Run1. HLT2 performs track reconstruction of tracks with $p_T > 500$ MeV/c just as in HLT1 as the first step. Then all VELO tracks are extrapolated to the IT and OT and combined with hits within search windows. The search windows in the second step account for bending tracks with $p_T > 80$ MeV/c. As a complementary method, standalone tracks in the OT and IT[106] are combined with standalone tracks in the VELO to

form long tracks [107]. Using two complementary algorithms improves the efficiency by a few percent.

As the next step in track reconstruction, fake tracks formed by random combination of hits and mismatch of upstream and downstream tracks are removed. This is done using a Kalman filter and a neural network trained using the TMVA package [108], [109] and is described in Ref. [110]. Finally the tracks are filtered to remove any clones which could have originated from the redundancy of algorithms.

The muon reconstruction procedure is the same are HLT1 except that HLT2 uses the full set of fitted tracks and not just the tracks with $p_{\mathrm{T}} > 500$ MeV/c. The reconstruction in the RICH detectors provides the main PID capability for the tracks. All the reconstructed tracks and all Cherenkov photons within both the RICH detectors are used in the algorithm.

The topological lines (discussed in the Run1 implementation) benefit from the improved track reconstruction and PID information available. Taking this into account the selection criteria has been re-optimized [111] and the MVA algorithm is used to discriminate signal and background. The performance of the topological lines as a function of p_{T} and B^+ lifetime is shown in Figure 3.21. The four topological lines used in Run1 are also used in Run2. The exclusive muon and dimuon lines are also the same as in Run1 and benefit from the improved muon identification procedure.

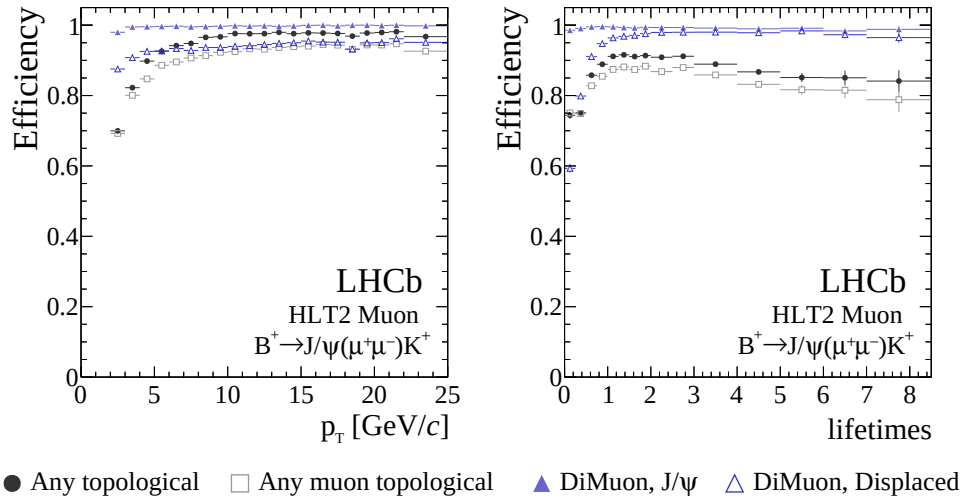

Figure 3.21: The selection efficiency of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ events as a function of p_{T} (left) and B^+ lifetime τ (right) for various HLT2 trigger lines in Run2 from Ref. [89]. The lifetime plot is binned in the x-axis in units of the B^+ lifetime in its rest frame.

3.5.4 Summary of the trigger system

The trigger lines used for this measurement to select $B^+ \to K^+\mu^+\mu^-$ candidates in the three different stages for Run1 and Run2 are shown in Table 3.2. The `DiMuonDetached` and

DiMuonDetachedHeavy HLT2 lines were omitted for the 2016 data set due to a known bug in the software. Between Run1 and Run2 the concept behind all the trigger lines have remained the same except for some performance improvements. The names of some lines have also been updated.

Level	Requirement	
	Run1	Run2
L0	L0Muon	L0Muon
HLT1	TrackMuon or TrackAll or DiMuonLow or DiMuonHigh	TrackMuon or TrackMVA or DiMuonLowMass or DiMuonHighMass
HLT2	TopoMu2BodyBDT or TopoMu3BodyBDT Topo2BodyBDT or Topo3BodyBDT or SingleMuon or DiMuonDetached or DiMuonDetachedHeavy	TopoMu2Body or TopoMu3Body or Topo2Body or Topo3Body or SingleMuon or DiMuonDetached* or DiMuonDetachedHeavy*

Table 3.2: The summary of the trigger lines used in this measurement to select $B^+ \to K^+\mu^+\mu^-$ candidates. *The DiMuonDetached and DiMuonDetachedHeavy were omitted when selecting candidates in the 2016 data set due to a known bug in the software.

3.6 Simulation

Simulated events are vital to help us understand the behaviour of the detector and its efficiency on the dimuon mass distribution. As will be described in Section 4.3, simulated $B^+ \to K^+\mu^+\mu^-$ candidates are used to train MVA classifiers to discriminate against combinatorial background. The simulated events used in this analysis has been centrally generated by the LHCb collaboration. The pp collisions are generated using PYTHIA 8 [112] with a LHCb configuration [113] and the B^+ decay is described by EVTGEN [114]. The Final State radiation (FSR) effects are generated using PHOTOS [115]. The GEANT4 toolkit [116] is used to describe the detector geometry. It also simulates the response of the detector to the propagation of the generated particles through the detector geometry as described in detail in Ref. [117]. The simulated events then experience the same reconstruction and selection procedure as the collected data. The simulated data sets are also sometimes referred to as MC in literature. In order to account for the small modelling differences of the detector occupancy, B^+ momentum vertex quality, trigger and PID, data-driven corrections are applied. This procedure is explained in Section 5.1. To perform the measurement discussed in this thesis four decays are simulated. These are the rare $B^+ \to K^+\mu^+\mu^-$ decays, the resonant $B^+ \to K^+ J/\psi(\mu^+\mu^-)$, $B^+ \to K^+\psi(2S)(\mu^+\mu^-)$ and $B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$ decays. The $B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$ are necessary to estimate the candidates where the π is misidentified as a K and is discussed further in Section 6.2.4.

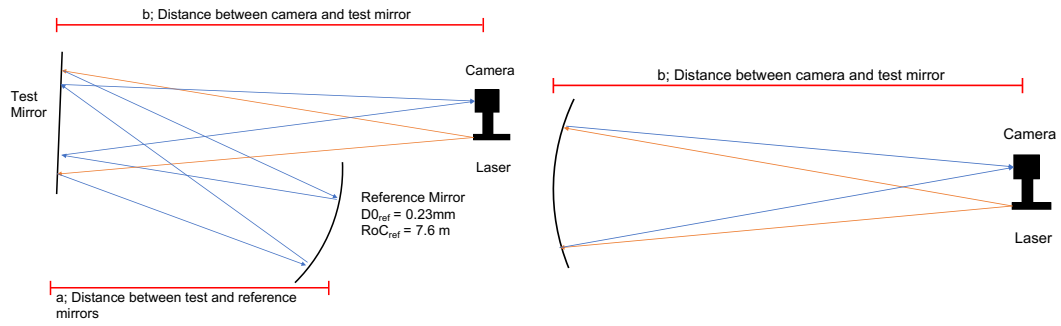

Figure 3.22: A schematic diagram of the set up used to test the flat (left) and spherical (right) mirror substrates for the upgrade of the RICH 1 sub-detector.

3.7 Upgrading the RICH-1 Sub-detector

As a part of activities carried out during the LHC shutdown period between Run2 and Run3 most of the LHCb detector was upgraded [118]. For the RICH 1 sub-detector: the optical components, the magnetic shielding, the gas enclosure and photon-detectors have been modified or replaced [119]. The optical components include the 4 spherical carbon fibre-based mirrors and 16 flat glass-based mirrors used to reflect and focus the Cherenkov light onto the photon detector plane.

The mirrors have been replaced as they suffer from radiation damage. In addition, the radius of curvature of the spherical mirrors needed to be modified to accommodate the new detector geometry. The mirrors were manufactured by third parties and had to be validated for quality assurance; this work will be highlighted as an aside in this chapter.

The 16 (+6 spare) flat mirror substrates were manufactured by Palacký University, Olomouc and are required to have a Radius of Curvature (RoC) > 90 m. The 4 spherical mirror substrates were manufactured by CMA, Arizona and are required to have a RoC of 3650 ± 10 mm to satisfy the geometric requirements of the gas enclosure. Both type of mirrors were also required to have a D0 below 2 mm. The D0 is defined as the diameter of the circle containing 95% of the reflected light from a point source. This value was motived by the pixel size of $2.88 \times 2.88\,\text{mm}^2$ of the photon detectors.

3.7.1 Measuring RoC and D0 of the Mirror Substrates.

The laser used is a diverging point source and is fixed relative to the camera's CCD which has a pixel width of 6.4×10^{-6} mm and is operated with a lens system. The camera-laser system is mounted on a movable platform which can move towards or away from the test mirror. The test setup for a flat and spherical mirror substrates is shown in Figure 3.22

For the flat substrates, a reference mirror is needed to focus the light onto the camera. The reference mirror has the know properties of $D0_{ref} = 0.23$ mm and $RoC_{ref} = 7.6$ m and is used to

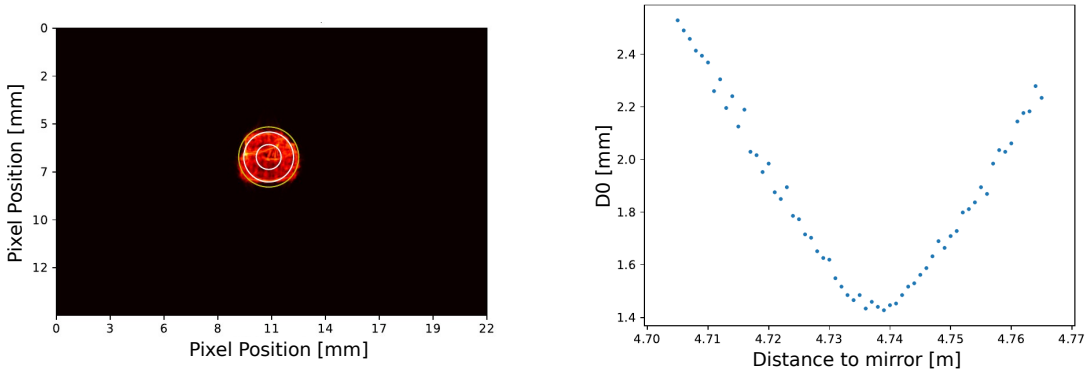

Figure 3.23: The image of the reflected laser spot at the focus distance (left) and the results of the D0 measurement at various distances between the mirror and camera (right). The example results shown here correspond to the flat mirror substrate tagged as OC083.

refocus the light back to the camera-laser system. This means that the light is doubly reflected by the test mirror and to account for this the measured D0 is corrected using

$$D0 = 0.5(D0_{meas} - D0_{ref}). \tag{3.2}$$

The camera-laser system is moved until the spot size seen by the camera is at its smallest, this means the reflected light is properly focused. This image is then used to calculate the D0 and RoC of the mirror. By measuring the distance between the camera and the test mirror, the RoC can be calculated using

$$RoC = \frac{2b}{1 - \frac{b}{RoC_{ref} - a}}. \tag{3.3}$$

Here a is the distance between the reference mirror and the test mirror, and b is the distance between the test mirror and camera's CCD.

Since no reference mirror is needed for the spherical mirror, the measured D0 is the value of the mirror. The RoC is also simply 2 times the distance between the camera's CCD and the mirror, RoC= $2b$.

To find the smallest spot size, multiple images are taken at differing distances between the camera and the mirror. Before measuring the D0, understanding and accounting for the background is needed. This is done in two steps, the first step reduces the global background arising from misfired pixels and the resting activation value of the pixels by using a ensemble of dark images taken just after collecting the D0 images. The ensemble is then used to calculate the average value of each pixel which is subtracted from the laser images. The second step runs the D0 measurement algorithm which integrate pixels in concentric circles centered at the center of mass of the image. Once the algorithm finds the largest circle containing 95% of

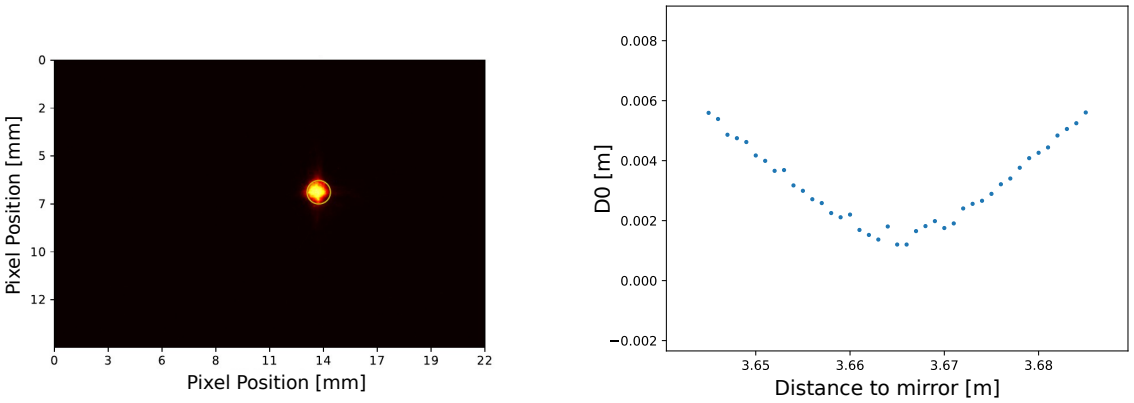

Figure 3.24: The image of the reflected laser spot at the focus distance (left) and the results of the D0 measurement at various distances between the mirror and camera (right). The example results shown here correspond to the spherical mirror substrate tagged as CMA3.

the signal, the pixels outside this circle are summed up together and averaged. This value is then subtracted from the entire image. This step ensures any residual ambient light present when the image was taken is accounted for. The D0 measurement algorithm is run again on the processed images and the image with the smallest $D0_{meas}$ is selected to be the focused image.

For the flat mirrors, Eq. 3.2 is used to calculate the D0 of the mirror and Eq. 3.3 is used to calculate the RoC. For spherical mirrors, $D0_{meas}$ is taken to be the D0 of the mirror and the RoC is calculated from the distance between camera and mirror. Example plots for flat and spherical mirror substrates at varying camera distances and the spot at the focal point are shown in Figures 3.23 and 3.24 respectively.

For each flat mirror, different positional arrangements of the mirrors were used to check the consistency of the results. These include moving the mirror to the left and right and inverting it upside-down before repeating the measurement. In all cases the results were within the required specifications, the summary of the results is shown in Figure 3.25.

The results for the tested carbon fibre-based spherical mirrors are shown in Figure 3.26. A good D0 within requested limits was achieved by all mirrors. While the radius of curvature was above the required tolerance by about 0.06 m, the good D0 result motivated the decision to accept them. Since the mirror substrates showed satisfactory results, the mirrors were then coated by the thin films group at CERN to achieve high reflectivity and then prepared for installation.

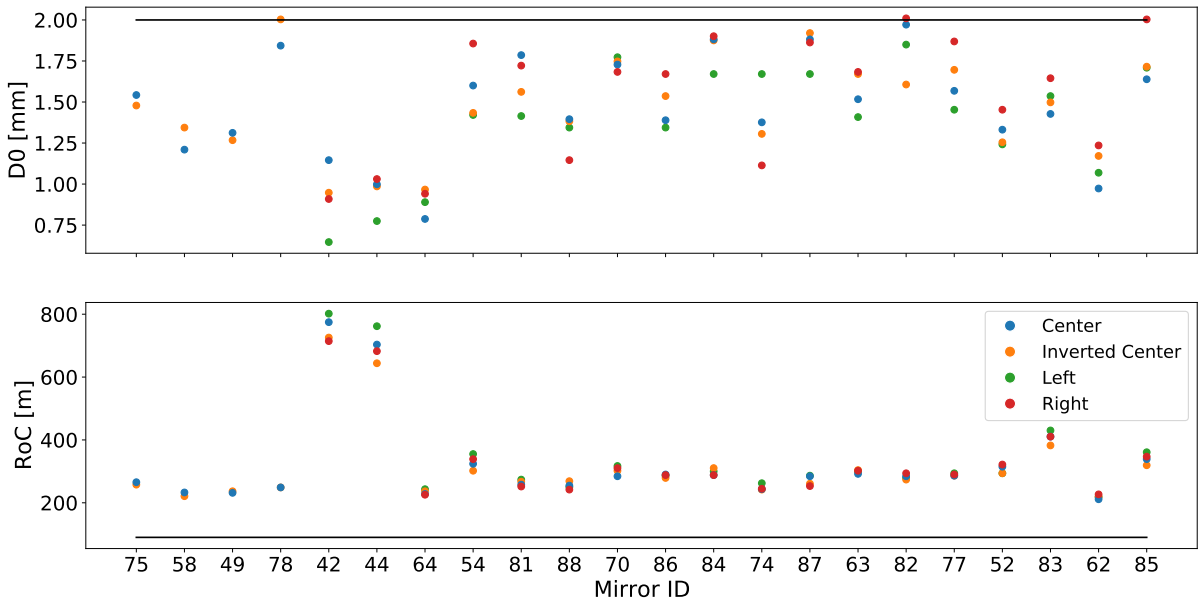

Figure 3.25: The consolidated summary of the quality assurance tests measuring the D0 (top) and RoC (bottom) for the flat mirror substrates. All tested mirrors were deemed to be within requested tolerance limits.

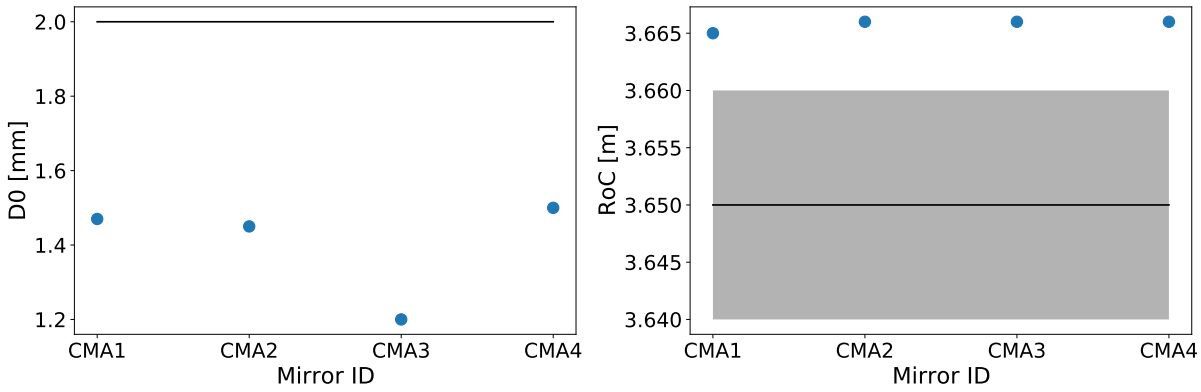

Figure 3.26: The consolidated summary of the quality assurance tests measuring the D0 (left) and RoC (right) for the spherical mirror substrates. The tolerance for the RoC is given by the gray band. All tested mirrors were accepted.

An Overview of the Analysis

The measurement presented in this thesis studies the dimuon spectrum of $B^+ \to K^+\mu^+\mu^-$ decays in order to determine any NP contribution. The local and non-local contributions are determined from data using the model developed in Section 2.4. The data collected at LHCb contains a plethora of particle decays. The method used to select $B^+ \to K^+\mu^+\mu^-$ candidates after removing combinatorial and misidentified background is discussed in Chapter 4. In Chapter 5 methods to account for the detectors effects such as the efficiency and resolution of the detector are described. These effects are calculated from simulation after correcting for differences between data and simulation.

Any remaining background contributions are estimated and a background model is developed in Chapter 6. The background model is added to the physics model to create the fit model and is validated. The results from the fit to Run1+Run2 data and a study of the systematic uncertainties is presented in Chapter 7. A phenomenological study of the τ scattering amplitude is also presented in Chapter 7 before concluding and highlighting the future work in Chapter 8.

CHAPTER

4 Selecting Candidates

The trigger requirements discussed in Section 3.5 perform crude selections to reduce the data size and save the information to disk. The algorithms are sophisticated enough to identify and select candidates with interesting kinematics. However, the collected data set includes a plethora of decay channels. Using the reconstructed track information, the data set needs to be further reduced and selected for genuine $B^+ \to K^+\mu^+\mu^-$ candidates. This process is done in a few steps.

The first step is referred to as "Stripping" within LHCb and the requirements are summarized in Section 4.1. The second step places requirements on the Particle Identification (PID) information which suppresses combinatorial and specific background. The requirements placed on the kaon and muon candidates are summarized in Section 4.2. The final step of data selection is to reduce the combinatorial background while keeping as much of the signal candidates as possible. The combinatorial background is formed by a random combination of tracks which happen to pass all the selections. The kinematics of combinatorial background and signal candidates are different and a MVA classifier is used to exploit these differences to efficiently discriminate between them. This procedure is detailed in Section 4.3.

The data set for Run1 was inherited from the previous iteration of this measurement, after the application of all necessary selection requirements. Therefore this chapter only discusses the selection procedure of data collected during Run2 of the LHC. More information regarding the selection of Run1 can be found in Ref. [12]. The selection requirements have been kept similar between the two data sets.

Candidate	Requirement
General	SPD Total hits < 600
B meson	IP χ^2 < 16 (best PV)
B meson	4700 MeV/c^2 < $m_{K\mu\mu}$ < 7000 MeV/c^2
B meson	DIRA < 14
B meson	FD χ^2 > 121
B meson	Vertex χ^2/ndf < 8
Dimuon	Vertex χ^2/ndf < 9
Dimuon	$m_{\mu\mu}$ < 7100 MeV/c^2
All Tracks	Ghost Probability < 0.4
All Tracks	IP χ^2 (minimum) > 9
Muon	`IsMuon` == True
Muon	$DLL_{\mu\pi}$ > -3

Table 4.1: A summary of the selection requirements imposed on candidate events to filter genuine $B^+ \to K^+\mu^+\mu^-$ events in.

4.1 Data Filtering (Stripping)

This stage of the data selection is done to reduce the data set to a manageable size while retaining potential $B^+ \to K^+\mu^+\mu^-$ candidates. For this purpose, the kinematic and topological properties of the reconstructed tracks are required to be consistent with that expected from a genuine $B^+ \to K^+\mu^+\mu^-$ candidates. These requirements are explained below, and are summarized in Table 4.1.

SPD Total hits - *General Requirement*

The number of hits in the Scinitillating Pad Detector (SPD) is required to be lower than 600 in order to remove events with high detector activity. High detector activity disproportionately uses a larger fraction of computing resources during reconstruction and this selection removes such events.

IP χ^2 (best PV)- *B meson*

The difference in χ^2 of the best Primary Vertex (PV) reconstructed including and excluding the reconstructed B meson candidate is required to be less than 16.

Three body mass $m_{K\mu\mu}$- *B meson*

The reconstructed mass of the B meson candidate $m_{K\mu\mu}$ is required to be between 4700 MeV/c^2 < $m_{K\mu\mu}$ < 7000 MeV/c^2. This selection also reduces the data to a manageable level. With the mass of the B meson being 5279.34 ± 0.12 MeV/c^2 [55], this range aims to ensure all the signal candidates are retained. It also retains some mass regions dominated by combinatorial

background which will be used in Section 6.2.3 to determine the combinatorial background under the signal peak.

DIRA - *B meson*

DIRA is the angle between the momentum vector of the B meson and the directional vector between the PV and the decay vertex (FD vector) of the B meson. This is required to be below 14 mrad.

Flight Distance χ^2- *B meson*

The Flight Distance (FD) χ^2 is the change in the vertex fit χ^2 when including the B meson child candidates. The decay vertex formed by the three final state particles should have a large FD χ^2 when compared with the PV as this implies that the B meson has a large FD. This parameter is required to be > 121.

Vertex χ^2/ndf - *B meson*

The χ^2_{vtx}/ndf for the B meson is the vertex fit χ^2 divided by the number of degrees of freedom (which is three) and is required to be below 8.

Vertex χ^2/ndf - *Dimuon system*

A requirement of χ^2_{vtx}/ndf to be less than 9 similar to the B meson is also given to the dimuon system.

Two body mass $m_{\mu\mu}$- *Dimuon system*

The dimuon mass $m_{\mu\mu}$ is required to be < 7100 MeV/c^2. This is a very generous requirement to select dimuon candidates.

Ghost probability - *Tracks*

A track is referred to as a ghost track if it is formed by a false combination of hits not belonging to the same particle. The probability of a track being a ghost track is calculated by the ghost probability algorithm which is detailed in Ref. [120]. This probability is required to be < 0.4.

min IP χ^2-*Tracks*

For the reconstructed tracks of the final state particles, IP χ^2 is the difference in χ^2 of the reconstructed PV including or excluding the track. Most of the combinatorial $K^+\mu^+\mu^-$ candidates arise from the PV and can be removed by requiring the minimum IP χ^2 of the tracks to be large. This is required to be > 9.

Particle	PID requirement
Kaons	(ProbNNK - ProbNNpi)>-0.1 and ProbNNK > 0.2
Muons	ProbNNmu > 0.25

Table 4.2: The PID requirements imposed on reconstructed data to select $B^+ \to K^+\mu^+\mu^-$ candidates. These requirements are identical between Run1 and Run2.

IsMuon - *Muon*

This requirement refers to the boolean information discussed in Section 3.4.1 which identifies if a track forms a pattern of hits in the muon system compatible with that of a muon and is required to be True.

$\text{DLL}_{\mu\pi}$ - *Muon*

$\text{DLL}_{\mu\pi}$ is the difference in the log-likelihoods between the muon and pion hypotheses for a track. More information about the method used to calculated the log-likelihoods is discussed in Section 3.4.1. This variable is required to be >-3 for the track to not be excessively *pion-like* and reduce misidentification probabilities.

4.2 Particle Identification Requirements

Knowing the identity of the final state particle tracks as a kaon or muon is a core requirement of this analysis. Particle Identification (PID) requirements are placed on the information provided by the hadron and muon identification systems discussed in Sections 3.3 and 3.4. These requirements considerably reduce the combinatorial background contribution from pions misidentified as kaons and combined with two random muons forming a B meson candidate. Through this technique, a major background component is reduced, as pions are the most abundant particle at proton colliders. It also suppresses so-called peaking backgrounds (*e.g.* $B^+ \to K^+\pi^+\pi^-$) which arise from decay modes with a final-state similar to $B^+ \to K^+\mu^+\mu^-$ but with one or more misidentified particles. These would form peaks in the $m_{K\mu\mu}$ distribution displaced from the B meson mass peak and are hence called peaking background.

The PID requirements imposed for this analysis are listed in Table 4.2. The variables ProbNNX with X = K,pi,mu are global PID variables which use information from RICH, muon and calorimeter systems. Deep neural networks are trained using this information to calculate the probability of a track being kaon, pion or muon. A discussion on these techniques can be found in Ref. [121] and [122]. The condition ProbNNK - ProbNNpi >-0.1 is imposed to select kaon tracks which are more *kaon-like* than *pion-like*. This used in conjecture with ProbNNK > 0.2 means that a good discrimination against pions can be achieved. The requirement of ProbNNmu > 0.25 is also imposed on the muon tracks.

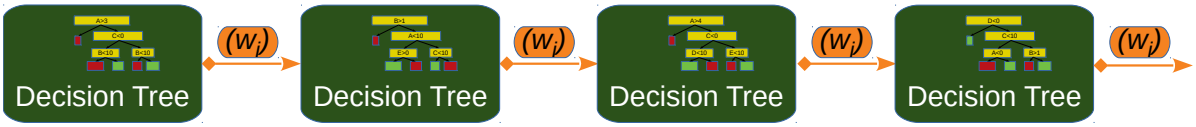

Figure 4.1: An illustration of the AdaBoost Algorithm using a series of decision trees. At each *boost* iteration the weights w_i of each candidate are updated.

4.3 Multivariate Classifier to Reduce Combinatorial Background

The data selection procedure discussed in the previous sections is effective in suppressing background candidates. However, it does not use all the available information such as the correlation between variables. As the penultimate step in the data selection procedure, an MVA method can be used to gain better separating power.

A boosted decision tree (BDT) [123] using an implementation of the AdaBoost algorithm [124, 125] is used through the scikit-learn python package [126]. Its working principle is detailed in Section 4.3.1. A BDT is trained to separate signal candidates and combinatorial background candidates. Labelled data sets are used with $B^+ \to K^+\mu^+\mu^-$ simulation acting as proxy for the signal candidates and data in the UMSB is used as proxy for combinatorial background candidates. The training procedure and performance of the BDT is described in Sections 4.3.2-4.3.5.

4.3.1 Overview of the AdaBoost Decision Tree algorithm

The Adaptive Boosting (AdaBoost) BDT algorithm is a supervised machine learning model, meaning it requires a training data set which has signal and background candidates correctly labelled. For this purpose it uses training variables which are commonly referred to as features. The algorithm works by using a specified number of decision trees in a series to create a strong classifier as shown in Figure 4.1.

Decision trees are weak classifiers and on their own do not offer a high predictive power. They also have a limitation in returning a globally optimal model as small changes in the data set could easily influence the final model generated. This can be overcome by using an ensemble method such as the AdaBoost algorithm.

4.3.1.1 Decison Tree Working Mechanism

A decision tree works by splitting the data set into two based on a threshold on one randomly chosen variable, an example condition would be `[B meson - FD` χ^2 `< 150]`. Each such decision is called a node and the initial decision is called the root node. The data set is continually divided and the tree grows until a node contains purely either signal or background. The growth of the tree can be interrupted by setting the total number of layers of such nodes. The number of layers is often referred to in literature as the maximum depth of the tree, the

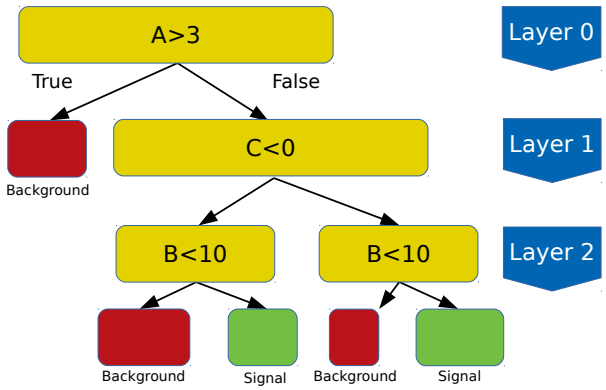

Figure 4.2: An illustration of a decision tree's attempt at classifying signal and background candidates. Here A,B,C are the exemplary MVA variables used to classify the data set.

end nodes are called leaf nodes. For this analysis the growth of the tree is stopped after one layer. Figure 4.2 shows an illustration of a decision tree.

At each node in order to pick the best split, the condition on the variable is chosen based on the Gini impurity. If a random sample is drawn from the data set and classified based on the number of signal and background, the Gini impurity gives the likelihood that it will be classified incorrectly. It can be written as,

$$\text{Gini}_{\text{node}} \equiv P_S(1-P_S) + P_B(1-P_B) \tag{4.1}$$

where P_S and P_B are the purity of the signal and background component and is defined by,

$$P_{S(B)} \equiv \frac{N_{S(B)}}{N_{tot}} \tag{4.2}$$

with $N_{S(B)}$ being the number of signal (background) candidates in the node and N_{tot} is the total number of candidates in the node.

The threshold θ to split the data set is chosen as the value of the variable which minimizes

$$L(\theta) = \frac{N_{\text{left}}}{N_{\text{current}}}\text{Gini}_{\text{left}} + \frac{N_{\text{right}}}{N_{\text{current}}}\text{Gini}_{\text{right}}, \tag{4.3}$$

with N_{current} being the total number of candidates in the current node and $N_{\text{left,right}}$ being the number of candidates in the left and right split nodes respectively for a threshold θ. This process is repeated until the leaf nodes are reached which for this analysis is a tree of depth one. This classification is not perfect and are only slightly better than random guessing. However they can be used within an ensemble method such as AdaBoost to create a better classifier.

4.3.1.2 AdaBoost Algorithm Working Mechanism

Multiple weak classifiers such as a decision tree can be used in series to form a powerful classifier. At each so-called boosting-iteration, all candidates in the data set are given a weight.

For the first iteration, all candidates are given the same weight of $1/N$ with N being the total number of candidates in the training set. This means that each candidate has the same level of attention on the first iteration. For every subsequent iteration the weights are updated such that all misclassified candidates get a higher weight. The weight of all correctly classified candidates are lowered such that the candidates that are difficult to classify get a higher attention by subsequent learners.

At each iteration the multiplicative weight factor for the next iteration is calculated according to,

$$W = \frac{1}{2} log \frac{1-\epsilon}{\epsilon}, \tag{4.4}$$

here ϵ is the error in identification and is given by,

$$\epsilon = \frac{\sum_{i \in N_{misid}} w_i}{\sum_i w_i}, \tag{4.5}$$

where the denominator is the weighted sum of all candidates. In the numerator N_{misid} is the set of misidentified candidates this includes signal candidates misidentified as background and vice versa. The numerator then is the weighted sum of all misidentified candidates[1].

With the multiplicative weight factor W, the weights w_i for each candidate are updated for the next iteration by,

$$w_i \rightarrow \begin{cases} w_i \cdot exp(W), & i \in N_{misid} \\ w_i, & i \notin N_{misid} \end{cases}, \tag{4.6}$$

implying that the weight of correctly identified candidates are unchanged while the weight of misidentified candidates are increased. The updated sum of all weights is normalized and used for the next iteration of the boosting.

In the overview of the working mechanism of a decision tree, the use of weighted candidates was ignored. To use a decision tree with weighted candidates, the $N_{S,B}$ quantities in Eq. 4.2 and Eq. 4.3 will be redefined as sum of the weights instead of number of candidates. The prediction of a candidate to be signal or background, referred to as the BDT score, is then calculated through an average of the sum of the binary predictions from all the decision trees.

4.3.2 Data preparation

The procedure used to set appropriate weights and configurations of the decision trees is referred to as training. In this analysis a maximum of 200 decision trees are used. Separate BDTs are trained for data collected in 2016, 2017 and 2018.

The BDT algorithm is trained on $B^+ \rightarrow K^+\mu^+\mu^-$ simulation used as proxy for signal candidates. The proxy for background is 10% of the reconstructed $m_{K\mu\mu}$ Upper Mass Sideband (UMSB) with potential $B^+ \rightarrow J/\psi K^+$ and $B^+ \rightarrow \psi(2S)K^+$ candidates removed. This means the

[1] In a classification problem with two classes such as in this analysis, ϵ is at most $\frac{1}{2}$.

MVA Variables		
B^+ IP χ^2 (own PV)	B^+ p_{T}	$m_{\mu\mu}$ IP (min.)
B^+ IP (own PV)	B^+ P	μ^- IP χ^2 (min.)
B^+ FD χ^2	K^+ IP χ^2 (min.)	μ^+ IP χ^2 (min.)
B^+ vertex χ^2	K^+ P	

Table 4.3: The kinematic and geometric variables use to train the MVA classifier to differentiate between signal and combinatorial background candidates.

background proxy is 10% of data after the requirements: $m_{K\mu\mu} \in [5600, 5900]\,\mathrm{MeV}/c^2$; $m_{\mu\mu} \notin [3050, 3150]\,\mathrm{MeV}/c^2$; $m_{\mu\mu} \notin [3650, 3725]\,\mathrm{MeV}/c^2$. The signal and background proxy candidates are combined and randomized. Only 80% of the mixed and randomized set is used for training the BDT, the rest is set aside as the testing set to validate the performance of the BDT and is discussed in Section 4.3.5. In order to account for data/simulation differences, for simulation the initial weight is set to the weights calculated in Section 5.1. The weight of data is set to 1. The procedure then continues as discussed in the beginning of Section. 4.3.1.2 with the weights normalized to the total number of candidates.

4.3.3 Choice of variables

The choice of training variables used to train the BDT is crucial. The more information the BDT has, the better its classification power. The variables used are summarized in Table 4.3 and their signal and background proxy distributions are shown in Figure 4.3.

From the B meson candidate, IP, IP χ^2 w.r.t the own PV, the end vertex χ^2 and the FD χ^2 are used. In addition the B meson transverse momentum is also used. These variables offer good discrimination power between signal proxy and background candidates as can be seen in Figure 4.3. While the signal proxy and background distributions of B, K momentum and K, $m_{\mu\mu}$, μ^+, μ^- IP χ^2 looks similar, in combination with other kinematic variables they also offer good discriminating power.

4.3.4 k-folding to reduce bias and increase statistics

Using the same BDT to evaluate the training data could potentially bias our sample. One way to avoid this would be to remove the training data set of the BDT from the rest of the analysis. This however has the issue of reducing the available statistics and needs to be accounted for in the later stages of the analysis. k-folding is a technique where the training data is split into $k = 8$ equal chunks. A BDT is trained on $k - 1$ chunks and evaluated on the remaining data. This procedure creates k independent BDTs each evaluated on the part of the data not used during training. The BDT score for the remaining data not used in the preparation of the BDT is set to be the average score from the 8 BDTs.

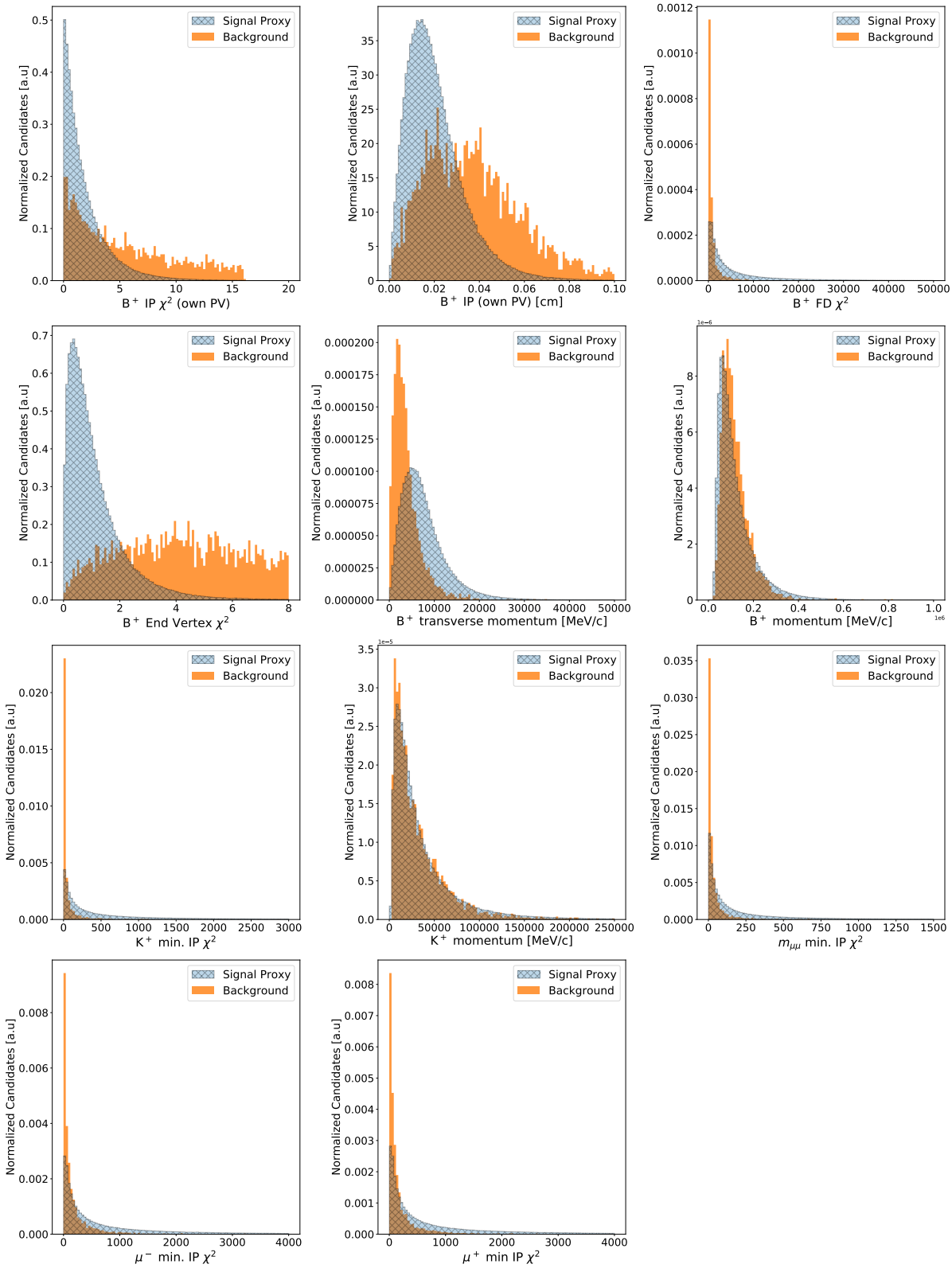

Figure 4.3: The signal and background distributions of the kinematics and geometric variables used to train the BDT. Shown here for illustration are the candidates from 2018.

4.3.5 BDT results and performance

To evaluate the performance of the BDT training, a Receiver Operating Characteristic (ROC) curve is plotted and is shown in Figure 4.4. A ROC curve plots the true positive rate against the false positive rate to measure the performance of a binary classifier. If the curve is more severe i.e. if the area under the curve is close to 1 it indicates at a well performing classifier. This is since a low false positive rate also has a very high true positive rate.

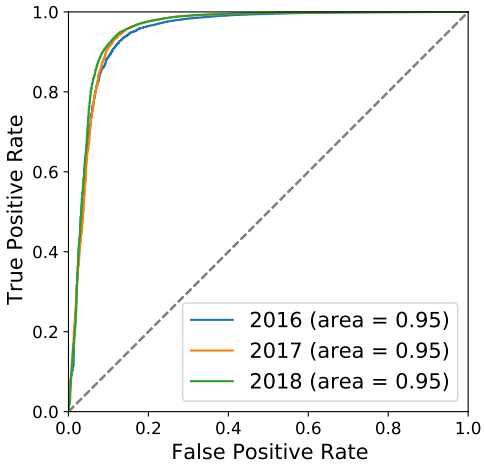

Figure 4.4: The Receiver Operating Characteristic (ROC) curve for 2016, 2017 and 2018. The area under the curve being close to 1 is an indication of a good BDT.

The BDT is also checked for overtraining, this check ensures that the BDT has not been heavily optimized for the training data set. If the BDT has been overtrained it will be able to predict the training data with high accuracy but perform poorly with an unseen data set. To check for overtraining, the trained BDT is used to classify the test data set. Figure 4.5 shows the BDT scores for the training and testing. A good separation between signal and background can be seen for all years. Since the testing and training distributions overlay each other well it can be concluded that the BDT is not overtrained and is generalizable to an unknown data set.

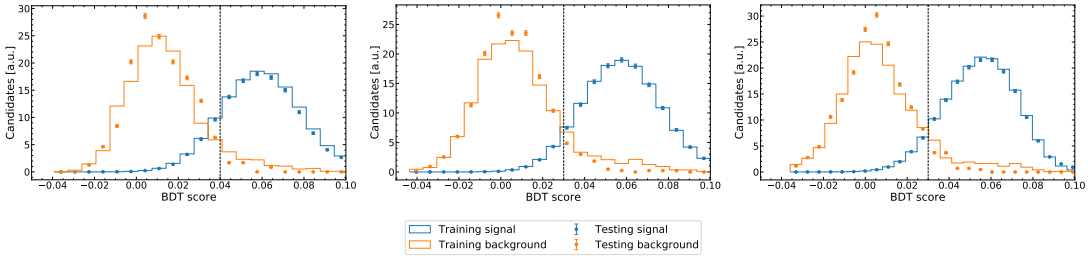

Figure 4.5: The distribution of BDT scores from the training/testing set for 2016 (left), 2017 (center) and 2018 (right) samples. The signal (blue) and background (orange) candidates are highlighted.

The BDT classifier gives a BDT score for each candidate quantifying if it is *signal-like* or

background-like. To select the optimal threshold value which separates signal and background, an optimization procedure is used. For each threshold value of the BDT, the number of signal and background (n_{sig}, n_{bkg}) candidates which pass this selection are estimated. The number of signal candidates, n_{sig} is estimated from how many $B^+ \to K^+\mu^+\mu^-$ candidates are expected based on the branching fraction $\mathcal{B}$ ($B^+ \to K^+\mu^+\mu^-$) and the integrated luminosity $\mathcal{L}_{int}$ of the year. This is given by,

$$n_{sig} = 2\mathcal{L}_{int} \cdot \sigma_{b\bar{b}} \cdot f_u \cdot \mathcal{B}(B^+ \to K^+\mu^+\mu^-) \cdot \frac{n^{MC}_{cand}(threshold)}{n^{MC}_{cand}(total)} \cdot G_{cut}, \tag{4.7}$$

where $\sigma_{b\bar{b}}$ is the cross-section of $b\bar{b}$ pair production, f_u is the hadronisation fraction for u quarks to create B^+. The number of $B^+ \to K^+\mu^+\mu^-$ simulated candidates which pass the threshold and total number of candidates are $n^{MC}_{cand}(threshold)$ and $n^{MC}_{cand}(total)$. Simulating particles through the detector is a very computationally expensive task and event generation is relatively not. In order improve the speed of the simulation, a selection is made before detector simulation on the generated events to remove any events which could be outside the detector's acceptance. This selection is called generator level cut and its efficiency is given by $G_{cut} = 0.176$.

The $m_{K\mu\mu}$ sideband is fitted with an exponential and extrapolated to the signal region to estimate the background yield n_{bkg} which is given by,

$$n_{bkg} = \epsilon^B_{BDT} \times n_{umsb} \cdot \frac{A_{sig}}{A_{sb}}, \tag{4.8}$$

here ϵ^B_{BDT} is the background BDT misidentification rate for a given threshold. n_{umsb} is the yield in $m_{K\mu\mu}$ UMSB and $A_{sig,sb}$ are the area under the exponential in the signal and the UMSB regions.

A very common approach to choose the optimum threshold is to use the BDT score which maximizes the figure of merit,

$$\frac{n_{sig}}{\sqrt{n_{bkg} + n_{sig}}} = \frac{S}{\sqrt{S+B}} \tag{4.9}$$

and is plotted in Figure 4.6 for 2016, 2017 and 2018.

At its maximum, the signal efficiency is high while the background misidentification rate is ~ 20% for all three years. Since the estimated number of signal candidates is high, a small fraction of the signal is sacrificed to reduce background further. Therefore, the threshold value chosen ensures the signal efficiency is always 90%. The background misidentification rates are 11%, 9% and 9% for cut values 0.04, 0.03 and 0.03 respectively for 2016, 2017 and 2018. The previous iteration of this analysis using Run1 data achieved a signal efficiency of 89% with a background misidentification rate of 6%.

The respective thresholds for each year are applied to get the BDT cut data set, removing most of the combinatorial background. The same procedure is applied to the rest of the simulated samples to maintain a consistent selection procedure. The BDT cut data set from the

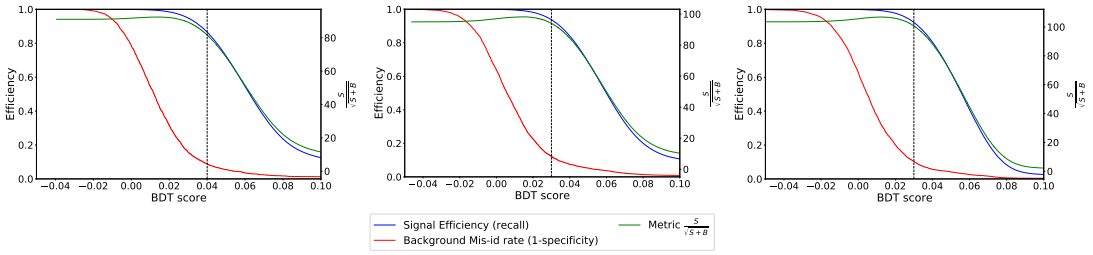

Figure 4.6: The signal efficiency (blue) and background misidentification (red) rates as a function of BDT scores for 2016 (left), 2017 (middle) and 2018 (right). The metric $\frac{S}{\sqrt{S+B}}$ is also plotted as a green line for reference. The dashed black line indicates the chosen cut value of the BDT which ensures a signal efficiency of 90%.

Run1 iteration is reused for this analysis. Some misidentified combinatorial background will remain in data and the method used to account for it is discussed in Section 6.2.3.

4.4 Removing Specific Backgrounds

With one or more particles misidentified, it is possible that other B meson decay modes with similar topology and kinematics to $B^+ \to K^+\mu^+\mu^-$ could pass the kinematic, PID and MVA based selections. This could be a particular large contribution if these specific decays have a significantly larger branching fraction that $B^+ \to K^+\mu^+\mu^-$ decays. This section identifies the specific backgrounds and describes the method used to remove them.

4.4.1 Pions misidentified as muons

One specific background contribution is from the decay $B^+ \to \pi^+\overline{D^0}(K^+\pi^-)$ which has a large branching fraction of about 1.8×10^{-4}. If the two final state pions decay in flight to muons or are misidentified, these decays can mimic signal $B^+ \to K^+\mu^+\mu^-$ candidates. This contribution can be suppressed by having a more stringent PID requirement. In order to achieve this the mass of the $K^+\mu^-$ system is recalculated under the hypothesis that the mass of the μ^- track is the mass of a π^-. The resulting $m_{K^+(\mu^-\to\pi^-)}$ mass distribution is shown in Figure 4.7 for 2016, 2017 and 2018. The peak seen at ~ 1864 MeV/c^2 on the BDT cut data set corresponds to the mass of the D^0 meson. A simple way to reduce this contribution would be to remove all candidates within the window $1850 < m_{K^+(\mu^-\to\pi^-)} < 1880\,\text{MeV}/c^2$. However an alternative approach is chosen to preserve as many of the genuine $B^+ \to K^+\mu^+\mu^-$ candidates as possible.

Within the window $1850 < m_{K^+(\mu^-\to\pi^-)} < 1880\,\text{MeV}/c^2$, the PID variable `ProbNNmu` of the μ^- track is required to be larger than 0.8. Outside this mass range the criterion `ProbNNmu` > 0.25 is maintained. Figure 4.7 shows the effect of this requirement and it can be seen that the structure caused by the D^0 candidates has reduced. The efficiency of this requirement on signal is greater than 99%.

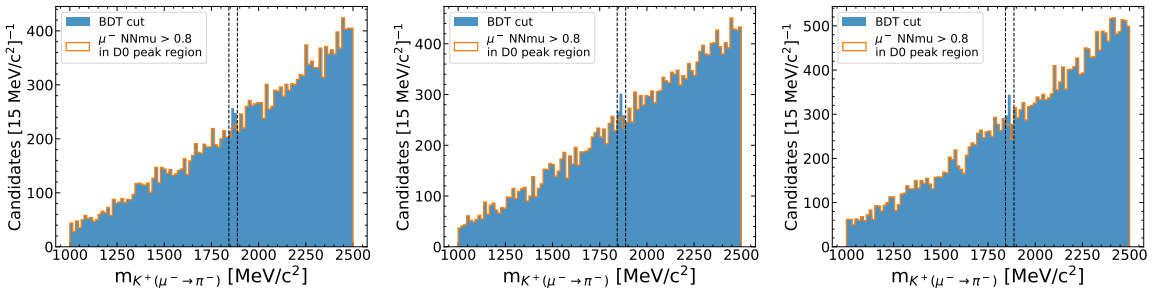

Figure 4.7: The $m_{K^+\mu^-}$ mass distribution under a mass hypothesis swap for μ^- to a π^-. Shown here are data distributions of 2016 (left), 2017 (middle) and 2018 (right). The peak seen within the dashed black lies in the BDT cut distribution correspond to D^0 candidates with the π^- misidentified as a μ^-. A more stringent PID requirement is placed on the μ^- track to remove these candidates and is shown by the orange distribution.

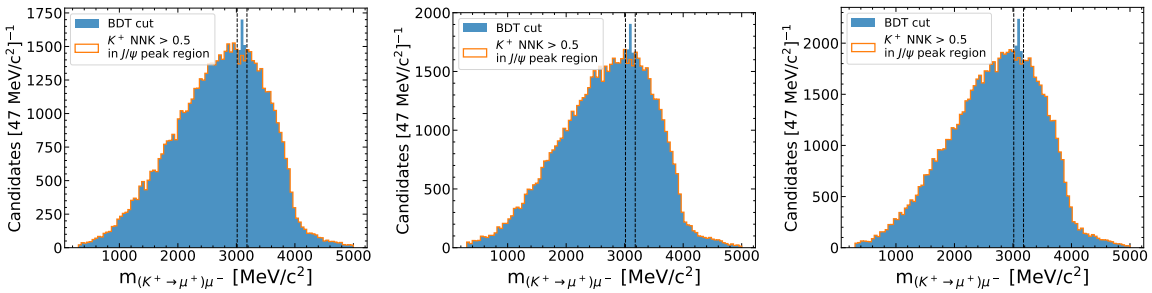

Figure 4.8: The $m_{K^+\mu^-}$ mass distribution under a mass hypothesis swap for K^+ to a μ^+. Shown here are data distributions of 2016 (left), 2017 (middle) and 2018 (right). The peak seen within the dashed black lies in the BDT cut distribution correspond to J/ψ candidates with the μ^+ misidentified as a K^+. A more stringent PID requirement is placed on the K^+ track to remove these candidates and is shown by the orange distribution.

4.4.2 Kaons and muons mass hypothesis swap

The decay $B^+ \to J/\psi K^+$ can form a misidentified background if the kaon swaps mass hypothesis with the same sign muon The large branching of $B^+ \to J/\psi K^+$ makes this more likely. Such swap candidates can be seen as a peak near the mass of the J/ψ in the $\mu^- K^+$ mass distribution under the $\mu^-\mu^+$ mass hypothesis shown. The mass hypothesis swapped distributions for the different Run2 years is shown in Figure 4.8. Similar to the pion misidentified scenario discussed above, a stronger PID requirement of `ProbNNK` > 0.5 is placed on the K^+ tracks. This requirement is placed around the J/ψ mass window $3037 < m_{(K^+\to\mu^+)\mu^-} < 3157\,\text{MeV}/c^2$ on the data set after the multivariate BDT selection. The peak can be seen to reduce with an efficiency on signal greater than 99%. Since a visible structure is not seen around the $\psi(2S)$ mass a requirement is not placed on it and any remaining background is assumed to be negligible.

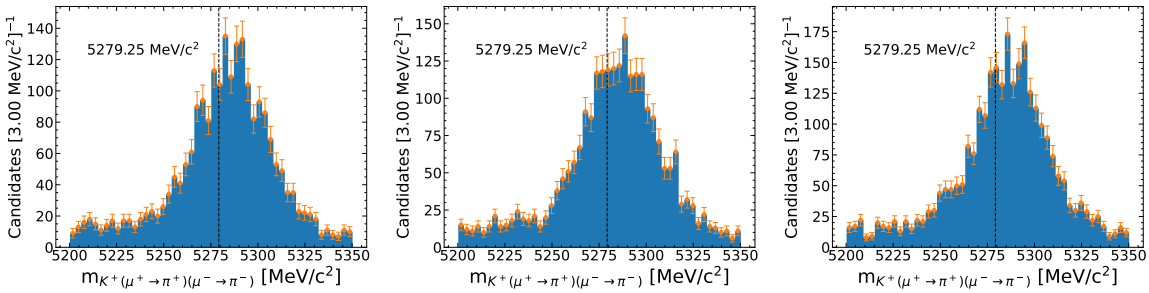

Figure 4.9: The $m_{K\mu\mu}$ mass distribution under a double mass hypothesis swap of a pion for both the muon tracks. Shown here are data distributions of 2016 (left), 2017 (middle) and 2018 (right) with mass. The known B meson mass of 5279.25 MeV/c^2 is highlighted by the dashed vertical line. No visible structure is seen under the alternative and therefore the specific background from $B^+ \to K^+\pi^+\pi^-$ candidates are assumed to be negligible.

4.4.3 Double misidentification of pions as muons

Another potential source of misidentification is from resonant and non-resonant $B^+ \to K^+\pi^+\pi^-$ with the dominant intermediate resonant states being $\omega(782)$ and $f(980)$. Here the two pions could be misidentified as muons, the total branching fraction of this decay mode $(5.10 \pm 0.29) \times 10^{-5}$ [55]. To check for this specific background the mass of $m_{K\mu\mu}$ system is recalculated with the mass hypothesis that both muon tracks are pions. The resulting $m_{K^+(\mu^+\to\pi^+)(\mu^-\to\pi^-)}$ mass distributions for 2016, 2017 and 2018 data are shown in Figure 4.9. Here an additional mass requirement of $m_{(\mu^+\to\pi^+)(\mu^-\to\pi^-)} \in [300, 2700]$ MeV/c^2 is placed. The existence of double misidentified candidates would show as a peak in the $m_{K^+(\mu^+\to\pi^+)(\mu^-\to\pi^-)}$ at the known B meson mass of 5279.25 MeV/c^2. No such candidates can be identified in Figure 4.9. The visible broad peaks are from $B^+ \to K^+\mu^+\mu^-$ candidates with the pion mass hypothesis on the muon tracks and are therefore shifted from the known B meson mass. Therefore, the double misidentified background is assumed to be negligible and no further selection requirement is imposed.

CHAPTER

5 Accounting for the Detector's Efficiency and Resolution effects

The detection and the selection requirements imposed on $B^+ \to K^+\mu^+\mu^-$ candidates sculpts the phasespace of the decay. To account for this, decays of B mesons are generated and simulated through the detector using algorithms mentioned in Section 3.6. The simulated tracks are then subject to an identical selection procedure as data (described in Chapter 4).

The simulation is imperfect and needs to be corrected for differences between data and simulation. This procedure is outlined in Section 5.1. Section 5.2 describes the functional form of the efficiency as a function of the mass of the dimuon system and how it is incorporated into the model of the differential decay rate. The mass resolution of the detector needs to be accounted for, especially to model accurately the large resonances with narrow natural widths: J/ψ and $\psi(2S)$. The method used to model the resolution effects is discussed in Section 5.3.

5.1 Correcting for differences between data and simulation

The simulation needs to be corrected for the kinematics of the B hadron, the trigger efficiency and the PID performance. The differences between data and simulation can be estimated by comparing distributions of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ candidates from data and simulation. All the simulated samples described in Section 3.6 for 2016, 2017 and 2018 are then corrected to account for these differences.

The corrections between data and simulation are done in three stages; corrections to the trigger efficiency, PID distributions and kinematic variables. The efficiency of the muon trigger within the detector acceptance in simulation are defined as the ratio of the number of candidate events triggered by a signal muon to the total number of events. In data however, there is no information about not triggered events, therefore the efficiency is defined as the number of candidates triggered by both signal μ^+ and μ^- over the number of events triggered on signal

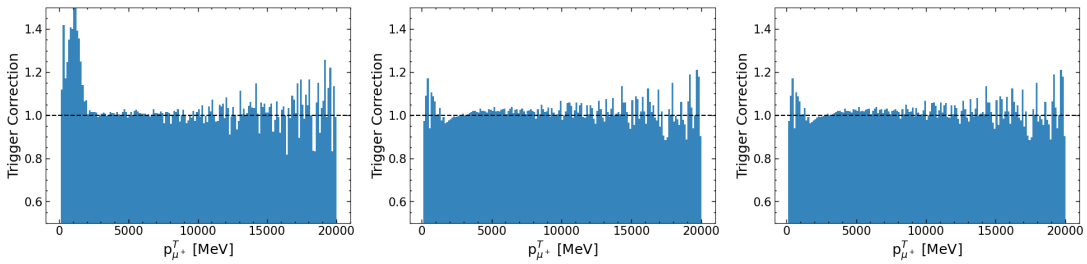

Figure 5.1: The ratio of the trigger efficiency in data over the efficiency in simulation for 2016 (left), 2017 (middle) and 2018 (right) as a function of μ^+ transverse momentum (p_T).

μ^+. The ratio of the two efficiencies as a function of the μ^+ p_T is calculated and shown in Figure 5.1 for 2016, 2017 and 2018. These distributions are used to correct for the differences between data and simulation. The corrections for all three years are estimated to be small. The efficiencies of the HLT are well modelled in the simulation and no correction is applied.

The second type of correction pertaining to the PID is estimated using PIDCalib2 [127]. This package uses a set of calibration samples from data ($B^+ \to K^+ J/\psi(\mu^+\mu^-)$ for muons and $D^{*+} \to D^0(K^-\pi^+)\pi^+$ for kaons) which are selected without the use of PID information. The efficiency of arbitrary PID requirements can be estimated from these samples. For muon and kaon tracks their respective calibration samples are used to create 2D distributions with sufficient granularity in momentum and transverse momentum. The efficiency of the PID requirements discussed in Section 4.2 as a function of the momentum and transverse momentum of the particle is estimated. This efficiency is used to correct the distribution of the simulation samples.

The final set of corrections are to correct for the differences in the distributions of multiplicity of the event, `nTracks`, the χ^2 of the decay vertex of the B^+ candidate (B^+ end vertex χ^2), and the momentum (p) and transverse momentum (p_T) of the B^+ candidate. Taking one variable at a time, the ratio between binned data and simulated samples is taken as the working point of the correction weight for the next variable. The correction weight is iteratively updated using the distributions of the four variables in series. These kinematic quantities are however correlated to each other and therefore this correction procedure is repeated three times, applying the corrections of the previous iteration to derive the corrections of the next, until convergence is reached for all variables. The resulting histograms show good agreement between data and simulation for all the variables.

The corrections for trigger and kinematic variables are then used to calculate the weights for the other simulated samples. The trigger weights have the smallest corrections followed by PID and kinematic weights. In order to limit the tail of the kinematic distribution from being unreasonably large, these weights have been capped at 10. The resulting distribution of correction weights for all Run2 $B^+ \to K^+\mu^+\mu^-$ simulated samples is shown in Figure 5.2. This weight is applied when training the BDT discussed in Section 4.3 and while calculating the efficiency in Section 5.2.

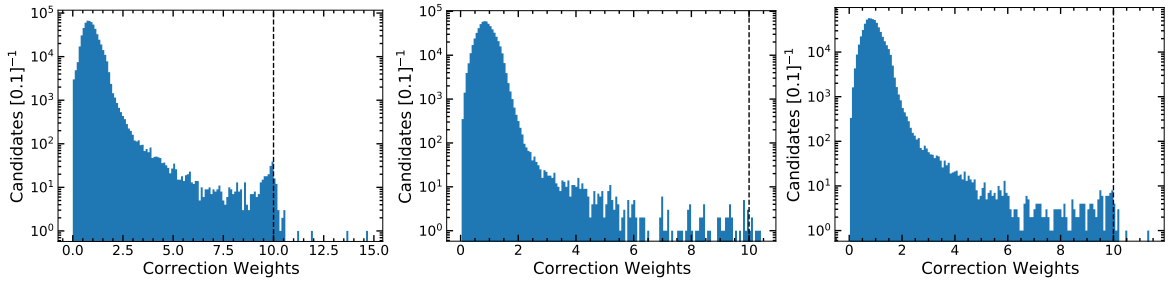

Figure 5.2: The distribution of the combined correction weights from left to right for 2016, 2017 and 2018 $B^+ \to K^+\mu^+\mu^-$ simulation. The dashed vertical line indicates the cap on the kinematic weights which in combination with the trigger and PID causes the peak around it.

The data and simulation distributions after applying BDT cuts with and without these corrections for 2016, 2017 and 2018 $B^+ \to K^+\mu^+\mu^-$ simulated candidates are shown in Figure 5.3, 5.4 and 5.5 respectively. The variables shown are the ones used to train the BDT. As can be inferred from these plots, the data/simulation corrections are small and have been corrected for.

5.2 Calculating the $m_{\mu\mu}$ dependent efficiency correction

The efficiency of the detector and selection requirements to collect the signal candidates needs to be accounted for. This correction is used to create the efficiency-corrected decay model. The dominant contributions to the trigger and selection efficiency are the p_{T} requirements on the muons and the IP requirements on the kaons and dimuon system.

The efficiency is calculated using two sets of simulated samples. The first is $B^+ \to K^+\mu^+\mu^-$ simulated candidates passed through a simulation of the detector (Section 3.6). Identical selection criteria similar to data are applied after correcting for differences between data and simulation. This simulated set of candidates is referred to as the reconstructed sample and contains 1,764,114 weighted candidates. The second set of simulated samples are the simulation of $B^+ \to K^+\mu^+\mu^-$ decays from EvtGen [114] using the same model used to generate fully reconstructed samples. This set is referred to as the generator level samples and 5,000,000 such candidates are simulated. The generator level samples simulate decays without any interaction with the detector and candidate selection requirement. By taking a ratio of these two sample distributions, the efficiency of the detector and candidate selection can be evaluated.

The normalized 2D $m_{\mu\mu}$ and $cos\theta_\ell$ distributions from the generator level and reconstructed sample are shown in Figure 5.6. As a reminder, θ_ℓ is the angle between the direction of the μ^+ (μ^-) and the direction of the B^+ (B^-) meson in the rest frame of the dimuon system. From these distributions it is clear that the differences between generator and reconstruction level candidates depend on $m_{\mu\mu}$ and $cos\theta_\ell$. The detector is more efficient at high $m_{\mu\mu}$ and $cos\theta_\ell$ close to the 0. This is primarily due to the large selection efficiency for high p_{T} muons which

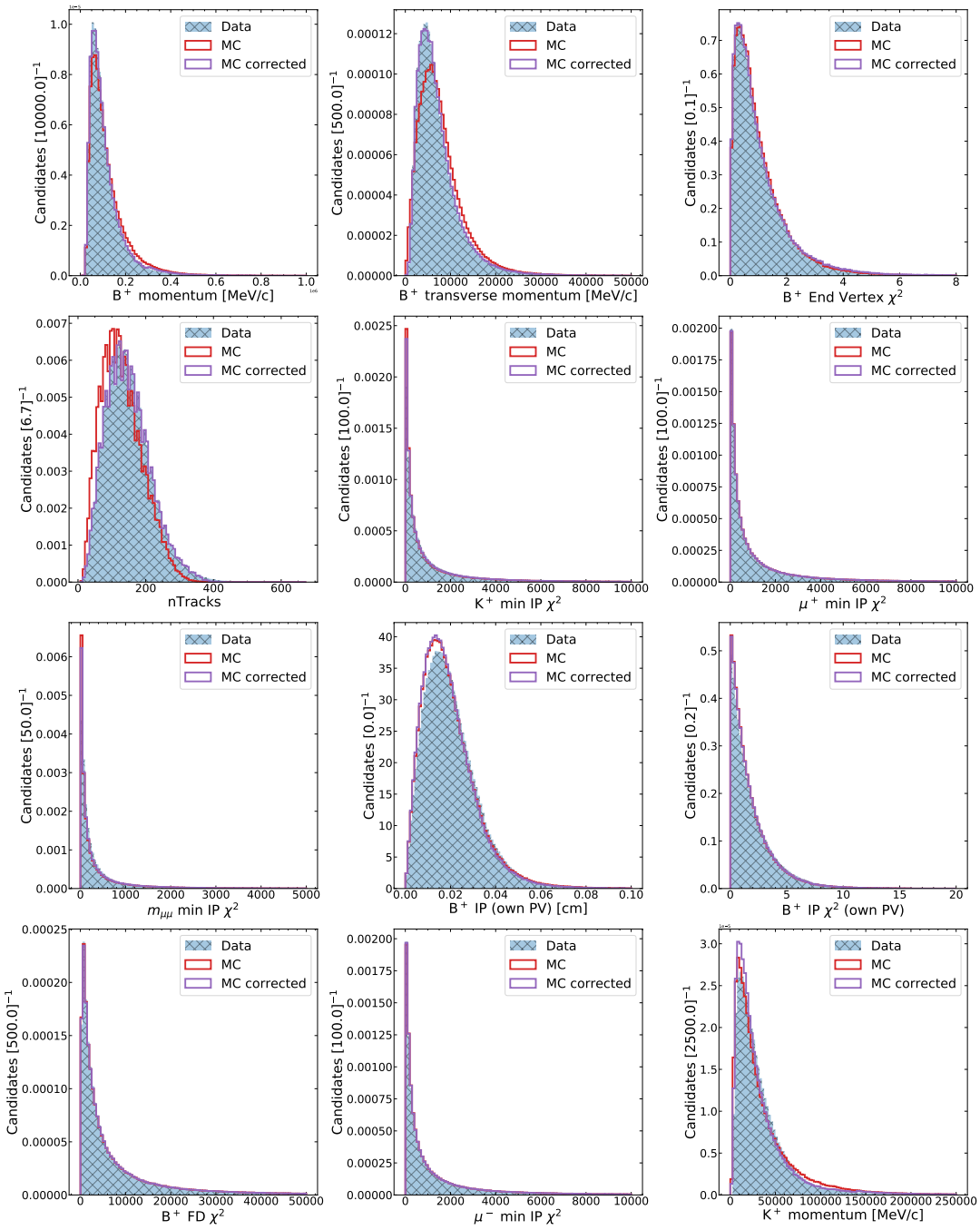

Figure 5.3: The distributions of various kinematic variables which are used to train the BDT discussed in Section 4.3 of 2016 data and $B^+ \to K^+\mu^+\mu^-$ simulated samples. The simulation corrected distributions include the trigger, PID and kinematic corrections discussed in the text.

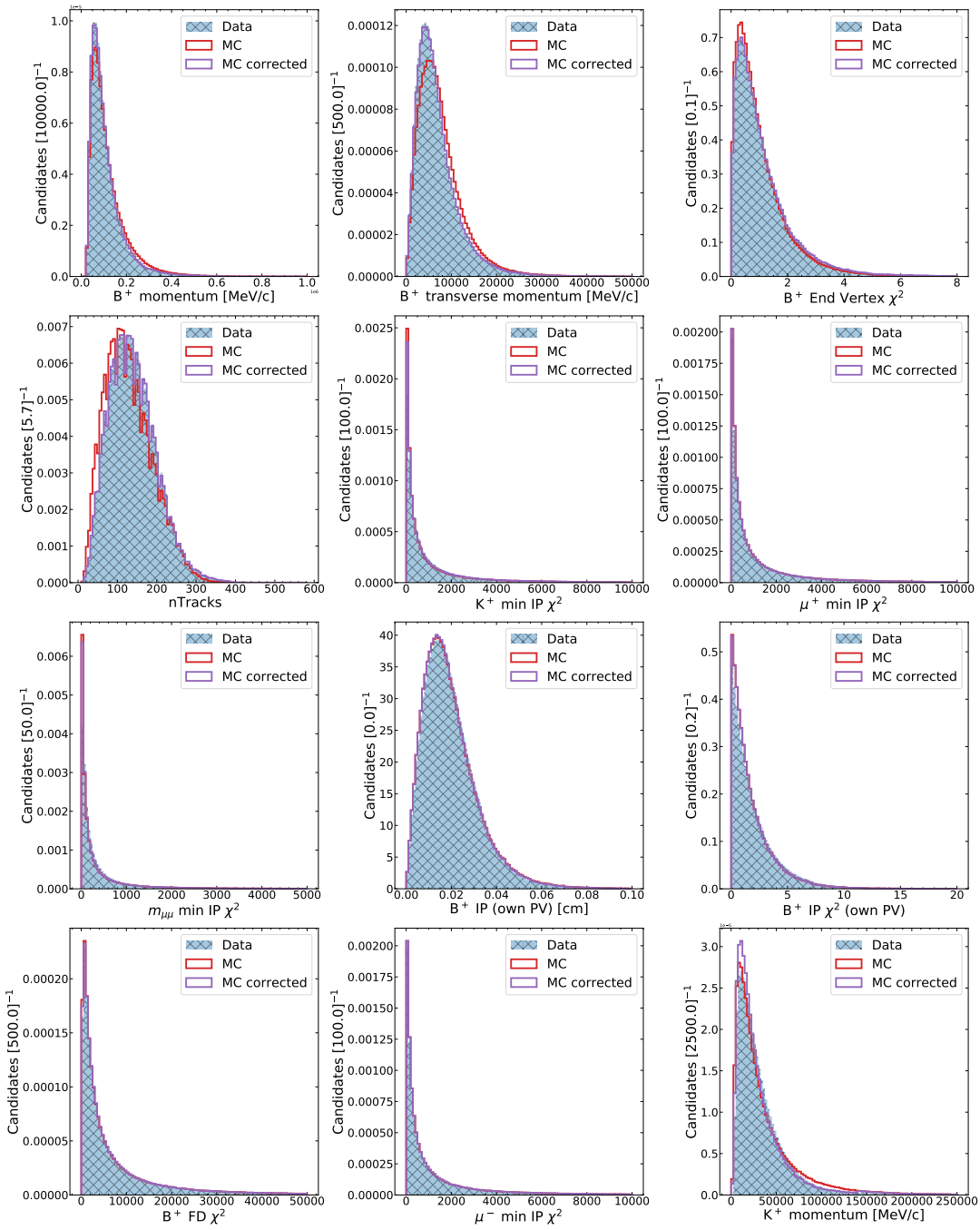

Figure 5.4: The distributions of various kinematic variables which are used to train the BDT discussed in Section 4.3 of 2017 data and $B^+ \to K^+\mu^+\mu^-$ simulated samples. The simulation corrected distributions include the trigger, PID and kinematic corrections discussed in the text.

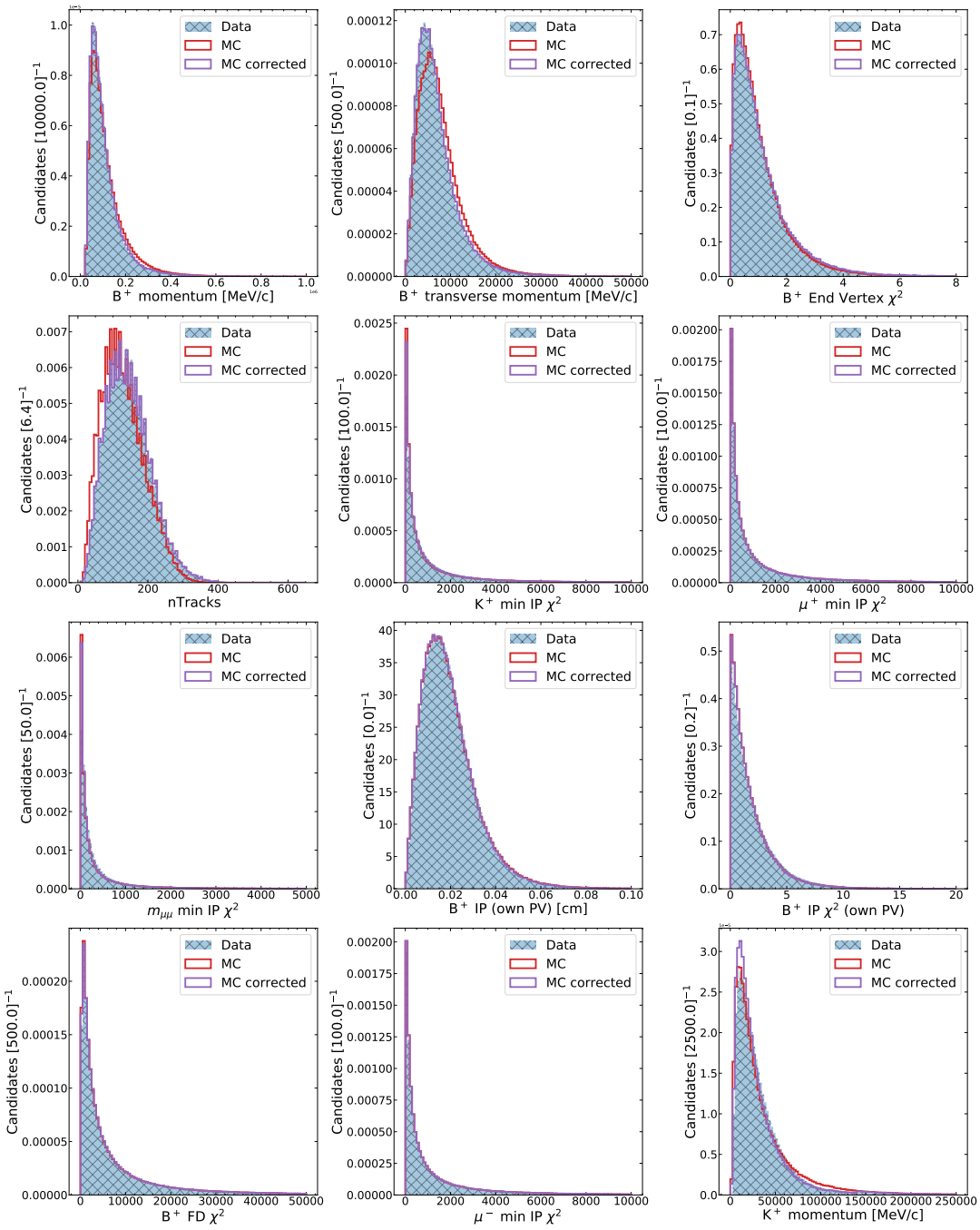

Figure 5.5: The distributions of various kinematic variables which are used to train the BDT discussed in Section 4.3 of 2018 data and $B^+ \to K^+\mu^+\mu^-$ simulated samples. The simulation corrected distributions include the trigger, PID and kinematic corrections discussed in the text.

drop at the highest values of $m_{\mu\mu}$. This drop comes from the strict IP requirement on an almost stationary kaon. The efficiency reduces for higher values of $|cos\theta_\ell|$ and is visible in these plots especially at low dimuon masses. At large $|cos\theta_\ell|$ ($\theta_\ell = 0, \pi$), in the lab frame this translates to having one fast and one slow moving muon which is difficult to detect as most of the momentum of the B meson is carried by the kaon and the fast muon. Due to the strong dependence of efficiency on both q^2 and $cos\theta_\ell$, both variables are used to evaluate the detector efficiency effects.

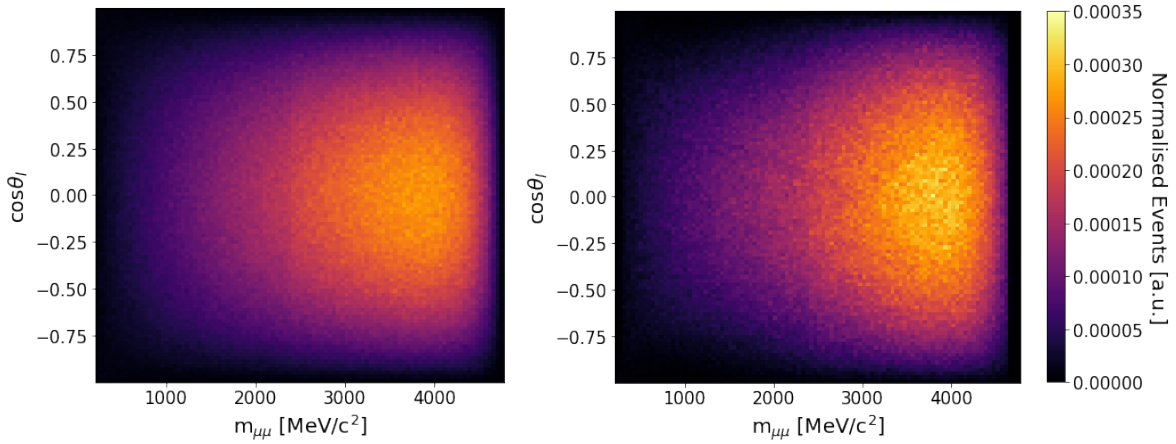

Figure 5.6: The normalized 2D distribution in $cos\theta_\ell$ and $m_{\mu\mu}$ from generator level (left) and data/simulation corrected reconstructed (right) simulation.

5.2.1 The angular dependence of efficiency

The differential decay rate used in this measurement has the angular dependence integrated out. The full double differential decay rate is,

$$\begin{aligned}\frac{d\Gamma_l}{dq^2 dcos\theta_\ell} \propto \sqrt{\lambda(q^2)}\beta\Big\{&\frac{\lambda(q^2)}{4}\beta^2(1-cos^2\theta_\ell)|C_{10}|^2 f_+^2(q^2)\\ &+\frac{m_l^2}{q^2}(M_B^2-M_K^2)^2|C_{10}|^2 f_0^2(q^2)\\ &+\frac{\lambda(q^2)}{4}(1-\beta^2 cos^2\theta_\ell)|\tilde{C}_9|^2 f_+^2(q^2)\Big\},\end{aligned} \tag{5.1}$$

where

$$\tilde{C}_9 = C_9^{\text{eff}} + 2C_7^{\text{eff}}\frac{m_b+m_s}{M_B+M_K}\frac{f_T(q^2)}{f_+(q^2)}. \tag{5.2}$$

The efficiency as a function of $cos\theta_\ell$ strongly depends on $m_{\mu\mu}$. Two example plots at two different bins of $m_{\mu\mu}$ are shown in Figure 5.7. This means that the three terms in Eq. 5.1 with a different dependence on $cos\theta_\ell$ do not have the same sensitivity to angular distribution effects. For this reason the efficiency correction is calculated separately for the three terms.

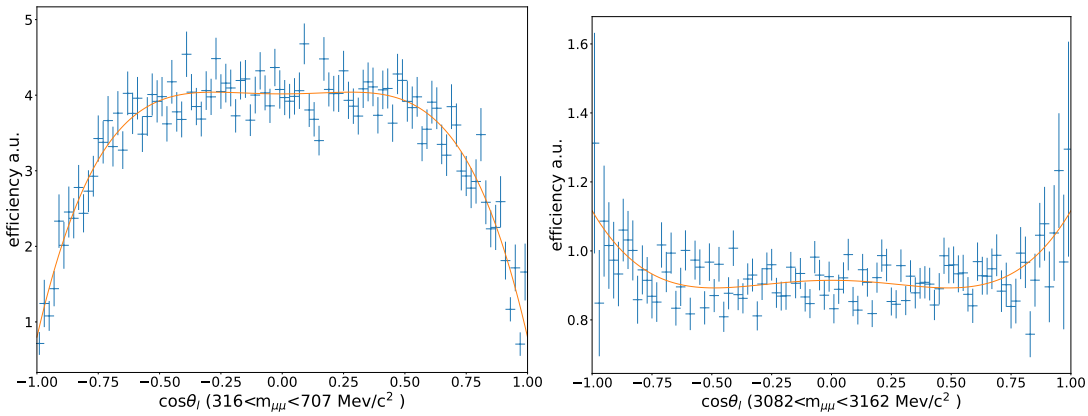

Figure 5.7: The efficiency effects in the $cos\theta_\ell$ distribution for $m_{\mu\mu}$ regions of $m_{\mu\mu}$ between [316, 707] MeV/c^2 (left) and [3082, 3162] MeV/c^2 (right) from $B^+ \to K^+\mu^+\mu^-$ simulated samples.

The efficiency correction proportional to the three terms are,

$$|C_{10}|^2 f_+^2(q^2): \qquad \mathcal{E}_{+C_{10}}(q^2) = \frac{\int_{-1}^{+1} \epsilon(q^2, cos\theta_\ell)(1 - cos^2\theta_\ell) dcos\theta_\ell}{\int_{-1}^{+1} ((1 - cos^2\theta_\ell) dcos\theta_\ell} \tag{5.3}$$

$$|C_{10}|^2 f_0^2(q^2): \qquad \mathcal{E}_{0C_{10}}(q^2) = \frac{\int_{-1}^{+1} \epsilon(q^2, cos\theta_\ell) dcos\theta_\ell}{\int_{-1}^{+1} dcos\theta_\ell} \tag{5.4}$$

$$|\tilde{C}_9|^2 f_+^2(q^2): \qquad \mathcal{E}_{+C_9}(q^2) = \frac{\int_{-1}^{+1} \epsilon(q^2, cos\theta_\ell)(1 - \beta^2 cos^2\theta_\ell) dcos\theta_\ell}{\int_{-1}^{+1} (1 - \beta^2 cos^2\theta_\ell) dcos\theta_\ell} \tag{5.5}$$

Here the numerator is the efficiency correction accounting for the appropriate $cos\theta_\ell$ dependence and the denominator is a normalisation constant. This correction is then multiplied by the corresponding term of the decay rate to obtain the efficiency-corrected differential decay rate model.

5.2.2 Evaluating the functional form of efficiency correction

The detection and reconstruction efficiency as a function of $cos\theta_\ell$ is evaluated by using $B^+ \to K^+\mu^+\mu^-$ simulated samples. This evaluation is performed in $m_{\mu\mu}$ using the definition $q^2 \equiv m^2_{\mu\mu}$ and later applied to the decay rate transformed to $m_{\mu\mu}$. The efficiency correction is calculated for the combination of all run periods by weighting the individual simulation samples according to the yields of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ candidates from data. This is done in order to match yields of the simulation samples from the different years to the expected yields in data to accurately represent the collected dataset. Since the efficiency correction is determined as a function of $m_{\mu\mu}$, obtaining the mixture weights at a singe point in $m_{\mu\mu}$ is by construction propagated to the rest of the decay phasespace.

The efficiency for the combined data periods is given by,

$$\epsilon(m_{\mu\mu}, cos\theta_\ell) = \sum_i \frac{\mathcal{N}_i^{J/\psi}}{\mathcal{N}_{2016}^{J/\psi}} \frac{N_i(m_{\mu\mu}, cos\theta_\ell)}{n_i(m_{\mu\mu}, cos\theta_\ell)}, \tag{5.6}$$

where the sum iterator i runs over the different data taking periods: Run1, 2016, 2017 and 2018. The yield of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ candidates from data is denoted by $\mathcal{N}_i^{J/\psi}$ with 2016 taken as the reference year. The number of candidates in the bin ($m_{\mu\mu}$, $cos\theta_\ell$) in fully reconstructed $B^+ \to K^+\mu^+\mu^-$ simulated samples is denoted by N_i and in $B^+ \to K^+\mu^+\mu^-$ generator level samples is denoted by n_i. The $m_{\mu\mu}$ and $cos\theta_\ell$ dependent efficiency is then integrated in $cos\theta_\ell$ as in Eqs. 5.5, 5.3, 5.4 to get the $m_{\mu\mu}$ dependent efficiency distributions. The three curves are shown in Figure 5.8. In order to emphasize the importance of implementing a $cos\theta_\ell$ dependent efficiency evaluation, Figure 5.8 also shows the efficiency evaluation $\epsilon(m_{\mu\mu})$ from generator level samples with $cos\theta_\ell$ integrated distributions. To get the functional forms of efficiencies, the distributions are fitted with a 4th order polynomial to get three efficiency curves $\mathcal{E}_{+C_{10}}$ $(m_{\mu\mu})$, $\mathcal{E}_{0C_{10}}$ $(m_{\mu\mu})$, $\mathcal{E}_{+C_9}$ $(m_{\mu\mu})$. These curves are then multiplied with their respective terms of the double differential decay rate and integrated in $cos\theta_\ell$ to give the efficiency corrected differential decay rate.

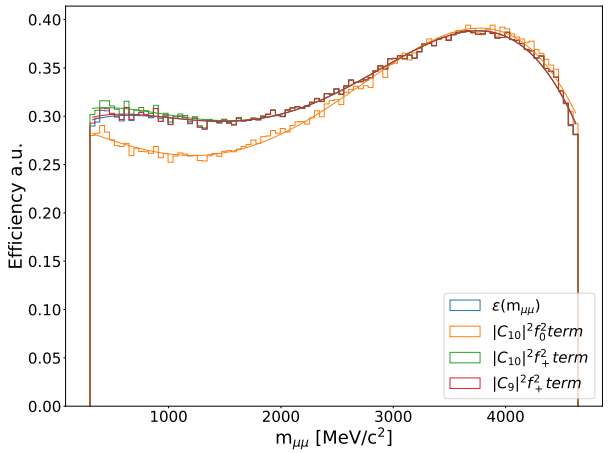

Figure 5.8: The efficiency dependence on the $m_{\mu\mu}$ spectrum calculated from reconstructed simulated events using weights from generator-level simulated events for the different angular terms as defined in the text.

5.3 Mass Resolution of the Detector

The natural widths of J/ψ is 0.093 MeV/c^2 and of $\psi(2S)$ is 0.29 MeV/c^2 [55], which are about 50 and 20 times smaller than the detector's dimuon mass resolution. The J/ψ peak from the model described in Section 6.1 compared with data after passing all the selections described in Chapter 4 is shown in Figure 5.9 by the orange shaded region. The scales of the distributions

are matched for illustration, but it can be seen that the natural J/ψ lineshape is too narrow to model the shape of the reconstructed resonance. Two methods are employed to handle this issue:

- Perform a kinematic fit to improve the detector resolution.
- Calculate the detector's resolution as a function of q^2 and perform a convolution to better model the data.

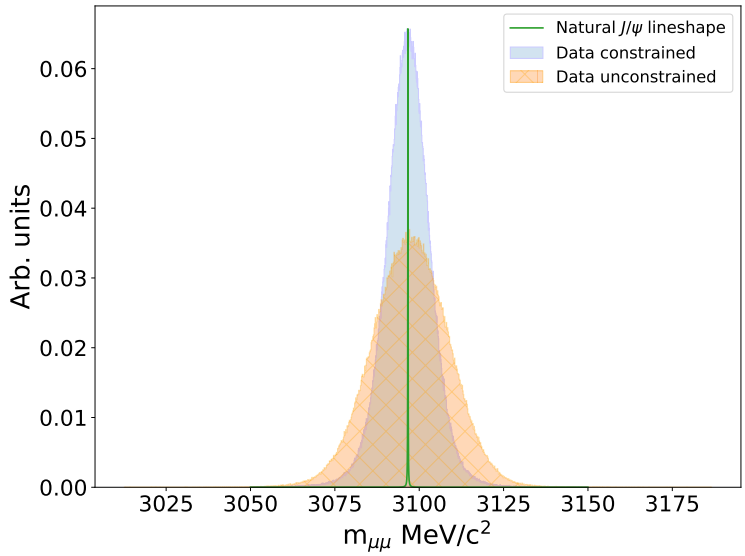

Figure 5.9: The natural line-shape of the J/ψ resonance (green) compared to 2018 data with (blue) and without (orange) the kinematic constraint on the $m_{K\mu\mu}$ mass.

5.3.1 Kinematic fit

A kinematic fit also known as a global decay chain fit as described in [128] can be performed on reconstructed tracks accounting for constraints present in the decay chain. This method improves the resolution of the computed final state particles' kinematic quantities.

In this measurement of $B^+ \to K^+\mu^+\mu^-$ decays, the reconstructed invariant mass of the $K^+\mu^+\mu^-$ system is constrained to the known particle data group (PDG) value of the B meson mass [55]. Performing a least squares fit using this information and allowing the measured momenta of the final state particles to vary in the fit, the momentum resolution of the muons can be improved thereby improving the $m_{\mu\mu}$ resolution. This can be seen in Figure 5.9 where the constrained distribution is denoted by the blue shaded region. The J/ψ resolution has improved compared to the unconstrained fit.

The dimuon mass resolution as a function of $m_{\mu\mu}$ is shown in Figure 5.10. The detector resolution shown here is estimated by calculating the Root Mean Square (RMS) width of the difference between true and reconstructed dimuon candidates from $B^+ \to K^+\mu^+\mu^-$ simulated decays in bins of dimuon mass. For the unconstrained distribution the resolution gets worse with increasing dimuon mass as the muon tracks possess a larger momentum. This is also true for the constrained distribution however, the resolution plateaus at 2500 MeV/c^2. The resolution improves at extreme dimuon masses since there is less phase space and the effect of the constraint is stronger.

For the rest of the thesis, the $m_{\mu\mu}$ mass is defined to describe the constrained dimuon mass distribution unless otherwise stated. While the kinematic fit improves the resolution, the $m_{\mu\mu}$ mass distribution is however still subject to the detector resolution effects. Encoding these effects to the physics model is the topic of discussion of the next section.

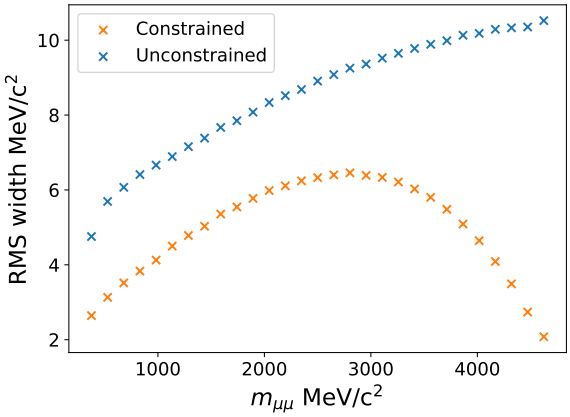

Figure 5.10: The dependence of the mass resolution on the dimuon mass (q^2 spectrum) calculated from Run1 and Run2 $B^+ \to K^+\mu^+\mu^-$ simulated samples. Shown here is the calculated RMS width of the difference between true and reconstructed candidates in bins of dimuon mass.

5.3.2 Convolution of the detector's resolution

To account for the detector's resolution a convolution is performed. The resolution of the detector as a function of dimuon mass $m_{\mu\mu}$ computed from simulation is shown in Figure 5.10.

The reconstructed dimuon mass distribution is described by,

$$P(m^{\text{rec}}_{\mu\mu}) = \mathcal{R}(m^{\text{rec}}_{\mu\mu}, m^{\text{true}}_{\mu\mu}) \circledast \left[\sum_i \mathcal{E}_i(m^{\text{true}}_{\mu\mu}) \cdot 2m^{\text{true}}_{\mu\mu} \frac{d\Gamma^i_\mu}{dq^2} \right] \tag{5.7}$$

where $m^{\text{rec}}_{\mu\mu}$ is the reconstructed dimuon mass, $m^{\text{true}}_{\mu\mu}$ is the true dimuon mass before simulating FSR effects and $\mathcal{R}$ is the resolution model of the detector. The sum runs over the three differen-

	dimuon range [MeV/c^2]	Dominant resonance
Region-1	300 to 1800	$\phi(1020)$
Region-2	1800 to 3400	J/ψ
Region-3	3400 to 4700	$\psi(2S)$

Table 5.1: The three regions chosen for the convolution and the dominant resonance within the region.

tial decay rate terms which each have a different efficiency correction $\mathcal{E}_i$ as discussed in Section 5.2. The factor $2m_{\mu\mu}^{\text{true}}$ comes from the transformation of q^2 to $m_{\mu\mu}^{\text{true}}$.

This convolution can be performed directly by evaluating the convolution integral,

$$(f \circledast \mathcal{R})(m_{\mu\mu}^{\text{rec}}) = \int_{-\infty}^{\infty} f(m_{\mu\mu}^{\text{true}}) \mathcal{R}(m_{\mu\mu}^{\text{rec}} - m_{\mu\mu}^{\text{true}}, m_{\mu\mu}^{\text{true}}) dm_{\mu\mu}^{\text{true}} \tag{5.8}$$

however, this is computationally very intensive. An alternative strategy is to perform the convolution using the Fast Fourier Transform (FFT) method expressed as,

$$f \circledast \mathcal{R} = F^{-1}(F(f)F(\mathcal{R})). \tag{5.9}$$

where F is the Fourier transformation and F^{-1} is the inverse. The limitation of using Fourier transforms to perform the convolution is that it cannot account for the detector's resolution to vary along the reconstructed dimuon mass spectrum. To mitigate this effect, the dimuon spectrum is split into three regions; 300-1800 MeV/c^2, 1800-3400 MeV/c^2 and 3400-4700 MeV/c^2 (subsequently referred to as region-1,2,3 respectively) and the convolution is performed independently in them. The convolution is implemented using TensorFlow [129] to harness its parallel processing and GPU capabilities.

5.3.3 The resolution models

The resolution model for all the three regions have been modelled as a Double Crystal Ball (DCB) added with a Gaussian. The choice of this model is validated using simulated $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \to K^+ \psi(2S)(\mu^+\mu^-)$ decays as these are the dominant contributions in each of the high dimuon mass regions. The mathematical formulation of a DCB function is presented in Appendix A. The tail parameter α_i defines the point at which the Gaussian transitions to a power law, and is chosen to be symmetric for the left and right Crystal Ball (CB) functions, *i.e.*, $\alpha_R \equiv -\alpha_L$. The resolution model values are estimated from fits to the distribution of the differences between the reconstructed and true dimuon mass ($m_{\mu\mu}^{\text{rec}}$- $m_{\mu\mu}^{\text{true}}$) in simulated events. Here the true mass distribution refers to the dimuon tracks before simulating the LHCb detector and FSR effects. This true mass of the dimuon system is calculated from the difference of the true mass of the B and K mesons, $m_B^{\text{true}} - m_K^{\text{true}}$.

For region-1, the resolution at $\phi(1020)$ is important as its natural width of 4 MeV/c^2 is comparable to the dimuon mass resolution at its pole mass. The resolution model is estimated

	Region-1	Region-2	Region-3
σ_{dcb}	2.81 ± 0.08	5.23 ± 0.10	4.34 ± 0.02
σ_{gauss}	4.71 ± 0.1	5.97 ± 0.06	5.15 ± 0.02
α_L	1.44 ± 0.08	1.16 ± 0.1	1.14 ± 0.02
n_L	1.99 ± 0.01	10.23 ± 1.7	11.28 ± 0.46
n_R	6.79 ± 0.07	31.61 ± 17.4	11.37 ± 0.47
f_{dcb}	0.52 ± 0.02	0.54 ± 0.06	0.53 ± 0.01

Table 5.2: The estimated values for the resolution model in the three dimuon mass regions from Run1 and Run2 combined simulation candidates. The α tail parameter of the DCB is forced to be symmetric, $\alpha_R \equiv -\alpha_L$.

from $B^+ \to K^+\mu^+\mu^-$ simulation with $m^{\text{rec}}_{\mu\mu}$ mass restricted to region-1. For regions-2 and -3, $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \to K^+\psi(2S)(\mu^+\mu^-)$ simulation is used to validate the model and the parameters are treated as fit parameters in the subsequent fit to data. Candidates only within $3096.7 < m^{\text{true}}_{\mu\mu} < 3098.4\,\text{MeV}/c^2$ for the $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ simulation and $3685.29 < m^{\text{true}}_{\mu\mu} < 3690.59\,\text{MeV}/c^2$ for the $B^+ \to K^+\psi(2S)(\mu^+\mu^-)$ simulations are selected. These windows were chosen in order to match the windows chosen by the simulation when generating the candidates.

An unbinned maximum likelihood fit is performed to ($m^{\text{true}}_{\mu\mu}$- $m^{\text{rec}}_{\mu\mu}$). The extracted values for the three regions for the Run1 and Run2 combined data sets are shown in Table 5.2 and the fit results displayed in Figure 5.11

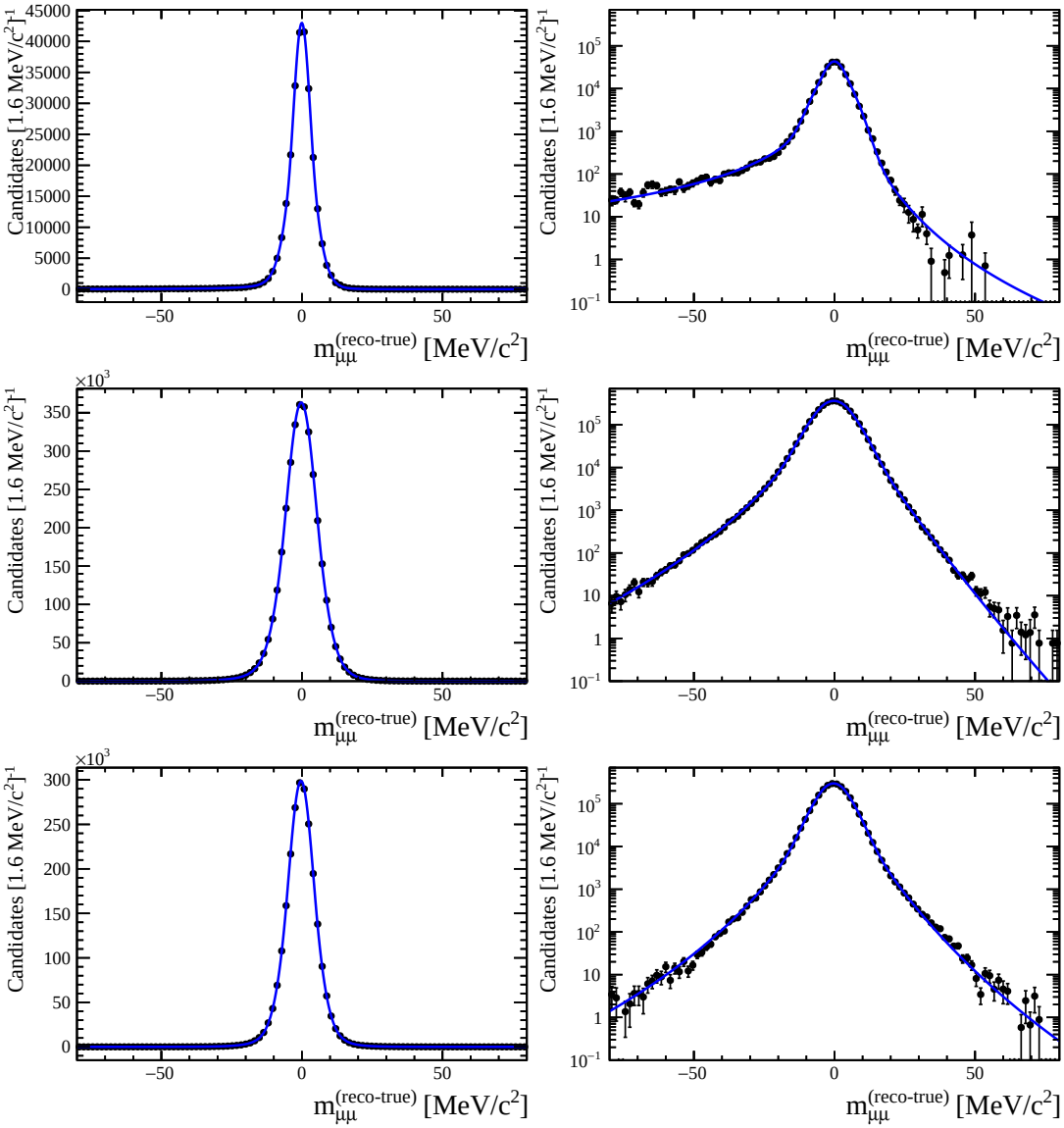

Figure 5.11: Fits to the $m^{\text{rec}}_{\mu\mu}$- $m^{\text{true}}_{\mu\mu}$ distribution for Run1 and Run2 combined simulation candidates. From the top, the different rows are fits to estimate the resolution model in region-1,2 and 3 using $B^+ \to K^+\mu^+\mu^-$, $B^+ \to J/\psi K^+$ and $B^+ \to \psi(2S)K^+$ simulation respectively. The two columns are the same plots in linear and log scale.

CHAPTER

6 Definition of the Background and Fit Model

6.1 The Differential Branching Fraction of $B^+ \to K^+\mu^+\mu^-$ decays

The analytical description of the invariant dimuon mass spectrum is described by the differential decay rate defined in Section 2.4. The analysis presented in this measurement performs an unbinned maximum likelihood fit to the invariant dimuon mass spectrum in the range $300 < m_{\mu\mu} < 4700$ MeV/c^2 to determine any NP effects which contribute to $b \to s\ell^+\ell^-$ transitions. As a reminder, the *CP* averaged differential decay rate of $B^+ \to K^+\mu^+\mu^-$ decays as a function of dimuon mass squared ($q^2 \equiv m^2_{\mu\mu}$) from Eq. 2.122 is,

$$\begin{aligned}\frac{d\Gamma_\mu}{dq^2} = &\frac{G_F^2\alpha^2|V_{tb}V_{ts}^*|^2}{2^7\pi^5}|\mathbf{k}|\beta_+\left\{\frac{2}{3}|\mathbf{k}|^2\beta_+^2\left|C_{10}f_+(q^2)\right|^2 + \frac{m_\mu^2(M_B^2-M_K^2)^2}{q^2M_B^2}\left|C_{10}f_0(q^2)\right|^2\right.\\ &\left. + |\mathbf{k}|^2\left[1-\frac{1}{3}\beta_+^2\right]\left|C_9^{\text{eff}}f_+(q^2) + 2C_7^{\text{eff}}\frac{m_b+m_s}{M_B+M_K}f_T(q^2)\right|^2\right\}, \end{aligned} \tag{6.1}$$

with all the variables defined exactly as defined in Sections 2.3-2.5.

The factors worth highlighting again are $f_{0,+,T}$, the hadronic form factors as discussed in Section 2.3.1 taken from [44]. Of the three form factors, the contribution from $f_+(q^2)$ is the largest and its coefficients are taken as fit parameters with Gaussian constraints based on their uncertainties from theoretical calculations taking into account their correlations. The other form factor coefficients are fixed. The coupling to the axial-vector current operator $\mathcal{O}_{10}$ is C_{10} and the coupling to the vector current operator $\mathcal{O}_9$ is C_9. The contribution from C_9 and the non-local hadronic contributions ($Y(q^2)$) are encoded in C_9^{eff}. The descriptions of these non-local effects are detailed in Section 2.3.2 and their magnitude (η_i) and phase (δ_i) are allowed to vary in the fit. Determining the couplings C_9 and C_{10} after explicitly accounting for the magnitude and phase of the non-local effects is the main result of this analysis. The effective contribution of the coupling to the electromagnetic dipole operator $\mathcal{O}_7$ and the operators $\mathcal{O}_{3..6}$ and $\mathcal{O}_8$ is given

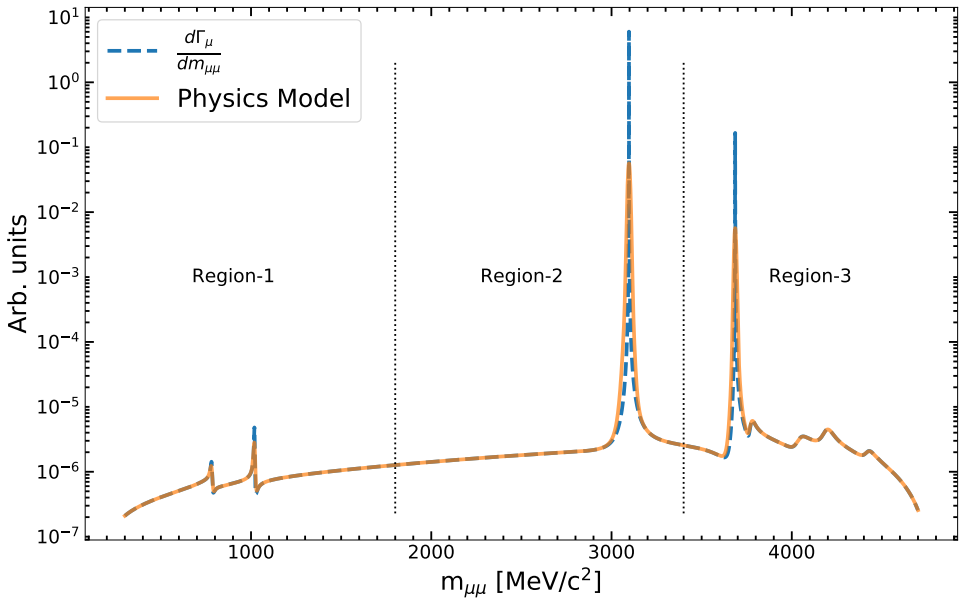

Figure 6.1: An illustration of the differential decay rate from equation Eq. 6.1 (blue) and the physics model (orange) with efficiency corrections and convolution performed in the three regions separately.

by C_7^{eff}. The SM values of these couplings as calculated in Ref. [39] are; C_7^{eff} = -0.304, C_9 = 4.211 and C_{10} = -4.103. Following the evaluation of the detector's efficiency and resolution model, the detector-effects corrected physics model is shown in Figure 6.1.

In this chapter, Section 6.2 develops the background model to account for the background contributions and combines it with the physics model. The thus created fit model is presented in Section 6.3 and validated in Section 6.4.

6.2 Accounting for the Background Contributions

The multivariate classifier discussed in Section 4.3 is designed to filter most of the combinatorial background. However, a small fraction was misidentified as signal candidates. Another background contribution is pions from $B^+ \to \pi^+\mu^+\mu^-$ decays misidentified as kaons. The method used to estimate and model these backgrounds is discussed here.

6.2.1 Measuring the signal candidate fraction

The signal candidate fraction can be estimated from the reconstructed B meson mass ($m_{K\mu\mu}$) distribution. The B meson candidates peak around 5279 MeV/c^2 with the combinatorial background spread over the full range in $m_{K\mu\mu}$. Another important contribution to be accounted for are $B^+ \to \pi^+\mu^+\mu^-$ with decays where the pion is misidentified as a kaon. This contribution manifests itself as a smaller peak at about 5330 MeV/c^2 in the $m_{K\mu\mu}$ distribution. This peak is

shifted from the nominal B meson mass due to the wrong mass hypothesis on a pion track. A maximum likelihood fit is used to extract the signal fraction and background fractions.

The $m_{K\mu\mu}$ mass fit is performed in the three dimuon mass regions discussed in Section 5.3 to better capture the small differences in the $m_{K\mu\mu}$ mass distribution along the dimuon spectrum.

6.2.2 The mass distribution model for reconstructed $m_{K\mu\mu}$

The three components of the $m_{K\mu\mu}$ distribution model are:

- The signal candidates from $B^+ \to K^+\mu^+\mu^-$ decays (inclusive of all local and non-local contributions)
- The remaining combinatorial background which manages to pass the BDT selections.
- Decays of $B^+ \to \pi^+\mu^+\mu^-$, $B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$ and $B^+ \to \pi^+\psi(2S)(\mu^+\mu^-)$ with the pion misidentified as a kaon.

The signal candidates are modelled using a DCB added with a narrow and wide Gaussian. The misidentified pion candidates are modelled using a DCB and an exponential function is used to model the combinatorial background. The analytical description of a DCB added with two Gaussian distributions is given in Appendix A. The tail parameters of the DCB; α_L, n_L α_R and n_R are extracted from a fit to $B^+ \to K^+\mu^+\mu^-$, $B^+ \to J/\psi K^+$ and $B^+ \to \psi(2S)K^+$ simulated candidates in region-1,2,3 respectively. The narrow and wide Gaussians share the same mean but are given different widths. The fractions between the two CB functions as well as the Gaussian widths and mean are determined directly from the fit to data as the simulation is known not to simulate these parameters correctly. The fractions between the signal and background components are also measured from data.

In order to account for the pions misidentified as kaons, a fit to the reconstructed $m_{K\mu\mu}$ mass of $B^+ \to J/\psi\pi^+$ simulated candidates is performed. A DCB is used to model the distribution, both tail and width parameters are estimated from this fit. The results of fits to the simulated samples are shown in Appendix B.

The result of the data fits to the $m_{K\mu\mu}$ mass distribution for Run1 and Run2 performed in the three dimuon mass regions are shown in Figure 6.2 and 6.3. The dashed blue curve is the signal contribution, the dashed red curve is the combinatorial background and the dotted green curve is the misidentified pion contribution. The combined number of the signal, misidentified and combinatorial background candidates within the signal region of $\pm$ 40 MeV/c^2 around the nominal B meson mass measured from the $m_{K\mu\mu}$ mass fits is shown in Table 6.1. The next steps involve determining the dimuon spectrum of the backgrounds within the $m_{K\mu\mu}$ signal region.

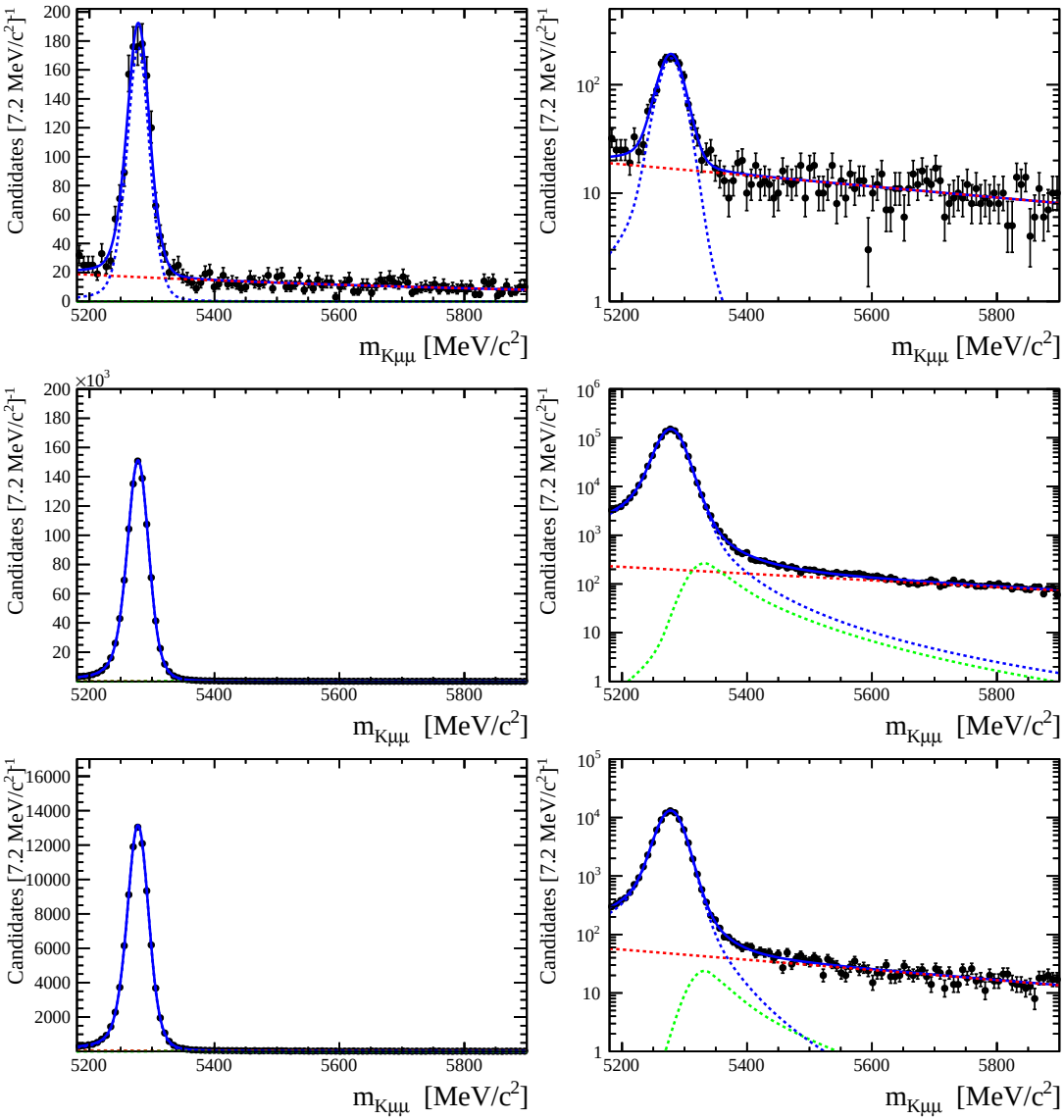

Figure 6.2: The fit results to the reconstructed $m_{K\mu\mu}$ mass distribution for Run1 data with the rows showing fits in the regions 1,2 and 3 in order as defined in the text. The columns depict y-axis in linear and log scales. The combinatorial background modelled as an exponential is given in dotted red and the contribution of pions misidentified as kaons is given in dotted green. The dotted blue curve is the DCB model discussed in Appendix A

	Run1	Run2
Signal	987130.19 ± 285.64	2529001.08 ± 357.27
Combinatorial Background	2908.01 ± 90.95	2805.26 ± 109.14
Pion misidentified	713.6 ± 90.73	1724.19 ± 102.59

Table 6.1: The combined signal and various background yields measured from $m_{K\mu\mu}$ mass fits in the three regions using Run1 and Run2 data.

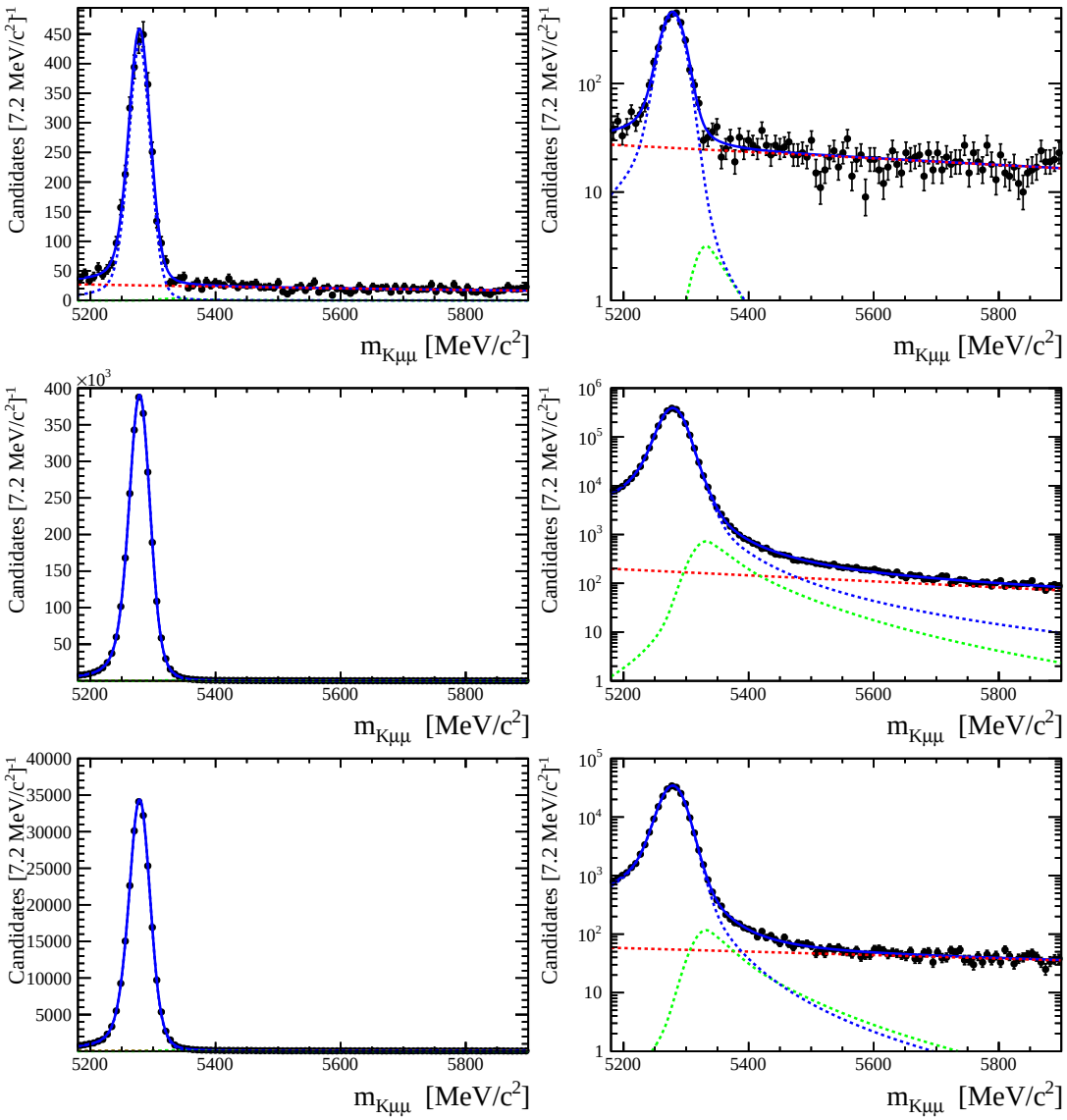

Figure 6.3: The fit results to the reconstructed $m_{K\mu\mu}$ mass distribution for Run2 data with the rows showing fits in the regions 1,2 and 3 in order as defined in the text. The columns depict y-axis in linear and log scales. The combinatorial background modelled as an exponential is given in dotted red and the contribution of pions misidentified as kaons is given in dotted green. The dotted blue curve is the DCB model discussed in Appendix A

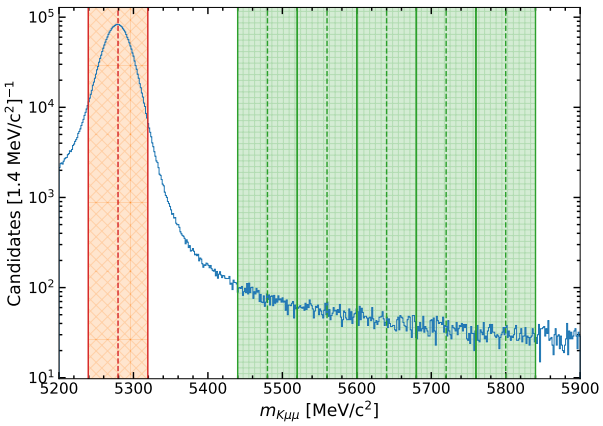

Figure 6.4: The $m_{K\mu\mu}$ mass distribution of Run2 data with the five UMSB windows marked by the shaded green region along with their constraint point. The orange shaded region is defined as the signal region.

6.2.3 Combinatorial background model

The dimuon spectrum of the combinatorial background can be extracted from the Upper Mass Sideband (UMSB). The UMSB is defined as the reconstructed $m_{K\mu\mu}$ mass region of (5440 - 5860 MeV/c^2) where the dataset is dominated by combinatorial background. However, the reconstructed $m_{\mu\mu}$ of the candidates from the UMSB cannot be relied on to estimate the shape of combinatorial background due to the presence of the kinematic constraint discussed in Section 5.3.1. The presence of this constraint distorts the dimuon mass spectrum.

In order to account for the effect of this mass constraint on the dimuon mass spectrum of the combinatorial background, five regions of the same width as the signal mass window (*i.e.* 80 MeV/c^2) are chosen and $m_{K\mu\mu}$ is constrained to the center of these windows. This mimics the effect of constraining the momenta of the reconstructed $B^+ \to K^+\mu^+\mu^-$ candidates in the signal region, to the nominal B meson mass. For illustration, the windows for Run2 data are highlighted on the $m_{K\mu\mu}$ mass distribution in Figure 6.4. In each of the five UMSB windows, the $m_{\mu\mu}$ distribution resulting from the kinematic fit of the $m_{K\mu\mu}$ system with the constraint point set to the centre of the window, is fitted using a parameterisation discussed below. The parameters of the $m_{\mu\mu}$ shape of the combinatorial background can then be determined by linearly extrapolating to the signal region each parameter of the $m_{\mu\mu}$ shape determined in the five windows of the UMSB.

The windows and the constraint point of the mass $m^{\text{con}}_{K\mu\mu}$, for Run1 and Run2 are shown in Table 6.2. The constraint points are different between the two run periods as the Run1 data set is inherited from the previous iteration of this measurement and different settings were used this time when processing the dataset. Due to this difference the background model shape is estimated separately using the same procedure for both run periods.

An ARGUS function [130] is used to parameterise the combinatorial shape over the range

Run1		Run2	
Constraint Mass (MeV/c^2)	Range (MeV/c^2)	Constraint Value (MeV/c^2)	Range (MeV/c^2)
5500	5460 to 5540	5480	5440 to 5520
5580	5540 to 5620	5560	5520 to 5600
5660	5620 to 5700	5640	5600 to 5680
5740	5700 to 5780	5720	5680 to 5760
5820	5780 to 5860	5800	5760 to 5840

Table 6.2: The different ranges in the Upper Mass Sideband (UMSB) and their constraint points for Run1 and Run2

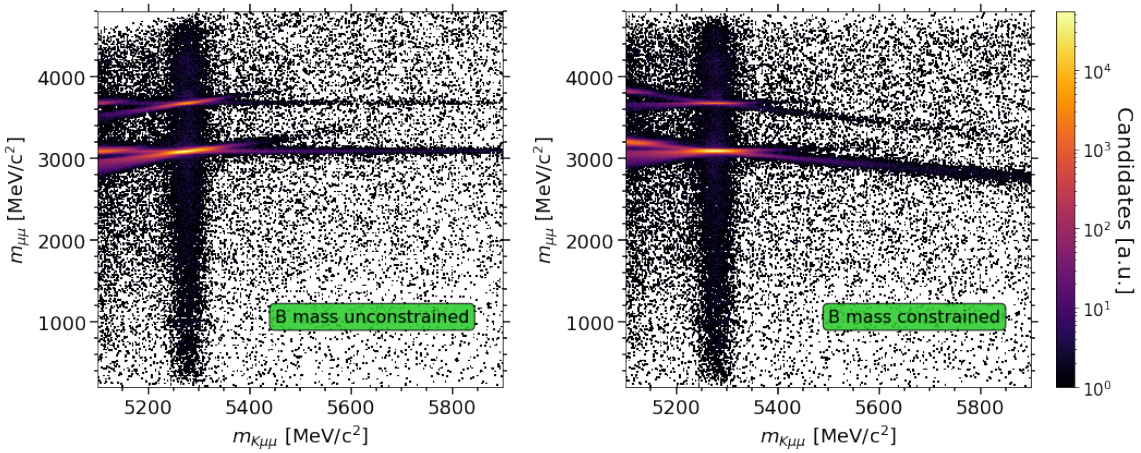

Figure 6.5: The 2D distribution of number of events in $m_{K\mu\mu}$ and $m_{\mu\mu}$ of Run2 data. The left plot is the dimuon distribution without the kinematic constraint on $m_{K\mu\mu}$ while the plot on the right has the $m_{K\mu\mu}$ mass constrained to the known B^+ PDG value when calculating $m_{\mu\mu}$.

$2m_\mu < m^{\text{rec}}_{\mu\mu} < m_{\text{t}}$ as,

$$\mathcal{P}_{\text{ARGUS}} = N(m^{\text{rec}}_{\mu\mu} - 300)\sqrt{1-\left(m^{\text{rec}}_{\mu\mu}/m_{\text{t}}\right)^2}\, e^{c\left(1-\left(m^{\text{rec}}_{\mu\mu}/m_{\text{t}}\right)^2\right)} \tag{6.2}$$

where $m_{\text{t}} = m^{\text{con}}_{K\mu\mu}$- M_K, N is a normalisation constant and c is the shape parameter which is estimated in the fit. There is also a significant combinatorial background contribution from J/ψ and $\psi(2S)$ resonances that are modelled as a double and single Gaussian respectively. A double Gaussian is the sum of two Gaussian PDFs added together according to a fraction which is estimated in the fit. One of the Gaussian is set to have a broader width than the other and this is done to model the core and tail of the peak effectively. The use of a double Gaussian was found to not be necessary for the $\psi(2S)$ as the peak is much smaller and a single Gaussian was sufficient.

Figure 6.5 shows the mass distribution of $m_{K\mu\mu}$ and $m_{\mu\mu}$. Analysing the plot without the mass constraint on the B^+ meson, the vertical band at $m_{K\mu\mu} \approx 5280$ MeV/c^2 is the signal $B^+ \to K^+\mu^+\mu^-$ decays including local and non-local contributions. The two long horizontal

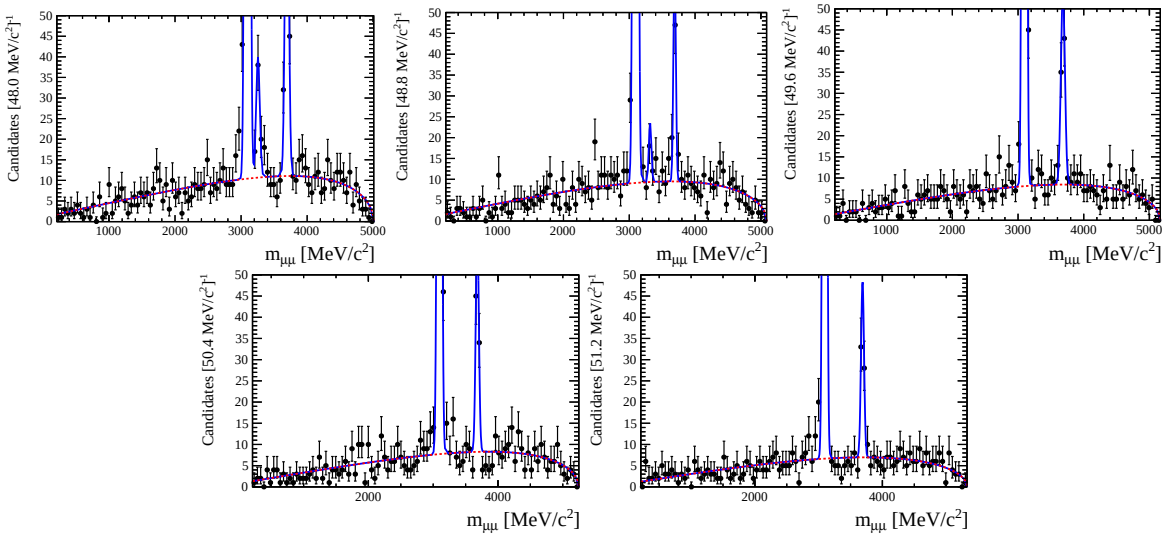

Figure 6.6: Fits to the Upper Mass Sideband (UMSB) using Run1 data in the five regions described in Table 6.2 with the $m_{K\mu\mu}$ mass constraint read in normal english reading order.

bands at $m_{\mu\mu}$ ~ 3000 and ~3700 MeV/c^2 are the J/ψ and $\psi(2S)$ combinatorial background. The two short bands are the intermediate resonance candidates of J/ψ and $\psi(2S)$. In the plot on the right it can be seen that the horizontal bands are rotated as a consequence of the kinematic constraint. The J/ψ signal candidates leak into the first two $m_{K\mu\mu}$ UMSB constraint regions between 5440 - 5620 MeV/c^2. Therefore, this contribution needs to be accounted for when estimating the background model from the UMSB. This is done by modelling it as a double Gaussian with a peak between 3225 - 3450 MeV/c^2 in the first two constraint regions.

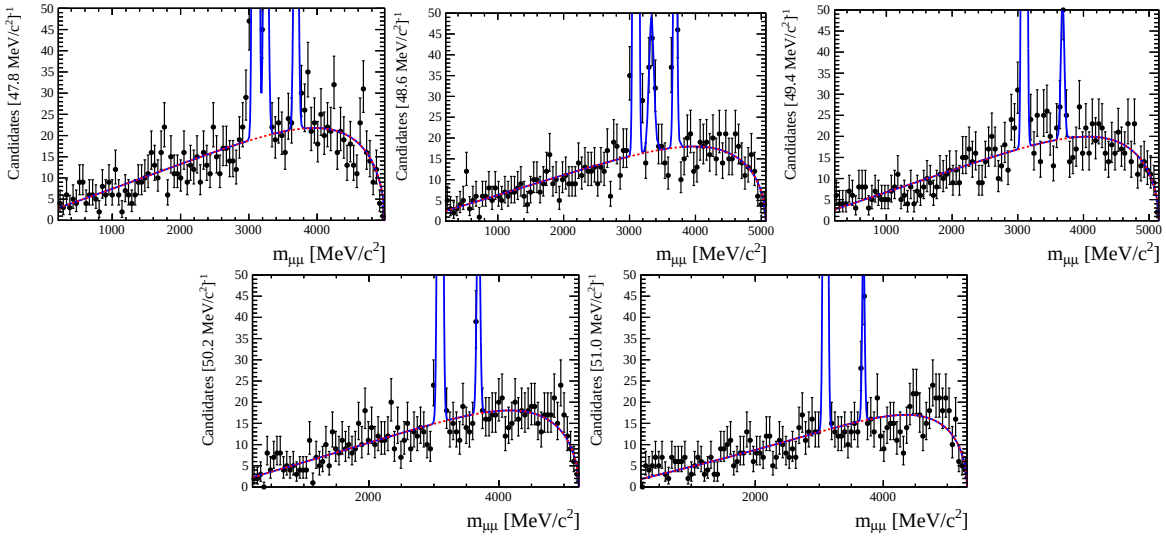

Figure 6.7: Fits to the Upper Mass Sideband (UMSB) using Run2 data in the five regions described in Table 6.2 with the $m_{K\mu\mu}$ mass constraint read in normal english reading order.

Fits to the five constraint regions using Run1 and Run2 data are shown in Figure 6.6 and 6.7 respectively. RooFit [131] was used to create the model and perform the fits. The shape parameters which need to be extrapolated to the signal region are; the ARGUS shape parameter c, the mean and widths of J/ψ and $\psi(2S)$ combinatorial and the J/ψ Gaussian fraction $G^{J/\psi}_{frac}$. A linear fit is performed to these values and the extrapolated value at the known B^+ mass (5279.25 MeV/c^2)[55] is calculated. The correlation between the different values is not taken into account. In order to estimate the errors on the extrapolated quantities correctly, the data set is bootstrapped 100 times and the fits are performed again. Bootstrapping is a procedure where the data set is resampled creating statistically semi-independent data samples from the original data. Replacement of values is allowed during bootstrapping. The total number of entries in each bootstrapped data set is fluctuated according to a Poisson distribution.

In addition to the shape parameters, the fractions between the J/ψ, $\psi(2S)$ and the ARGUS combinatorial contribution needs to be estimated in the signal region as well. Recall from Section 6.2.2 that the dimuon range was split into three in order to better model the shape of the B^+ mass peak. For this reason, the signal fraction is known separately in the three dimuon mass regions. Therefore, the combinatorial background fractions of J/ψ and $\psi(2S)$ are calculated within region-2 and region-3 of the dimuon mass spectrum respectively. This is then extrapolated to the signal region to be combined with the signal fractions from the appropriate regions to define the combinatorial background model (B_{comb}). The J/ψ and $\psi(2S)$ combinatorial fractions within specific region of the dimuon spectrum are not fit quantities in the combinatorial background fits. In order for the extrapolated value to account for the fit uncertainties, bootstrapped data sets are used to estimate these fractions, the median of the extrapolated value is taken as the nominal value. The result of the extrapolation procedure is shown in Figure 6.8 and 6.9 for Run1 and Run2 respectively. The error bands correspond to the 1 σ errors and are the 16th and 84th percentile of the distribution of extrapolated values from fits to bootstrapped datasets.

6.2.4 Background from pions misidentified as kaons

In the signal region of interest, the other dominant background sources are $B^+ \to \pi^+\mu^+\mu^-$, $B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$ and $B^+ \to \pi^+\psi(2S)(\mu^+\mu^-)$ decays with the pion misidentified as a kaon. The branching fractions of these decay modes are 1.75×10^{-8}, 2.3×10^{-6} and 1.95×10^{-7} respectively [55]. The fraction which manage to pass the PID requirements is estimated in the $m_{K\mu\mu}$ mass fits in Section 6.2.2 and is shown by the green dotted curves in Figures 6.2 and 6.3.

In region-1, the number of misidentified candidates are from non-resonant $B^+ \to \pi^+\mu^+\mu^-$ decays and is of $\mathcal{O}$ (1), in region-2 and 3 the misidentified candidates are of $\mathcal{O}$ (1000) and $\mathcal{O}$ (100) respectively. Due to the small relative branching fraction of the non-resonant modes and since it is spread over the full $m_{\mu\mu}$ mass range, the non-resonant contribution is assumed to be negligible and not studied further. This assumption means that the misidentified candidates in

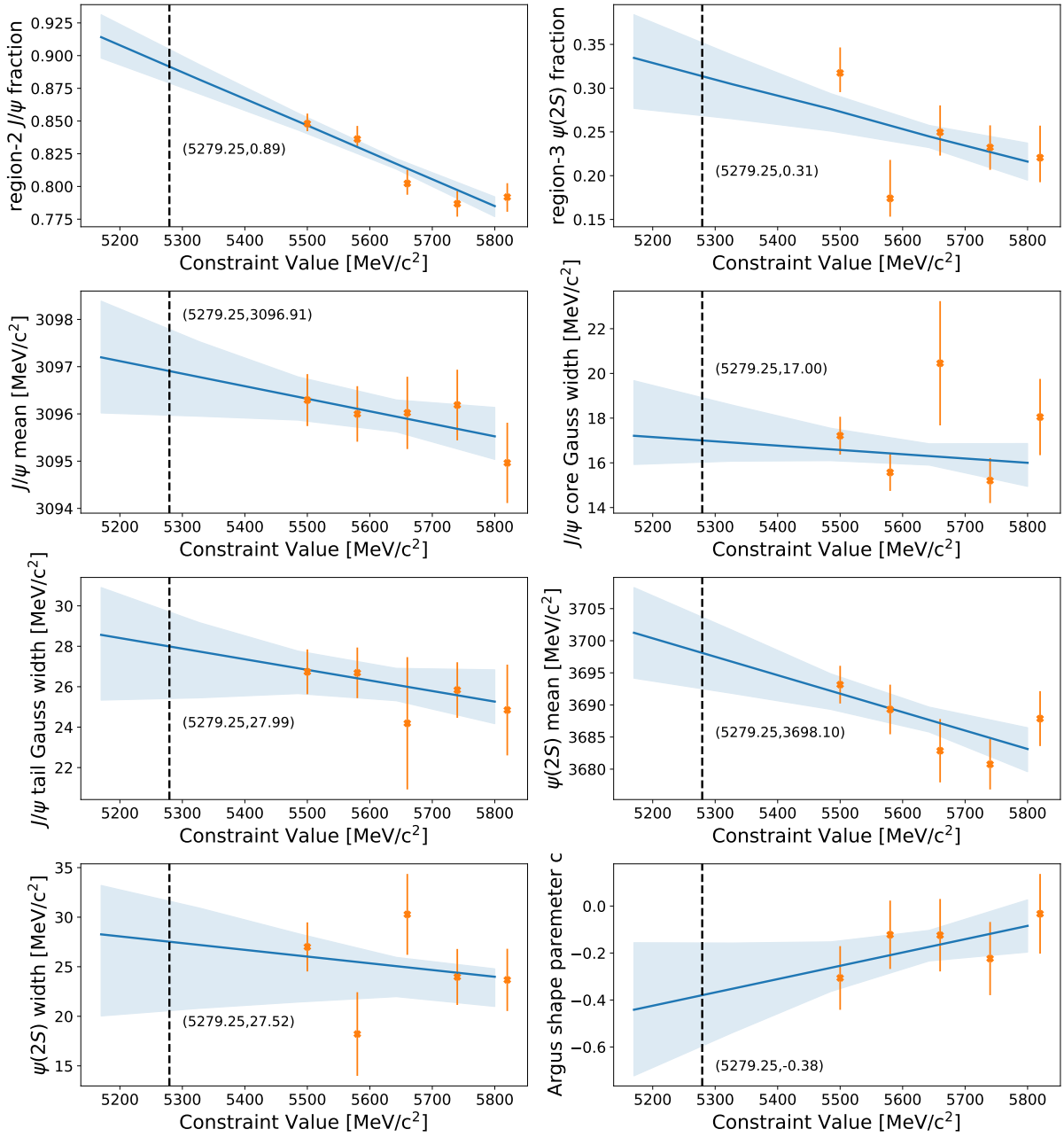

Figure 6.8: Run1 extrapolated results of the yield fractions and shape parameters for the combinatorial model. The orange points are the result from fits to the true data set. The lightblue bands are formed from the 16th and 84th percentile of the bootstrapped results corresponding to a 1σ error band. J/ψ and $\psi(2S)$ fractions were take from the median value of bootstrapped data at the B meson mass while the rest of the parameters were extrapolated from the fit to data.

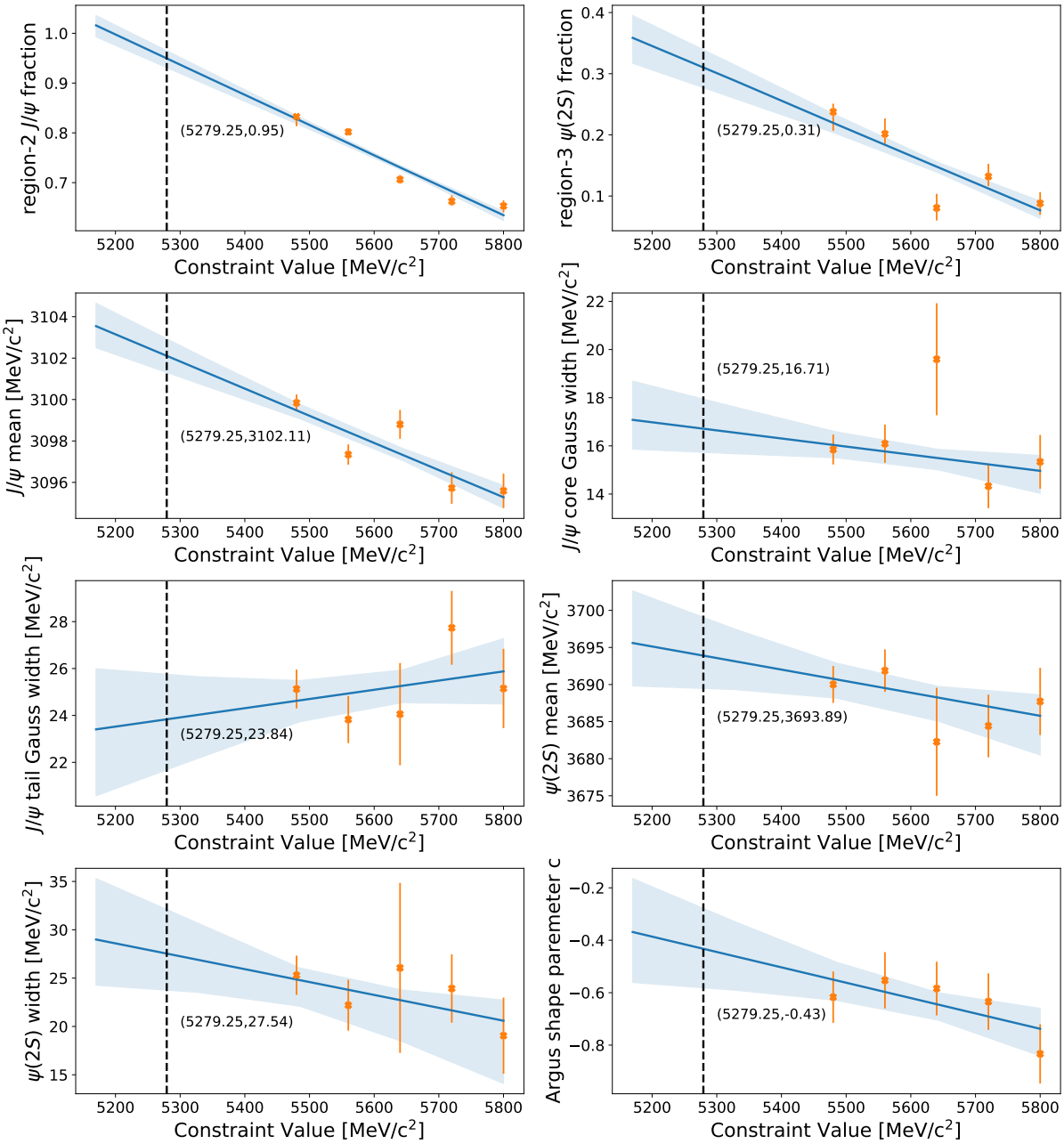

Figure 6.9: Run2 extrapolated results of the yield fractions and shape parameters for the combinatorial model. The orange points are the result from fits to the true data set. The lightblue bands are formed from the 16th and 84th percentile of the bootstrapped data results corresponding to a 1σ error band. J/ψ and $\psi(2S)$ fractions were take from the median value of bootstrapped at the B meson mass while the rest of the parameters were extrapolated from the fit to data.

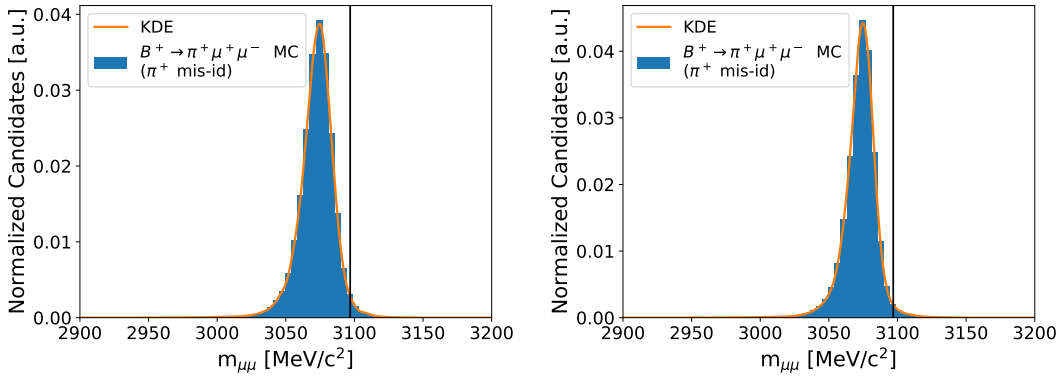

Figure 6.10: Dimuon mass distribution from $B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$ simulated candidates from Run1 (left) and Run2 (right). A KDE is used to estimate the shape of the misidentified background, the mass of the J/ψ resonance is marked by the vertical black line.

region-2 and 3 are only from the J/ψ and $\psi(2S)$ resonant modes. From the $m_{K\mu\mu}$ mass fits the total misidentified background is estimated to be 50% of the total combinatorial background for Run1 and Run2.

To determine the shape of the pion misidentified background (B_π), the dimuon distribution of $B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$ simulation from within the signal region is fitted with a Gaussian KDE as shown in Figure 6.10 for Run1 and Run2. A KDE allows the estimation of a Probabilty Density Function (PDF) in a non-parametric way using essentially a sum of kernel functions for each data point. Here a Gaussian kernel is used and the width of this kernel is optimized for the input distribution. In Figure 6.10, the peak of this distribution is shifted to the left of the known J/ψ mass of 3096.9 MeV/c^2 due to kaon mass hypothesis on a pion track during track reconstruction. The shape of the $\psi(2S)$ misidentification background is assumed to be identical to the J/ψ shape with a shift in the dimuon mass by 589.2 MeV/c^2 corresponding to the difference between the known J/ψ and $\psi(2S)$ mass.

6.2.5 The full background model

Using the measured fractions of combinatorial and misidentified background (f_{comb} and f_π) in Section 6.2.2, the extrapolated shape for both the run periods is shown in Figure 6.11. Here, the bands correspond to the spread of the extrapolated bootstrapped model. The noise seen at 2750 MeV/c^2 arises from the KDE fit and is negligible so no further correction is applied. The PDF of the total background contribution is then modelled as the sum of the individual component PDFs according to their relative fractions. Furthermore, since eventually the Run1 and Run2 data sets will be combined for the final fit, their respective background PDFs are combined based on their relative fractions. The full background model is shown in Figure 6.12 where the Run1 and Run2 contributions are also highlighted. With this estimate, everything

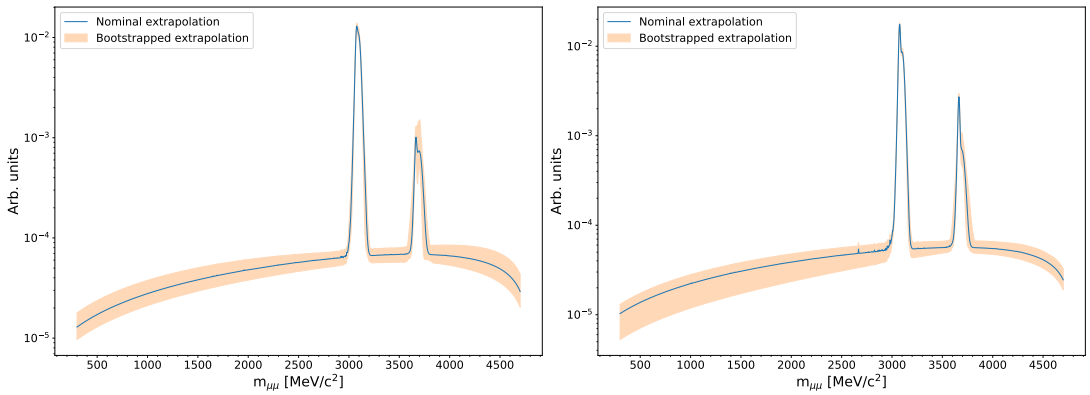

Figure 6.11: The extrapolated nominal background model (blue) and the spread of extrapolated model from bootstrapped data set (orange band) for Run1 (left) and Run2 (right) data set.

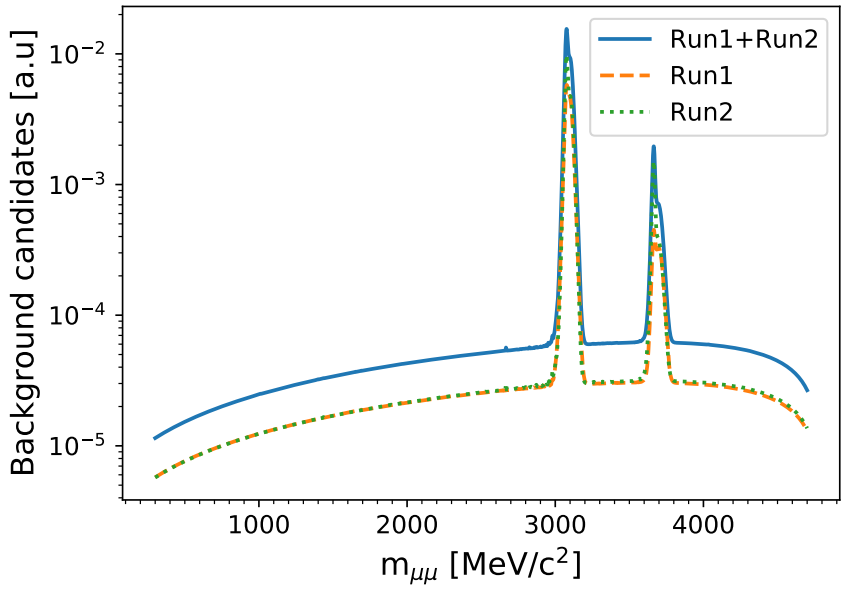

Figure 6.12: The background contribution to the dimuon mass spectrum estimated from the UMSB and $m_{K\mu\mu}$ mass fits from Run1 and Run2 data.

needed to fit the differential decay rate to the dimuon spectrum is in place. In the next section the fit strategy is discussed along with studies performed to validate the fitter.

6.3 Implementation of the Fitter

With the physics model defined and the detector's efficiency estimated along with the background contributions, a fit to data can be performed. The combined Run1 and Run2 data set contains 3524407 candidates within the fit window ($300 < m_{\mu\mu} < 4700\,\text{MeV}/c^2$) and signal region ($5239.25 < m_{K\mu\mu} < 5319.25\,\text{MeV}/c^2$). A summary of the various components of the fit

model and the maximum likelihood estimation method is presented in Appendix C.

Wilson coefficients C_9 and C_{10} are fit parameters and C_7^{eff} is fixed to its SM value of C_7^{eff} = -0.304 [39]. The values of C_9 and C_{10} visible to the experimenter are shifted by random value in order to avoid any potential observation bias until the fitter is fully understood. The form factor coefficients of $f_+(q^2)$, $b^+_{0,1,2}$ are fit parameters subject to Gaussian constraints based on the value and uncertainties including their correlations from [44]. The form factor coefficients of $f_0(q^2)$ and $f_T(q^2)$ are fixed to the values given in [44]. This can be done as they have a limited impact on the shape of the differential branching fraction defined in Eq. 6.1.

The pole mass and width of all the resonances, except for J/ψ and $\psi(2S)$, are initialized to their known values. Table 6.4 shows values for the light 1P non-local states and Table 6.5 shows the values for the charmonium 1P and 2P non-local states. Parameters varied with Gaussian constraints based on their uncertainties are highlighted. Since J/ψ and $\psi(2S)$ have a large magnitude they are very sensitive to the position of their pole mass. To account for any residual uncertainty on the momentum scale in the data, the pole masses of J/ψ and $\psi(2S)$ are taken as a fit parameters in the fit and treated as nuisance parameters. The natural width of the J/ψ and $\psi(2S)$ are fixed to the known values since they are very narrow and detector resolution effects dominate.

The resolution parameters for region-2 and region-3 are included as fit parameters as the presence of the large narrow resonances (J/ψ and $\psi(2S)$) means that it is crucial to not rely on simulation to avoid potential mismodelling. The parameters include the two width parameters, the fraction between the DCB and Gaussian f_{dcb} and the α parameter describing the tails of the resolution. The tail parameters n_L and n_R are taken from the fits to simulation discussed in Section 5.3.3 and are given in Table 5.2. A description of these parameters can be found in Appendix A where a mathematical formulation of the DCB line shape is also shown. The resolution parameters of region-1 are not estimated from data as the resolution effects are smaller and the estimated values from simulation are assumed to give a sufficiently accurate model. The resolution fit parameters are summarised in Table 6.6.

The Branching Fraction (BF) $\mathcal{B}(B^+ \to J/\psi K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)$ is used to normalise the local and other intermediate non-local contributions. The value $\mathcal{B}(B^+ \to J/\psi K^+) = (9.95 \pm 0.32) \times 10^{-4}$ which accounts for isospin asymmetries in the production of B^+ at the $\Upsilon(4S)$ resonance is used [132]. Multiplied with $\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.96 \pm 0.03) \times 10^{-2}$ taken from Ref. [55] then gives the $\mathcal{B}(B^+ \to K^+ J/\psi(\mu^+\mu^-))$. The uncertainties on these values will be translated to systematic uncertainties on the measured values. The magnitudes of the intermediate states $\rho(770)$, $\omega(782)$ and $\phi(1020)$ are fixed to the values in Table 6.4 since their branching fractions are very well known while their phases are fit parameters. The magnitudes and phases of the charmonium resonances: J/ψ, $\psi(2S)$, $\psi(3770)$, $\psi(4040)$, $\psi(4160)$ and $\psi(4415)$ are estimated from the fit. Since the 2P non-local states (DD, DD^* and D^*D^*) have a similar amplitude structure, in the first instance, these states are combined to have an equal contribution. This means that there is only

Type	Parameter Name	Value	Remarks
Wilson	C_9	-	Fit Parameter
	C_{10}	-	Fit Parameter
	C_7^{eff}	-0.304 [39]	Fixed
Form Factor	b_0^+	0.466 ± 0.014 [44]	Gauss Constrained
	b_1^+	-0.89 ± 0.13 [44]	Gauss Constrained
	b_2^+	-0.21 ± 0.55[44]	Gauss Constrained

Table 6.3: A summary of the local parameters which define the fit model. Highlighted in yellow are the parameters estimated from the fit, and highlighted in blue or red are treated as fit parameters with Gaussian constraints or fixed respectively.

Type	Parameter Name	Value	Remarks
$\rho(770)$	Mass	775.3 MeV/c^2 [55]	Fixed
	Width	149.1 MeV/c^2 [55]	Fixed
	Phase	-	Fit Parameter
	BF	1.7×10^{-10} [55]	Fixed
$\omega(782)$	Mass	782.7 MeV/c^2 [55]	Fixed
	Width	8.5 MeV/c^2 [55]	Fixed
	Phase	-	Fit Parameter
	BF	4.9×10^{-10} [55]	Fixed
$\phi(1020)$	Mass	1019.4 MeV/c^2 [55]	Fixed
	Width	4.3 MeV/c^2 [55]	Fixed
	Phase	-	Fit Parameter
	BF	2.5×10^{-9} [55]	Fixed

Table 6.4: A summary of the non-local parameters for the light 1P states which define the fit model. Highlighted in yellow are the parameters estimated from the fit, and highlighted in blue or red are treated as fit parameters with Gaussian constraints or fixed respectively.

one magnitude and one phase fit parameter for the 2P states.

6.4 Validation of the Fitter

As with any parameter estimation through a fit to data, it is important to establish the level of bias and statistical coverage of the measurement and its uncertainties. The typical way of validating the fit performance is through what is known as a "pull study"

To perform a pull study, a PDF is created using the physics model, background model and detector efficiency discussed earlier. Pseudodata commonly referred to as a toy data set is generated from this PDF and the fit model summarised in Section 6.3 is used to fit to it. The values for the fit parameters are taken from Ref. [12], the previous iteration of this measurement keeping C_9 and C_{10} at the SM values. This in principle simulates the experiment but using

Type	Parameter Name	Value	Remarks
J/ψ	Mass	3096.6 MeV/c^2 [55]	Fit Parameter
	Width	0.09 MeV/c^2 [55]	Fixed
	Phase	-	Fit Parameter
	BF	$(5.93 \pm 0.19) \times 10^{-5}$ [55, 132]	Fixed
$\psi(2S)$	Mass	3685.9 MeV/c^2 [55]	Fit Parameter
	Width	0.29 MeV/c^2 [55]	Fixed
	Phase	-	Fit Parameter
	Magnitude (BF)	-	Fit Parameter
$\psi(3770)$	Mass	3773 MeV/c^2 [133]	Fixed
	Width 1	0.29 MeV/c^2 [55]	Fixed
	Width 2	27.2 MeV/c^2 [55]	Fixed
	Phase	-	Fit Parameter
	BF	$(4.7 \pm 1.3) \times 10^{-9}$	Gauss Constrained
$\psi(4040)$	Mass	4039 ± 1 MeV/c^2 [133]	Gauss Constrained
	Width	80 ± 10 MeV/c^2 [133]	Gauss Constrained
	Phase	-	Fit Parameter
	Magnitude (BF)	-	Fit Parameter
$\psi(4160)$	Mass	4192 ± 7 MeV/c^2 [133]	Gauss Constrained
	Width	72 ± 12 MeV/c^2 [133]	Gauss Constrained
	Phase	-	Fit Parameter
	Magnitude (BF)	-	Fit Parameter
$\psi(4415)$	Mass	4421 ± 4 MeV/c^2 [133]	Gauss Constrained
	Width	62 ± 20 MeV/c^2 [133]	Gauss Constrained
	Phase	-	Fit Parameter
	Magnitude (BF)	-	Fit Parameter
2P States	Phase	-	Fit Parameter
(DD, D^*D^*, DD^*)	Magnitude (BF)	-	Fit Parameter

Table 6.5: A summary of the non-local parameters for the charmonium 1P and 2P states which define the fit model. Highlighted in yellow are the parameters estimated from the fit, and highlighted in blue or red are treated as fit parameters with Gaussian constraints or fixed respectively.

Type	Parameter Name	Value	Remarks
Region-1	σ_{dcb}	2.81	Fixed
	σ_{gauss}	4.71	Fixed
	α_L	1.44	Fixed
	n_L	1.99	Fixed
	n_R	6.79	Fixed
	f_{dcb}	0.52	Fixed
Region-2	σ_{dcb}	-	Fit Parameter
	σ_{gauss}	-	Fit Parameter
	α_L	-	Fit Parameter
	n_L	10.23	Fixed
	n_R	31.61	Fixed
	f_{dcb}	-	Fit Parameter
Regions-3	σ_{dcb}	-	Fit Parameter
	σ_{gauss}	-	Fit Parameter
	α_L	-	Fit Parameter
	n_L	11.28	Fixed
	n_R	11.37	Fixed
	f_{dcb}	-	Fit Parameter

Table 6.6: A summary of the resolution parameters which define the fit model. Highlighted in yellow are the parameters estimated from the fit. The fixed parameter values highlighted in red are taken from Table 5.2.

data generated with known parameter values. The fitter is studied by analysing how the fitted parameter values differ from the generation values. For this the pull statistic for each parameter X is defined as,

$$X_{pull} = \frac{X_{true} - X_{fit}}{X_{error}}. \tag{6.3}$$

Here X_{true} is the generation value for the toy, X_{fit} and X_{error} are the fitted value and its estimated error, in this case reported by the Hessian calculation of the covariance matrix at the minimum of the likelihood. For each parameter the pull statistic is calculated from 200 toy data sets. The pull statistic distributions would be Gaussian distributed with a mean of 0 and width of 1 for all the parameters if the fit is working well.

The total number of entries in the toy data set are fluctuated according to a Poisson distribution with the expectation value being the total number of candidates in data within our fit window. The constraining point of parameters with Gaussian constraints are also Gaussian fluctuated according to their errors. This is done in order to simulate the uncertainty in the external measurement which provides these values [134].

The distribution of the pull statistic for C_9, C_{10} and the distribution of the fitted values is shown in Figure 6.13. The distributions are fitted with a Gaussian function shown by the orange curve. The summary of the pull distributions for all the parameters of interest is shown

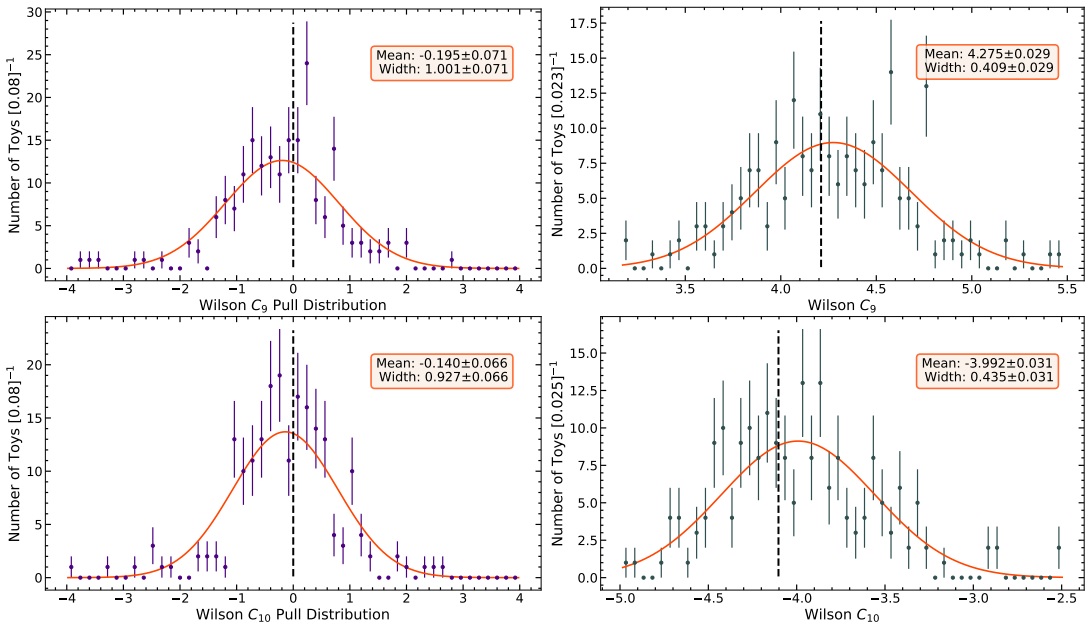

Figure 6.13: The distribution of the pull statistic (left) and fitted values (right) for Wilson coefficients C_9 (top) and C_{10} (bottom). The toys were generated at the SM values for C_9 = 4.211 and C_{10} = -4.103.

in Figure 6.14. In this plot the central value is the mean of the Gaussian function fitted to the pull distributions in Figure 6.13 and the length of the error bars are the width of the Gaussian. The colour band on the central value and the edges are the associated errors from the fits to the pull distribution.

From the summary plots it can be noticed that all the parameters have pull distributions within uncertainties except for Wilson C_9, C_{10} and $\delta_{\psi(2S)}$. The distribution of pulls and fitted values for $\delta_{\psi(2S)}$ is shown in Figure 6.15. Due to the high correlation of the Wilson coefficients with the other parameters, the small bias present is unsurprising. Since the expected precision of these parameters are high, the residual bias can be corrected for and accounted as a systematic effect. Since the residual distributions of the $\delta_{\psi(2S)}$ is accurate within the available statistics, the bias in the pull statistic is indicative of an asymmetric error interval.

Another complimentary method to study the fitter is by analysing the coverage of distribution of the fitted value. The coverage test checks the number of times the Hessian error intervals around the best fit point containing the true value follows a Gaussian expectation for the interval. The coverage test plots for these parameters in various n-σ intervals is shown in Figure 6.16 Here error bars are estimated assuming a binomial distribution as the true value can fall in or out of a given interval. The curves show the expected fractions at a given n-σ interval.

The over-coverage present for C_{10} and $\delta_{\psi(2S)}$ is another strong indicator of an asymmetric error interval. The coverage plot would show an improved behaviour if the intervals are

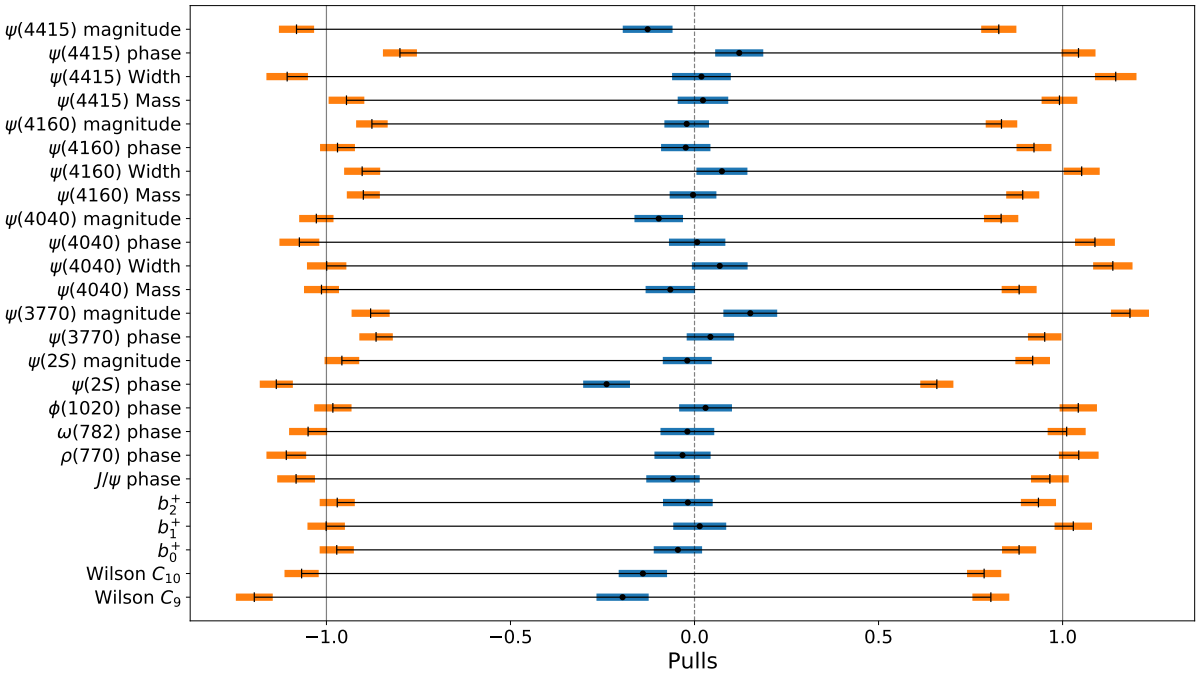

Figure 6.14: The summary of the pull study using 200 toys generated at the values from the previous measurement. SM values are used for the Wilson coefficients. The central point and the error bars indicate the mean and width of the pull distributions. The color band on these points indicate the error on the mean and width from the fit to the pull distributions.

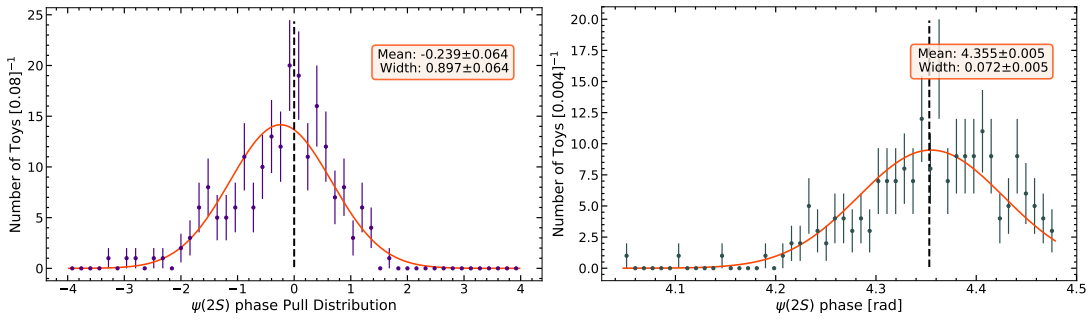

Figure 6.15: The distribution of the pull statistic (left) and fitted values (right) for the phase of $\psi(2S)$ using 200 toys. The toys are generated at the previously measured value of $\delta_{\psi(2S)} = 4.353$ rad. [12].

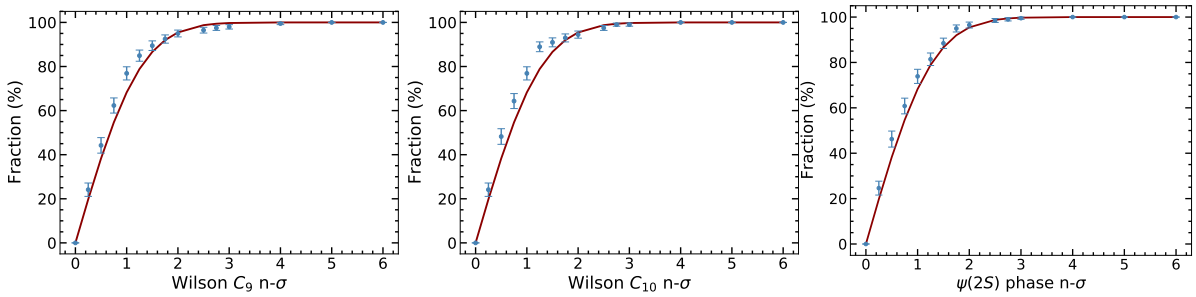

Figure 6.16: The coverage check of C_9 (left), C_{10} (centre) and $\delta_{\psi(2S)}$ (right) using 200 toys generated at the values from the previous measurement. SM values are used for the Wilson coefficients. The blue data points are the calculated coverage from fit to toy data sets and the brown curve is the expected coverage at the n-σ interval.

defined by profiling a particular parameter to various $\Delta\log\mathcal{L}$ values. However, given the complexity of the fit, this would require extreme computational power. From the fit to data, the error intervals of these parameters will need to estimated from likelihood profiles.

In order to understand the effect of the 2P states, 200 toys with a 2P magnitude η_{2P} = 0.2 and 2P phase δ_{2P} = -3 rad were generated, the resulting summary of the pull distributions is shown in 6.17. This shows a similar structure to the plot in Figure 6.14. The coverage test of C_9 , C_{10} , $\delta_{J/\psi}$ and $\delta_{\psi(2S)}$ are shown in Figure 6.18. The 1-σ error intervals of these parameters are accurate within the allowed statistical precision, reiterating at a need for appropriate bias correction for C_9 and C_{10} and error interval estimates from likelihood profiles. When the number of fit parameters is restricted to only the local parameters and the dominant non-local parameters, the distribution of the pull statistic has improved as shown in Figure 6.19.

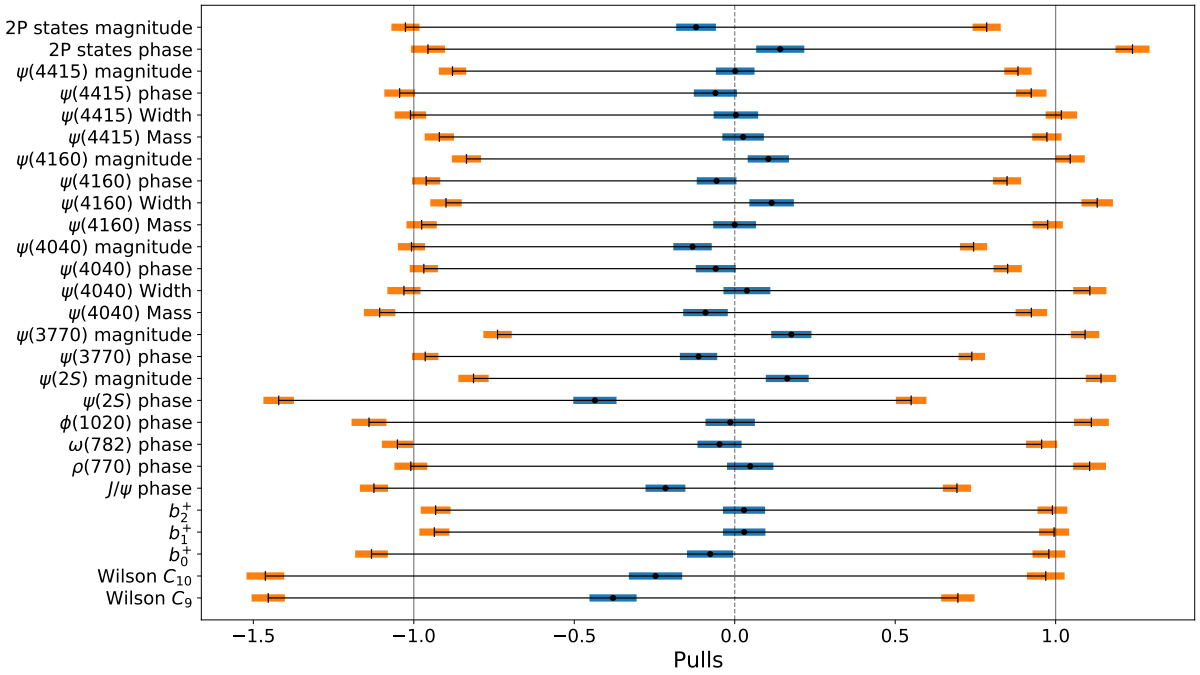

Figure 6.17: The summary of the pull study using 200 toys generated by including and fixing the 2P state magnitude to 0.2 and phase to -3 rad. The central point and the error bars indicate the mean and width of the pull distributions. The color band on these points indicate the error on the mean and width from the fit to the pull distributions.

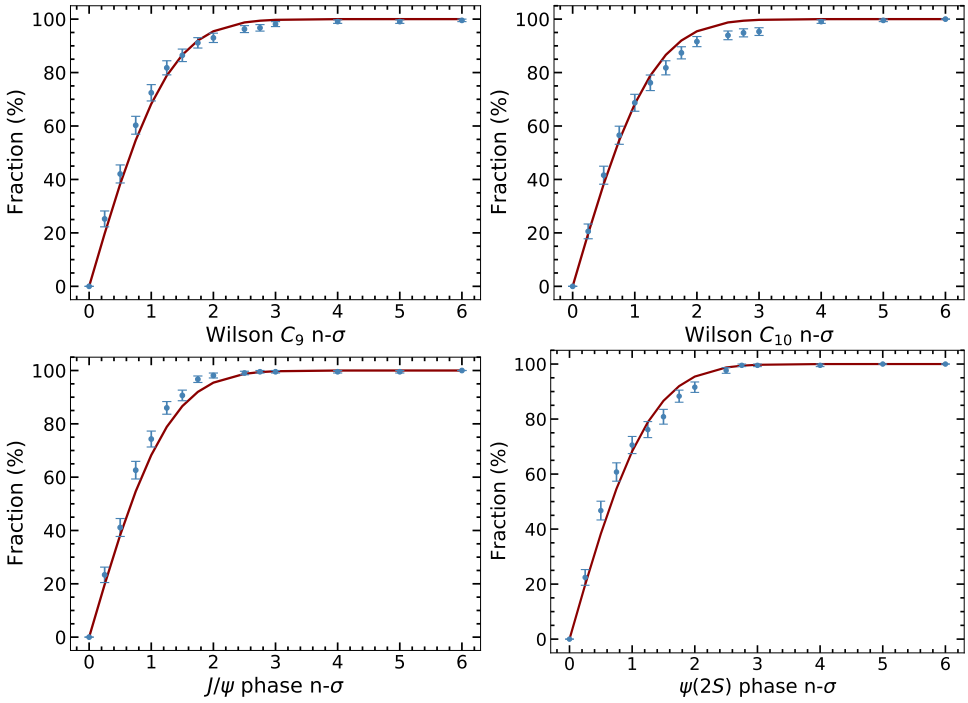

Figure 6.18: The coverage check of C_9 (top left), C_{10} (top right), $\delta_{J/\psi}$ (bottom left) and $\delta_{\psi(2S)}$ (bottom right) using 200 toys using 200 toys. The blue data points are the calculated coverage from fit to toy data sets and the brown curve is the expected coverage at the n-σ interval. The toys are generated with SM values for Wilson coefficients C_9 = 4.211 and C_{10} = -4.103 and previously measured value of $\delta_{J/\psi}$ = 4.623 rad and $\delta_{\psi(2S)}$ = 4.353 rad. [12]. The 2P states are included with a magnitude of 0.2 and phase of -3 rad.

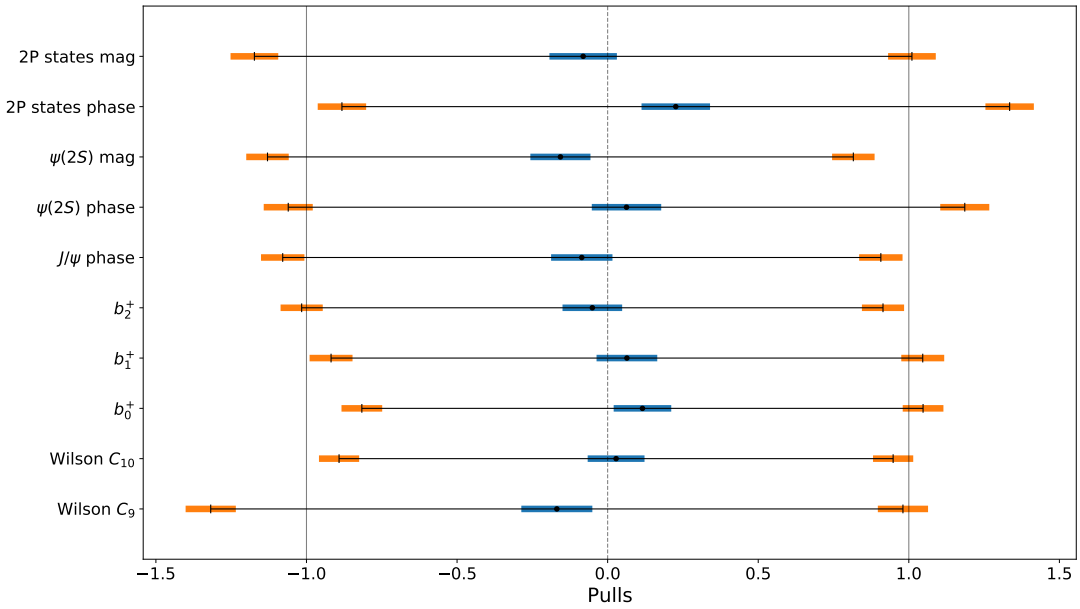

Figure 6.19: The summary of the pull study using 100 toys and only allowing the Wilson and form factor coefficients; 2P state and $\psi(2S)$ magnitude; J/ψ, $\psi(2S)$ and 2P phases to vary. The central point and the error bars indicate the mean and width of the pull distributions. The color band on these points indicate the error on the mean and width from the fit to the pull distributions.

CHAPTER

7 Results

Having determined the resolution model, the background model and signal yield, Chapter 6 described the procedure of the fit to the dimuon invariant mass distribution and how the fit was validated. The fit performed to LHCb Run1 and Run2 data is shown in Figure 7.1 and the results presented in Table 7.1 and Table 7.2. For the purpose of facilitating the discussion in this chapter, this result will be referred to as the nominal result. At this stage of the analysis review, the key parameters of interest, C_9 and C_{10} , which could reveal the existence of any potential BSM physics are hidden from the experimenter to avoid potential observation biases. This procedure is commonly referred to as blinding. Note that throughout this thesis, the quoted values of Wilson C_9 and C_{10} contains the redefinition $C_9 \equiv C_9 + C_9^{'}$ and $C_{10} \equiv C_{10} + C_{10}^{'}$, the sum of the unprimed and primed coefficients of the chirality flipped operators (See Section 2.2). The analysis is performed to a CP averaged data which makes it implicitly insensitive to Im[C_9] and Im[C_{10}].

Some early observations from the fit shown in Figure 7.1 are that the interference effect between the local and non-local contributions is large with a sizable effect even in the range $m_{\mu\mu} < 2800$ MeV/c^2. This is in contrast to the previous iteration of this measurement from the LHCb collaboration presented in Ref. [12]. The major difference between the two measurements is the use of a dispersion model which allows for a mechanism to include the 2P non-local charmonium states (DD, DD^*, D^*D^*) in a theoretically coherent way. The measured 2P state magnitude is $\eta_{2P} = 0.67 \pm 0.20$ with a phase difference of $\delta_{2P} = -3.05 \pm 0.13$ rad w.r.t. to the local contribution. The other potential source of large interference is from J/ψ and $\psi(2S)$ 1P states, both of which have a very large branching fraction. The phase difference of the J/ψ and $\psi(2S)$ with the local contribution are $\delta_{J/\psi} = -1.71 \pm 0.08$ and $\delta_{\psi(2S)} = -2.81 \pm 0.58$ respectively. These numbers reveal a fully destructive interference from the $\psi(2S)$ and 2P states with the local contribution but a minimal interference with the J/ψ.

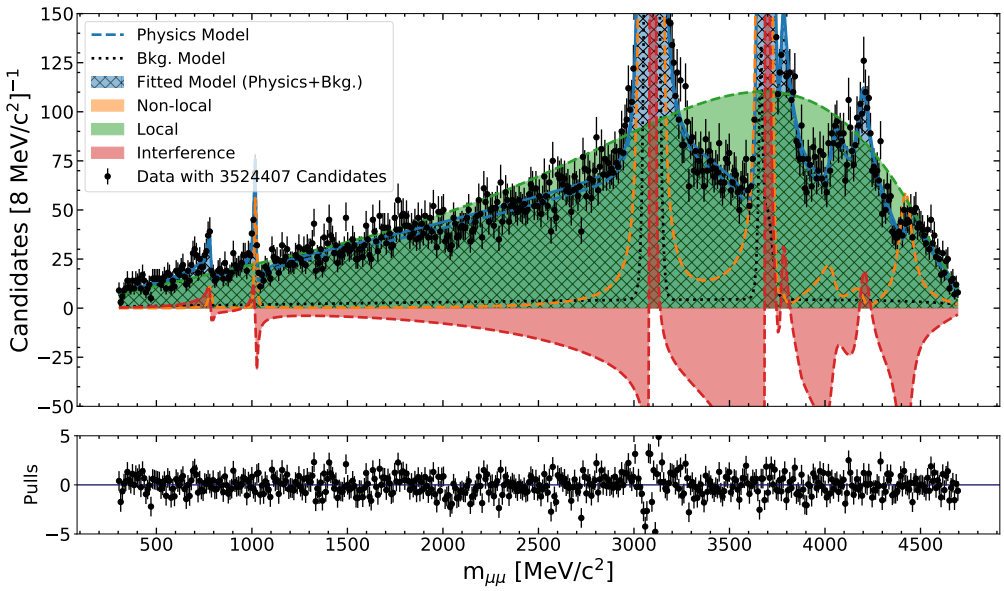

Figure 7.1: The fit to the dimuon invariant mass of $B^+ \to K^+\mu^+\mu^-$ decays using the fit model described in Chapter 6 and data collected by the LHCb experiment during Run1 and Run2. The background (Bkg.) model and the convolved physics model are highlighted. The local contributions of the physics model which are parametrised by the Wilson and form factor coefficients are shown by the shaded green region. The non-local effects produced by the intermediate 1P and 2P states and their interference with the local contribution are shown in orange and red respectively. The fit parameters obtained from this fit are treated as the nominal result.

In Figure 7.1, the sub axis below the main plot shows the pulls of the binned data. The pull values in each histogram bin are defined as,

$$\text{Pulls} = \frac{C_{\text{data}} - C_{\text{fit model}}}{C_{\text{error}}}, \tag{7.1}$$

where C_{data} and C_{fitmodel} are the number of candidates in data and as estimated by the fit model, C_{error} is the Poissonian error of the binned data. At ~3100 MeV/c^2, the deviations from 0 are caused by small mismodelling effects of the peak of the resolution model which has a visible impact due to the very narrow and large J/ψ resonance. Figure 7.2 shows the data and fitted model within the invariant dimuon mass range of 2800-3400 MeV/c^2. As is also visible, the effect of this mismodelling is minimal at the tail ends of the peak where the impact on the local contribution is largest. Therefore, this effect can be safely ignored.

Motivated by a structure visible at ~ 2000 MeV/c^2 in Run1 data set in Ref. [12], a similar observation can also be made with the Run1+Run2 data set in Figure 7.1. This structure could potentially hint at broad vector resonances states such as $\rho(1700)$, $\omega(1650)$ and $\phi(1680)$ which

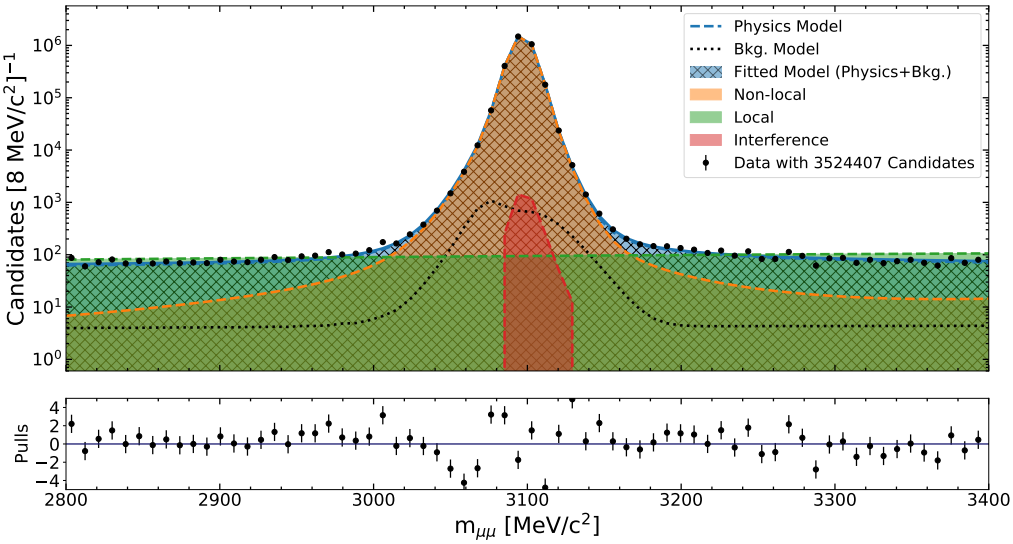

Figure 7.2: The fit to the dimuon invariant mass of $B^+ \to K^+\mu^+\mu^-$ decays using the fit model described in Chapter 6 and data collected by the LHCb experiment during Run1 and Run2. This figure shows the data and fitted model from Figure 7.1 and is displayed in the invariant dimuon mass range around the J/ψ resonance, 2800-3400 MeV/c^2, using a logarithmic scale for the y axis.

could interfere with the local contribution. The significance of the presence of such states will be studied as a systematic.

7.1 The Observed Degeneracy of 1P states

The fit model contains degeneracies, the solutions with the smallest difference in $\log\mathcal{L}$ values compared to the nominal result are from $\psi(4415)$ and J/ψ 1P states with $\Delta\log\mathcal{L} = 0.08$ and 0.18 respectively. The results from the fit to data at these alternative solutions for small $\psi(4415)$ and positive J/ψ phase are shown in Figure 7.3 and the values presented in Table 7.1. As can be noted from Table 7.1, the values of Wilson C_9 and C_{10} do not change between the degenerate solutions while some of the non-local parameters show notable deviations. A discussion on the degeneracy in $\psi(4415)$ is presented first followed by a discussion on the J/ψ degeneracy.

In order to study this degenerate behaviour the likelihood profiles of the parameters are determined. The likelihood profile of a parameter is obtained by systematically fixing the particular parameter to different values around the most likely value and calculating the $-\log\mathcal{L}$ at the best fit point of the remaining parameters of the model. Two parameters simultaneously profiled will yield a 2D likelihood surface.

From Figure 7.3 it is visible that the alternative solution for $\psi(4415)$ has a smaller magnitude.

Parameter		Nominal Result $\Delta\log\mathcal{L} = 0$	Smaller $\psi(4415)$ $\Delta\log\mathcal{L} = 0.08$	Positive J/ψ phase $\Delta\log\mathcal{L} = 0.18$
Wilson C_9		16.16 ± 0.42	16.16 ± 0.22	16.18 ± 0.29
Wilson C_{10}		-52.03 ± 0.26	-52.04 ± 0.27	-52.04 ± 0.44
b_0^+		0.448 ± 0.010	0.448 ± 0.010	0.448 ± 0.010
b_1^+		-1.12 ± 0.07	-1.12 ± 0.05	-1.13 ± 0.09
b_2^+		-1.00 ± 0.37	-0.99 ± 0.31	-1.00 ± 0.33
$\rho(770)$ phase	rad	-0.69 ± 0.38	-0.70 ± 0.38	-0.69 ± 0.39
$\omega(782)$ phase	rad	-0.06 ± 0.24	-0.06 ± 0.23	-0.05 ± 0.25
$\phi(1020)$ phase	rad	0.34 ± 0.21	0.34 ± 0.21	0.36 ± 0.21
J/ψ Mass	MeV/c^2	3096.668 ± 0.004	3096.668 ± 0.004	3096.668 ± 0.004
J/ψ phase	rad	-1.71 ± 0.08	-1.72 ± 0.08	1.74 ± 0.13
$\psi(2S)$ Mass	MeV/c^2	3685.88 ± 0.01	3685.88 ± 0.01	3685.88 ± 0.01
$\psi(2S)$ phase	rad	-2.81 ± 0.58	-2.84 ± 0.53	3.00 ± 0.74
$\psi(2S)$ BF		$(4.75 \pm 0.15) \times 10^{-6}$	$(4.75 \pm 0.01) \times 10^{-6}$	$(4.74 \pm 0.02) \times 10^{-6}$
$\psi(3770)$ phase	rad	-2.87 ± 0.33	-2.90 ± 0.28	2.98 ± 0.36
$\psi(3770)$ BF		$(2.23 \pm 0.51) \times 10^{-9}$	$(2.23 \pm 0.51) \times 10^{-9}$	$(2.25 \pm 0.60) \times 10^{-9}$
$\psi(4040)$ Mass	MeV/c^2	4039.00 ± 1.03	4039.00 ± 1.00	4038.99 ± 1.00
$\psi(4040)$ Width	MeV/c^2	80.44 ± 9.54	80.48 ± 9.76	80.58 ± 9.65
$\psi(4040)$ phase	rad	-3.02 ± 0.24	-3.09 ± 0.24	2.89 ± 0.28
$\psi(4040)$ BF		$(2.30 \pm 0.99) \times 10^{-9}$	$(2.28 \pm 0.85) \times 10^{-9}$	$(2.42 \pm 1.12) \times 10^{-9}$
$\psi(4160)$ Mass	MeV/c^2	4196.28 ± 6.73	4196.17 ± 5.59	4195.77 ± 5.32
$\psi(4160)$ Width	MeV/c^2	71.68 ± 9.81	71.71 ± 9.91	71.57 ± 10.21
$\psi(4160)$ phase	rad	-2.14 ± 0.26	-2.29 ± 0.22	-2.48 ± 0.27
$\psi(4160)$ BF		$(3.97 \pm 0.87) \times 10^{-9}$	$(3.76 \pm 0.85) \times 10^{-9}$	$(4.01 \pm 1.26) \times 10^{-9}$
$\psi(4415)$ Mass	MeV/c^2	4420.12 ± 3.97	4420.10 ± 3.98	4420.20 ± 3.98
$\psi(4415)$ Width	MeV/c^2	78.47 ± 4.90	78.34 ± 14.88	79.64 ± 16.17
$\psi(4415)$ phase	rad	1.71 ± 0.23	2.17 ± 0.29	1.48 ± 0.39
$\psi(4415)$ BF		$(9.60 \pm 3.96) \times 10^{-9}$	$(2.56 \pm 1.42) \times 10^{-9}$	$(9.65 \pm 7.43) \times 10^{-9}$
2P States phase	rad	-3.05 ± 0.13	-3.05 ± 0.15	2.96 ± 0.30
DD BF		$(2.34 \pm 1.37) \times 10^{-9}$	$(2.18 \pm 0.67) \times 10^{-9}$	$(2.37 \pm 0.96) \times 10^{-9}$
D^*D^* BF		$(1.48 \pm 0.87) \times 10^{-9}$	$(1.38 \pm 0.43) \times 10^{-9}$	$(1.50 \pm 0.60) \times 10^{-9}$
DD^* BF		$(8.92 \pm 5.22) \times 10^{-9}$	$(8.29 \pm 2.57) \times 10^{-9}$	$(9.04 \pm 3.64) \times 10^{-9}$

Table 7.1: The measured parameters from the fits to the dimuon invariant mass of $B^+ \to K^+\mu^+\mu^-$ decays using the fit model described in Chapter 6 and data collected by the LHCb experiment during Run1 and Run2. The nominal results and the two other degenerate solutions discussed in Section 7.1 are shown with the values of Wilson C_9 and C_{10} having a random offset from the measured values. The quoted errors are Hessian error intervals estimated from the fit. The Branching Fraction (BF) have been calculated from the measured magnitude using Eq. 2.124. The parameters with notable differences w.r.t. the nominal result are highlighted in purple.

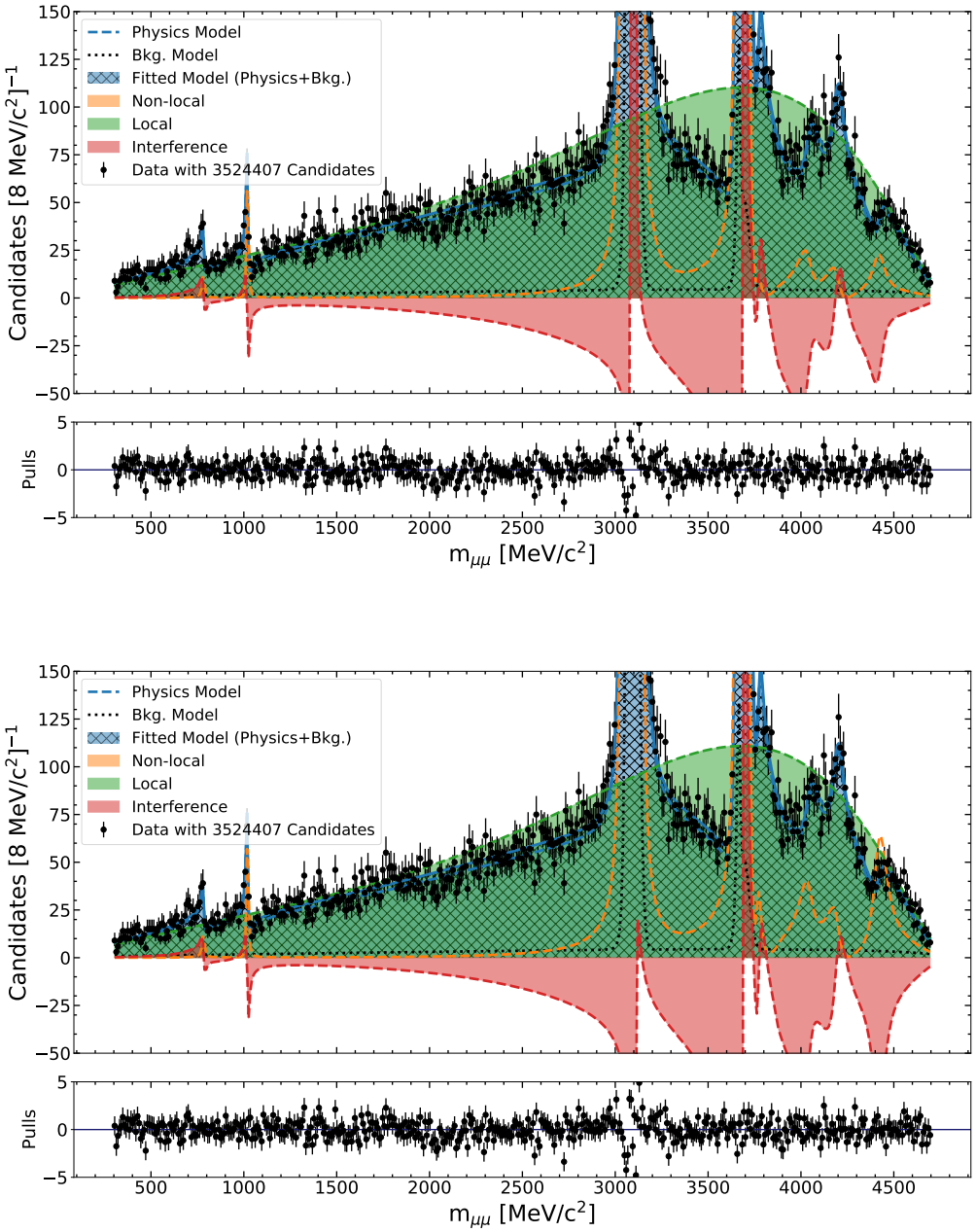

Figure 7.3: The degenerate solutions from fits to the dimuon invariant mass of $B^+ \to K^+\mu^+\mu^-$ decays using the fit model described in Chapter 6 and data collected by the LHCb experiment during Run1 and Run2. (Top) The degenerate solution of $\psi(4415)$ with a smaller magnitude and a $\Delta\log\mathcal{L} = 0.08$ compared with the nominal result. (Bottom) The degenerate solution of J/ψ with a positive phase and a $\Delta\log\mathcal{L} = 0.18$ compared with the nominal result.

Region	Parameter		Nominal Result
Region-2	σ_{dcb}	MeV/c^2	5.78 ± 0.04
	σ_{gauss}	MeV/c^2	7.71 ± 0.10
	α_L		1.13 ± 0.02
	f_{dcb}		0.83 ± 0.02
Region-3	σ_{dcb}	MeV/c^2	4.47 ± 0.21
	σ_{gauss}	MeV/c^2	6.07 ± 0.11
	α_L		0.91 ± 0.08
	f_{dcb}		0.59 ± 0.08

Table 7.2: The measured resolution parameters from invariant dimuon mass fits to data.

The 2D likelihood surface determined from a likelihood profile of $\psi(4415)$ magnitude ($\eta_{\psi(4415)}$) and phase ($\delta_{\psi(4415)}$) is shown in Figure 7.4. On this 2D surface, the two degenerate solutions are seen to lie within the 68.3% contour. While the difference in the phases between the two solution appears to be small, there is a clear difference in the magnitudes. Figure 7.4 also shows the 1D likelihood profiles of the $\eta_{\psi(4415)}$ and $\delta_{\psi(4415)}$. This degeneracy arises from a symmetry in the structure of a Breit-Wigner amplitude and its interference with the local and 2P states. Presently no precise external input on the $\psi(4415)$ state is available and it is hoped that more data would help break this degeneracy.

The smallness of the bump which splits the two solutions and the closeness of their $\log\mathcal{L}$ values is a cause of instability as the minimizer could descend into either minimum. This feature is of importance later in order to understand the correlations with the other parameters. A similar effect is also present for the other high $m_{\mu\mu}$ resonances: $\psi(3770)$, $\psi(4040)$ and $\psi(4160)$ but the difference in magnitudes and $\log\mathcal{L}$ values are large so that the instability caused is small.

The other observed degeneracy is from the phase of the J/ψ resonance. This degeneracy arises from a symmetry in the structure of two Breit-Wigner amplitudes, namely the J/ψ and $\psi(2S)$ and is irrespective of the underlying contribution. To demonstrate this, the left plot in Figure 7.5 shows a 1D likelihood profile of J/ψ phase ($\delta_{J/\psi}$) using a very simple toy data set generated using only the J/ψ, $\psi(2S)$ and a flat phasespace. In this figure, the two solutions for $\delta_{J/\psi}$ are visible and a similar structure is also seen when the $\psi(2S)$ phase ($\delta_{\psi(2S)}$) is profiled. The underlying contribution could potentially help break this degeneracy given enough data.

This degeneracy is a known effect and was also seen in the previous iteration of this measurement where a 4-fold degeneracy was observed with the $\delta_{J/\psi}$ and $\delta_{\psi(2S)}$ either positive or negative in the parameter range $[-\pi, \pi]$ rad. The right plot in figure 7.5 shows the 1D likelihood profiles of $\delta_{J/\psi}$ in the positive and negative regions of $\delta_{\psi(2S)}$ distinctly showing the four minima. While the values of C_9 and C_{10} did not vary much between the four solutions, the non-local parameters were dependent on it. As a result of this observation four set of results

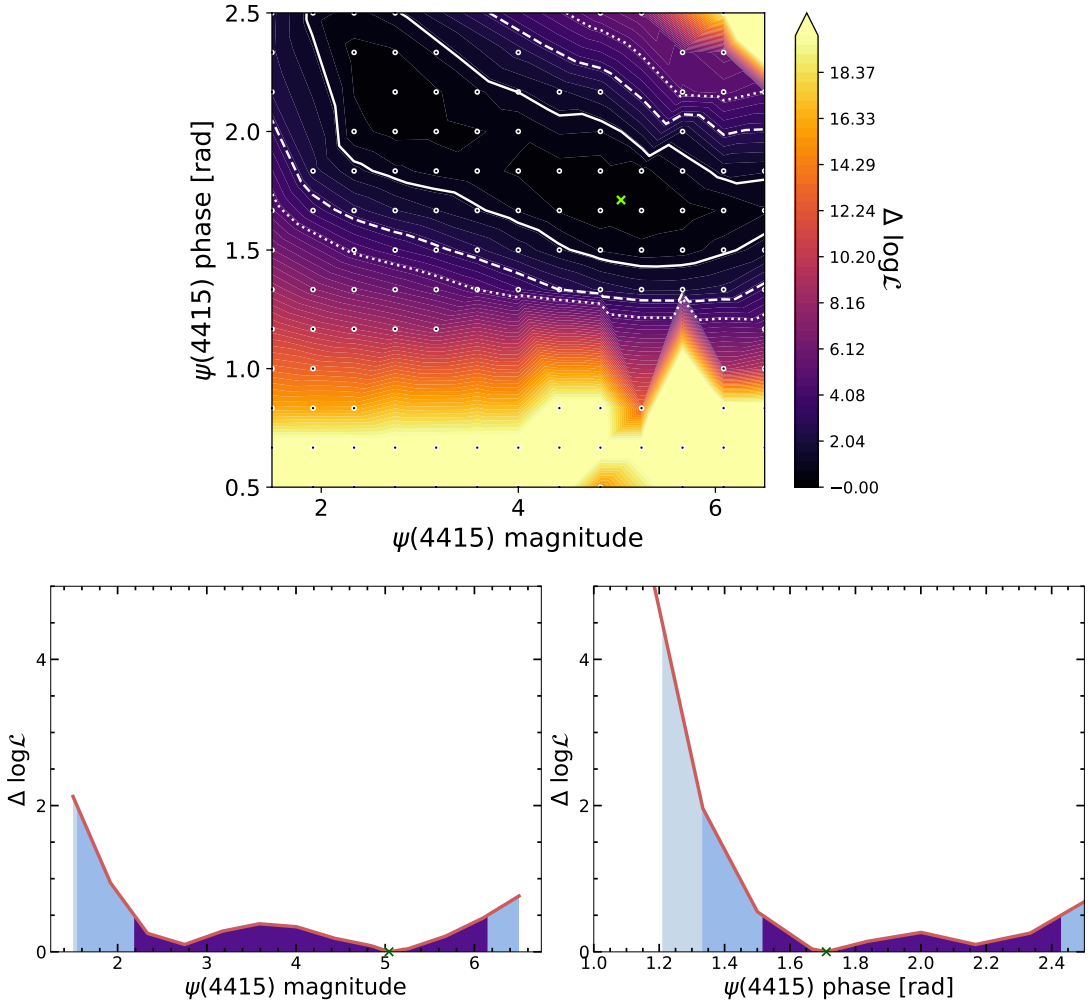

Figure 7.4: The 2D (top) and 1D (bottom) likelihood profile of the fit parameters $\psi(4415)$ magnitude and phase. The nominal result is marked by the green cross. In the 2D likelihood profile, the profiled points are marked by the grid of dots and the colour bar linearly interpolates between these points, the solid, dashed and dotted white contours mark the 68.3%, 95.5% and 99.7% statistical intervals respectively. From the 1D likelihood profiles, the measured values are $\eta_{\psi(4415)} = 5.04^{+1.1}_{-2.9}$ and $\delta_{\psi(4415)} = 1.71^{+0.7}_{-0.2}$. The 68.3%, 95.5% and 99.7% statistical intervals are highlighted by the progressively lighter blue shaded regions.

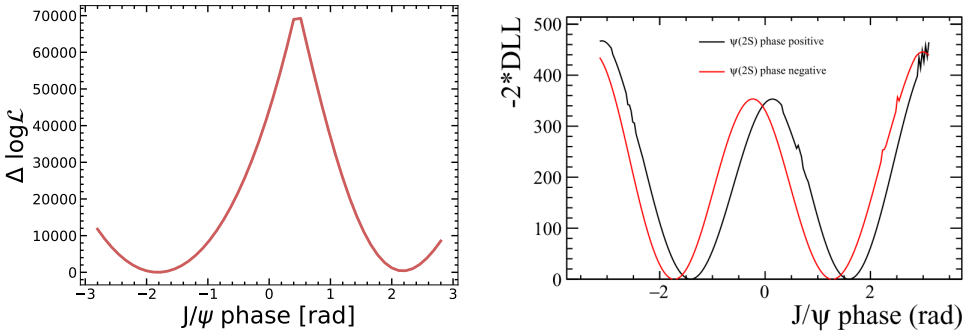

Figure 7.5: (Left) The likelihood profile of J/ψ phase using a toy data set generated with a J/ψ and $\psi(2S)$ component over a flat phasespace in place of the local contribution. No other intermediate state is included in the description of the PDF. (Right) The 1D likelihood profiles of J/ψ phase with $\psi(2S)$ phase positive or negative from Ref. [12]. The four minima correspond to the four published degenerate results.

were published for the non-local contributions.

In this measurement, the inclusion of a broad, non-resonant 2P amplitude has an interesting effect. The 4-fold degeneracy in $\delta_{J/\psi}$ and $\delta_{\psi(2S)}$ has reduced to a 2-fold one as shown in Figure 7.6. The 2D likelihood profile of $\delta_{J/\psi}$ and $\delta_{\psi(2S)}$ in the parameter range of $[-\pi,\pi]$ rad clearly shows this, the results from Ref. [12] are marked by the solid red points with error bars. Note that the phase values are cyclic, meaning the parameter space above π rad is equivalent to the parameter space above $-\pi$ rad. Therefore, for a negative or positive J/ψ phase, the two contours at $\psi(2S)$ phase around $-\pi$ rad and π rad are in fact two halves of the same contour. The introduction of a mechanism to account for the 2P states has not only revealed the presence of a large 2P state contribution, the precision of the $\delta_{\psi(2S)}$ has deteriorated enough that the two solutions cannot be separated. The significance of the presence of the 2P states is presented in Section 7.2.

The difference in the $\log\mathcal{L}$ between the two minima is $\Delta\log\mathcal{L} = 0.18$ with the solution at negative $\delta_{J/\psi}$ treated as the nominal result. The 1D likelihood profiles of $\delta_{J/\psi}$ and $\delta_{\psi(2S)}$ phase at the nominal result is shown in Figure 7.6. The very broad likelihood surface leads to a large statistical uncertainty on this parameter. The effect of the 2P states on the $\psi(2S)$ is larger when compared to the J/ψ since the mass of DD, DD^* and D^*D^* is closer to $m_{\psi(2S)}$.

7.2 The Significance of the 2P states

Before attempting the fit to data the contribution from the 2P states was expected to be negligible and was the primary motivation to exclude these contributions in the fit model of the previous iteration. In Ref. [50] it is estimated that 2P state magnitude has an upper limit of $\eta_{2P} \leq 0.2$. This value comes from saturating the perturbative charm loop calculation

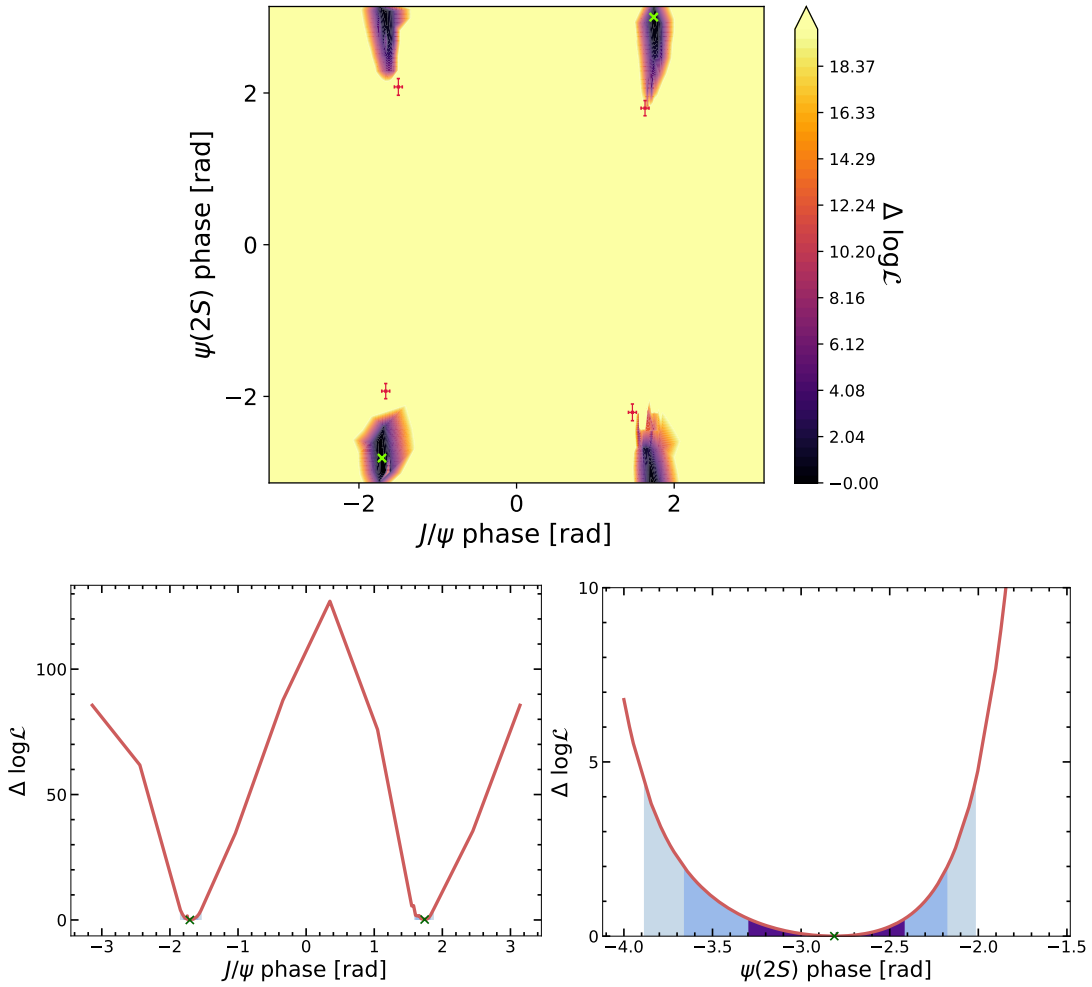

Figure 7.6: The 2D (top) and 1D (bottom) likelihood profiles of the fit parameters J/ψ and $\psi(2S)$ phase. The nominal result and positive J/ψ phase result are marked by a green cross. In the 2D likelihood profile, the solid red points with error bars mark the results from Ref. [12]. From the 1D likelihood profiles, the measured values of $\delta_{\psi(2S)}$ = -2.81$^{+0.4}_{-0.5}$. The 68.3%, 95.5% and 99.7% statistical intervals are highlighted by the progressively lighter blue shaded regions. Note here that the phase is cyclic *i.e.* values below $-\pi$ are equivalent to values below π.

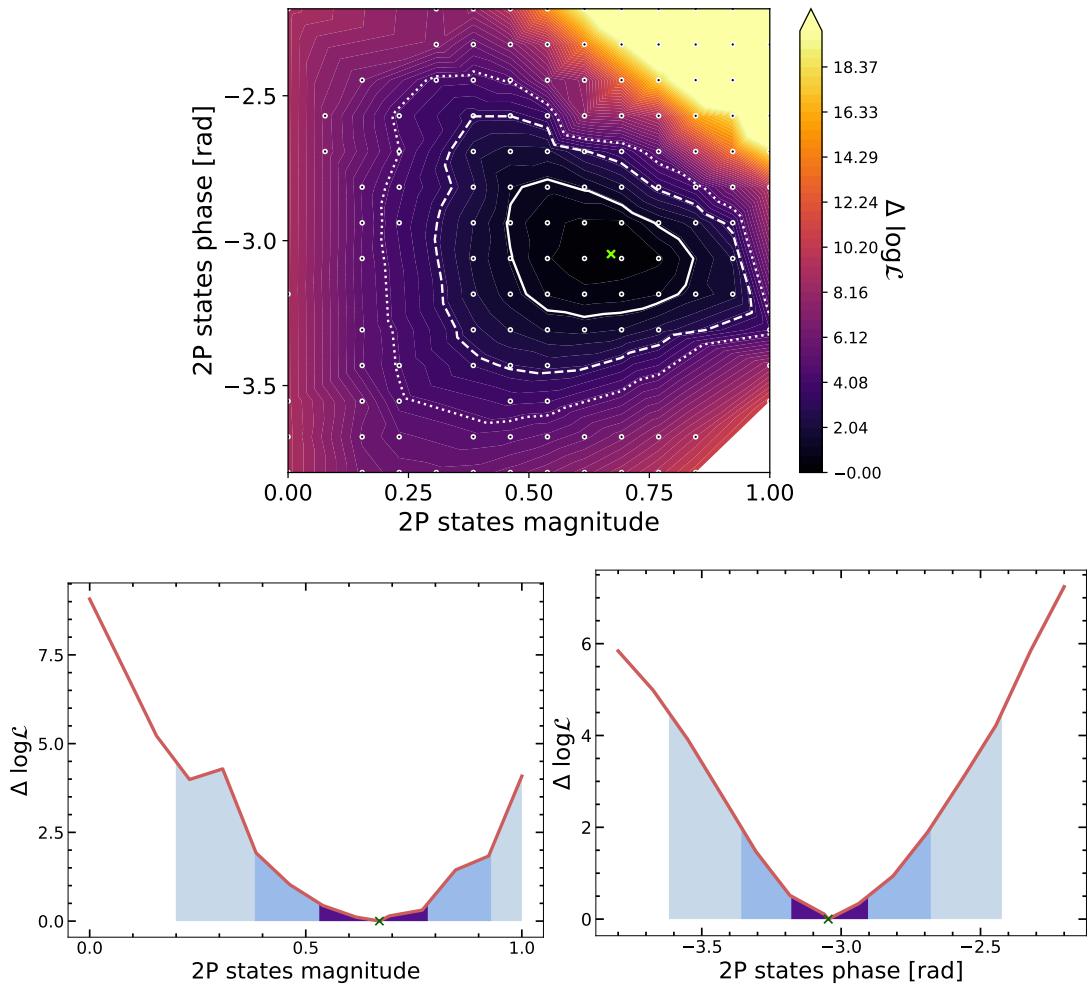

Figure 7.7: The 2D (top) and 1D (bottom) likelihood profiles of the fit parameters 2P states magnitude and phase. The nominal result is marked by the green cross. In the 2D likelihood profile, the profiled points are marked by the grid of dots and the colour bar linearly interpolates between these points, the solid, dashed and dotted white contours mark the 68.3%, 95.5% and 99.7% statistical intervals respectively. From the 1D likelihood profiles, the measured value of 2P magnitude, $\eta_{2P} = 0.67^{+0.11}_{-0.14}$ and phase δ_{2P} = -3.05 ± 0.13 rad . The 68.3%, 95.5% and 99.7% statistical intervals are highlighted by the progressively lighter blue shaded regions.

with only the DD^* S-wave contribution, giving a magnitude $\eta_{\bar{D}} = 2(C_2 + C_1/3) = 0.2 \pm 0.2$. As such, according to this calculation, the individual exclusive meson contribution should be significantly smaller that the inclusive quark contribution.

In Figure 7.7 the 2D and 1D likelihood surfaces from the profile of the η_{2P} and δ_{2P} are shown. From these likelihood profiles, *each* of the 2P states is measured to have a magnitude[1] of $\eta_{2P} = 0.67^{+0.11}_{-0.14}$. The preference of data for a large destructive 2P state is very clear. The likelihood difference between the null hypothesis of $\eta_{2P} = 0$ and the measured value is $\Delta \log \mathcal{L}$

[1] A reminder that in this measurement it is assumed that all 2P states have an equal contribution

= 9.07. Using Wilk's theorem [135] the significance of the 2P states is 3.68σ with a p-value of 0.0001. The measured value of all the fit parameters at the null hypothesis is given in Table 7.3.

The measurement of a large η_{2P} is certainly intriguing and offers more information to understand the difficulty to calculate non-local effects. The large destructive interference caused by it has the potential to explain the anomalies seen in B meson decays which were discussed in Section 2.6.

7.3 Measurement of the Wilson coefficients

The measurement of the Wilson coefficients after explicitly accounting for the non-local contributions will reveal any NP effect present. While it is currently not possible to study the presence of such an effect, the correlation of Wilson C_9 and C_{10} can be studied. In Figure 7.8 the 2D and 1D profiles of C_9 and C_{10} are shown. The blank region in the top left corner is due to the failed fits caused by the instability discussed in Section 7.1. From the 1D likelihood profiles, the 1-σ statistical precision can be estimated and are C_9 = 16.16 ± 0.3 and C_{10} = -52.03 ± 0.3. To summarise, the parameters with the 68.3% error intervals estimated from 1D likelihood profiles are presented in Table 7.4.

The model to include the 2P states has yet another interesting and important consequence. For comparison, the 2D likelihood profile from the previous iteration is shown in Figure 7.9. In the previous iteration, the likelihood surface of C_9 and C_{10} formed a semicircle which was expected from the approximate proportionality of $\frac{d\Gamma}{dq^2} \propto |C_9|^2 + |C_{10}|^2$. When the 2P states are excluded a similar behaviour is also seen with the current data set. With the 2P states however, the correlation between C_9 and C_{10} has reduced to the disc-like structure seen in Figure 7.8. The consequences of this is profound as it questions our understanding of the extent of the contribution of non-local states and their interaction with the local contribution. Comparing the nominal result and the result when the 2P states are excluded (η_{2P} =0), the value of C_9 changes by ~0.1 units while C_{10} changes by ~2 units (SM values are ~ ±4). This observation is also of great importance since it implies that the 2P amplitudes, which interfere with C_9, have a much larger impact on C_{10}.

7.4 A Toy Study of the Nominal Result

To understand the correlations between the parameters and investigate the consequences of the observed degeneracies, a toy study is conducted. For this toy study, 1000 toy data sets are generated using a PDF defined by the parameters from the nominal fit to data. Figure 7.10 shows the distributions of the pull statistic and fitted values for C_9, C_{10} and $\delta_{\psi(2S)}$. From these figures, the fitted values of C_9 and C_{10} certainly contain a bias. The double peak structure in the distributions of C_9 is peculiar, however, a more interesting effect is seen in the distributions of $\delta_{\psi(2S)}$. The fitted values have a narrow distribution around the nominal result and a wide tail

Parameter		No 2P result $\Delta\log\mathcal{L} = 9.07$
Wilson C_9		16.05 ± 0.46
Wilson C_{10}		-50.21 ± 0.74
b_0^+		0.466 ± 0.010
b_1^+		-0.87 ± 0.04
b_2^+		-0.34 ± 0.27
$\rho(770)$ phase	rad	-0.49 ± 0.35
$\omega(782)$ phase	rad	-0.07 ± 0.24
$\phi(1020)$ phase	rad	0.34 ± 0.20
J/ψ Mass	MeV/c^2	3096.668 ± 0.004
J/ψ phase	rad	-1.55 ± 0.02
$\psi(2S)$ Mass	MeV/c^2	3685.88 ± 0.01
$\psi(2S)$ phase	rad	2.49 ± 0.14
$\psi(2S)$ BF		$(4.77 \pm 0.01) \times 10^{-6}$
$\psi(3770)$ phase	rad	-3.11 ± 0.12
$\psi(3770)$ BF		$(1.28 \pm 0.35) \times 10^{-9}$
$\psi(4040)$ Mass	MeV/c^2	4038.86 ± 0.99
$\psi(4040)$ Width	MeV/c^2	86.45 ± 9.20
$\psi(4040)$ phase	rad	-3.02 ± 0.13
$\psi(4040)$ BF		$(1.93 \pm 0.60) \times 10^{-9}$
$\psi(4160)$ Mass	MeV/c^2	4190.01 ± 5.31
$\psi(4160)$ Width	MeV/c^2	71.32 ± 9.49
$\psi(4160)$ phase	rad	-2.08 ± 0.19
$\psi(4160)$ BF		$(2.42 \pm 0.60) \times 10^{-9}$
$\psi(4415)$ Mass	MeV/c^2	4420.21 ± 3.89
$\psi(4415)$ Width	MeV/c^2	92.50 ± 14.61
$\psi(4415)$ phase	rad	1.84 ± 0.06
$\psi(4415)$ BF		$(3.64 \pm 0.70) \times 10^{-8}$
2P States phase	rad	-
DD BF		-
D^*D^* BF		-
DD^* BF		-

Table 7.3: The measured parameters from the fits to the dimuon invariant mass of $B^+ \to K^+\mu^+\mu^-$ decays using the fit model described in Chapter 6 with no 2P states and data collected by the LHCb experiment during Run1 and Run2. The nominal results and the two other degenerate solutions discussed in Section 7.1 are shown with the values of Wilson C_9 and C_{10} having a random offset from the measured values. The quoted errors are Hessian error intervals estimated from the fit. The Branching Fraction (BF) have been calculated from the measured magnitude using Eq. 2.124.

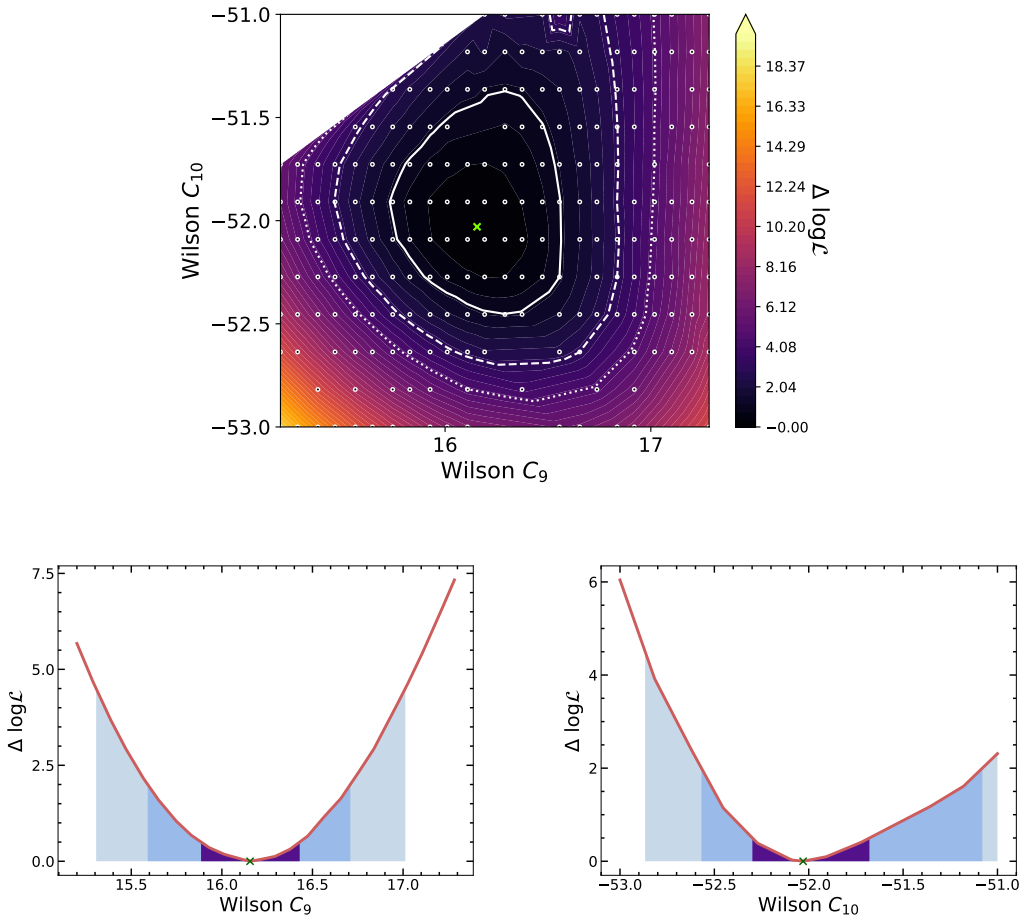

Figure 7.8: The 2D (top) and 1D (bottom) likelihood profiles of the fit parameters Wilson C_9 and C_{10} . The nominal result is marked by the green cross. In the 2D likelihood profiles, the profiled points are marked by the grid of dots and the colour bar linearly interpolates between these points, the solid, dashed and dotted white contours mark the 68.3%, 95.5% and 99.7% statistical intervals respectively. From the 1D likelihood profiles, the measured value of Wilson C_9 = 16.16± 0.27 and C_{10} = -52.03 $^{+0.35}_{-0.27}$. The 68.3%, 95.5% and 99.7% statistical intervals are highlighted by the progressively lighter blue shaded regions.

Parameter	Measured Value
Wilson C_9	16.16 ± 0.27
Wilson C_{10}	-52.03 $^{+0.35}_{-0.27}$
$\psi(2S)$ phase	-2.8$^{+0.4}_{-0.5}$ rad
$\psi(4415)$ magnitude	5.0$^{+1.1}_{-2.9}$
$\psi(4415)$ phase	1.7$^{+0.7}_{-0.2}$ rad
2P magnitude	0.67 ± 0.20
2P phase	-3.05 ± 0.13 rad

Table 7.4: The 68.3% error interval estimated from 1D likelihood profiles.

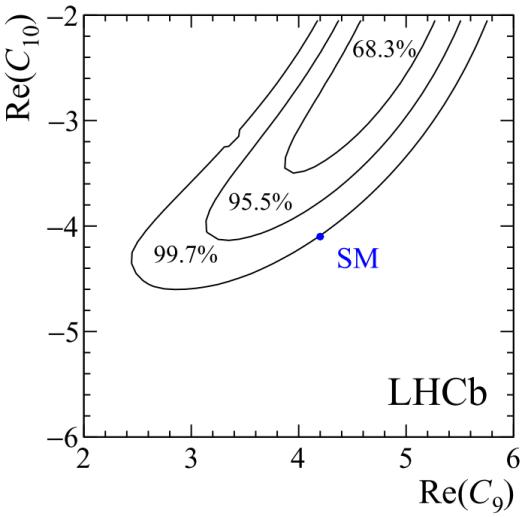

Figure 7.9: The 2D likelihood profile of the fit parameters Wilson C_9 and C_{10} from Ref. [12]. The SM value is marked by the blue dot, the 68.3%, 95.5% and 99.7% statistical intervals are marked by the contours.

on both sides. Another observation in the distributions of Wilson C_9 is the cluster of data points very close to the nominal result away from the peak of the distribution. This could indicate at the presence of two or more multidimensional likelihood surfaces with their minima close to each other. The reader may recall the discussion on the degeneracy of the 1P states in Section 7.1 which investigated exactly such surfaces.

In order to understand this phenomenon better, the 2D distributions of the $\delta_{\psi(2S)}$ with a selection of the other fit parameters are shown in Figure 7.11. From these distributions, the tail ends of the $\delta_{\psi(2S)}$ show both correlation and anti-corelation with a large cluster of points towards the center. This effect is most prominent in the 2D distribution of the $\delta_{\psi(2S)}$ and $\eta_{\psi(4415)}$ shown in Figure 7.12. Here the correlation and anti-correlation is distinctly visible from the "X"-pattern formed. This pattern is precisely the consequence of the two very close minima

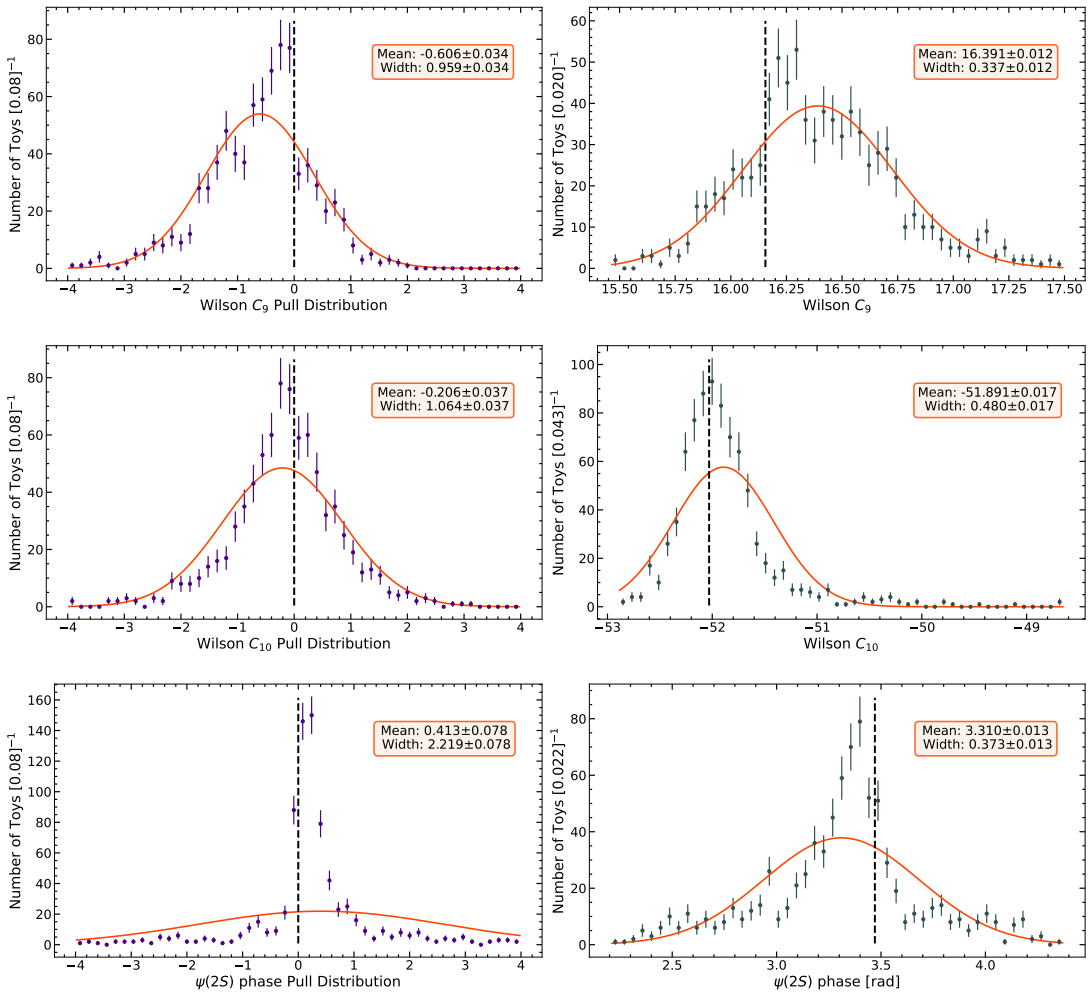

Figure 7.10: The distribution of the pull statistic (left) and fitted values (right) of Wilson C_9 (top), C_{10} (middle) and $\delta_{\psi(2S)}$ (bottom) from fit to 1000 toy data set generated using the PDF defined by the parameter values from the nominal result.

seen in the 1D likelihood profile of the $\eta_{\psi(4415)}$ in Figure 7.4. A similar correlation between $\eta_{\psi(4415)}$ and the other parameters is not seen. The wide 1D likelihood profile of the $\delta_{\psi(2S)}$ means that the different behaviours of both these minima are clearly visible.

The best fit points of the set of toys generated at the nominal fit result sample the other minima within the 1-σ interval of the nominal fit parameters. It is also statistically possible that in different toy data sets, the true minimum is not the measured best fit points. To be able to always find the global minimum would require a 1D likelihood profile of $\eta_{\psi(4415)}$ to each of the toy data sets, which is computationally unfeasible.

A particularly interesting observation is that the $\psi(2S)$ and $\psi(4415)$ amplitudes are far apart in the $m_{\mu\mu}$ spectrum. The correlations arise from the broad 2P states. Studying the Wilson C_9 distribution shown in Figure 7.10, the consequence of this degeneracy can be visualized.

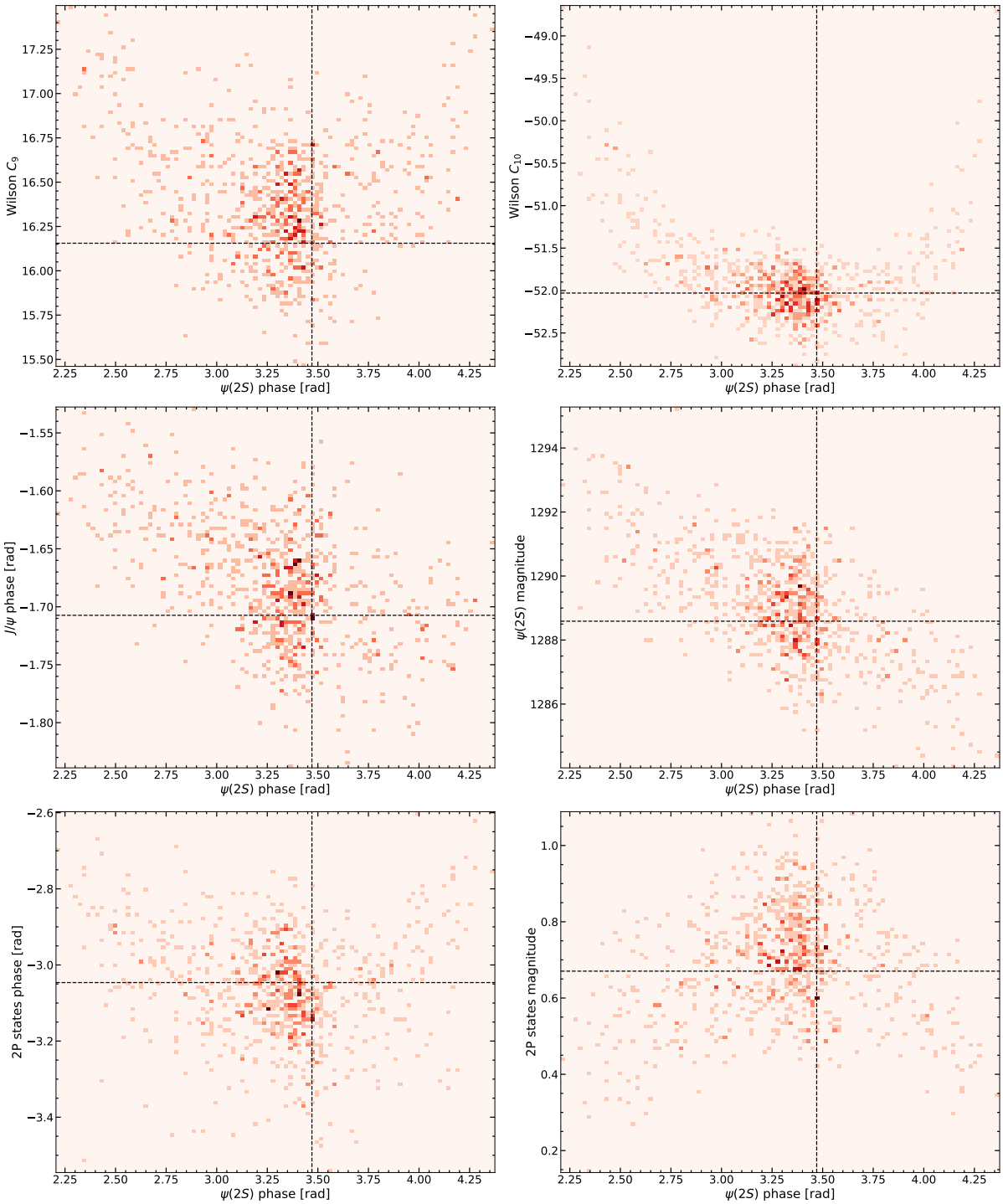

Figure 7.11: The 2D distributions of the fitted values of $\delta_{\psi(2S)}$ and another parameter from fits to 1000 toy data sets generated using a PDF defined by the nominal result. The nominal result is marked by the dashed black line.

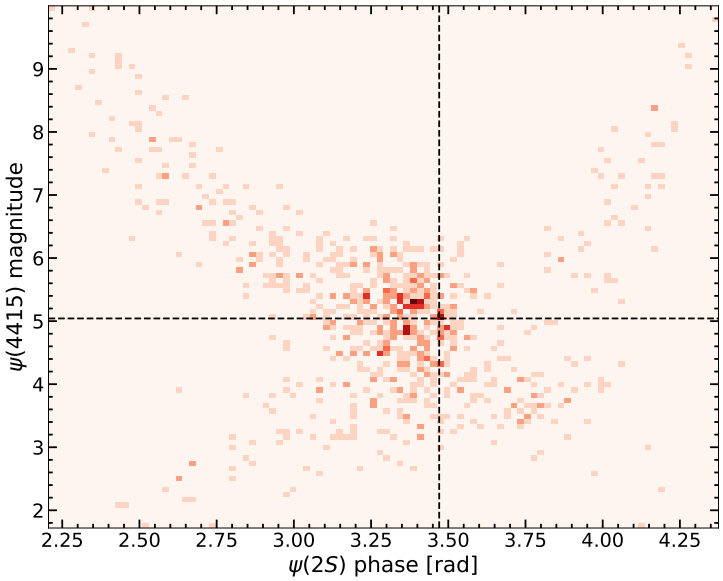

Figure 7.12: The 2D distributions of the fitted values of $\delta_{\psi(2S)}$ and $\eta_{\psi(4415)}$ from fits to 1000 toy data sets generated using a PDF defined by the nominal result. The nominal result is marked by the dashed black line.

The effect of a simple selection of $\eta_{\psi(4415)} > 4.5$ on Wilson C_9 can be seen in Figure 7.13. This selection explicitly picks toys with $\eta_{\psi(4415)}$ closer to the nominal result. By removing events with the lower $\eta_{\psi(4415)}$, the cluster closer to the nominal value of C_9 remains while the wider peak which is the main cause of the bias has reduced. Further investigation is needed in order to ascertain the full effects of this degeneracy on the nominal result values of Wilson C_9 and C_{10}. A method albeit computationally intensive would be to perform a 3D likelihood profile of C_9, C_{10} and $\eta_{\psi(4415)}$ which would map out the full likelihood surface encoding all the important correlations.

7.4.1 Goodness of Fit

The goodness of the fit to data can be ascertained using the fit to toy data sets which have been generated using the nominal result. In Figure 7.14, the distribution of likelihood values calculated from the toy data and fitted model at their best fit points is shown. The value of the likelihood from data using a fit model at the nominal result is also marked. The small mismodelling effects of the peak of the J/ψ resonance in data has a large impact on likelihood values due to the large number of candidates. Including the core of the resonance, the likelihood values between real data and toy data sets is not expected to agree. Therefore, in order to calculate the likelihood values shown in this figure, real and toy data within the invariant dimuon mass range 3050-3110 MeV/c^2 have been removed. This choice is justified since the impact of data within this window on Wilson coefficients C_9 and C_{10} is negligible. It is clear

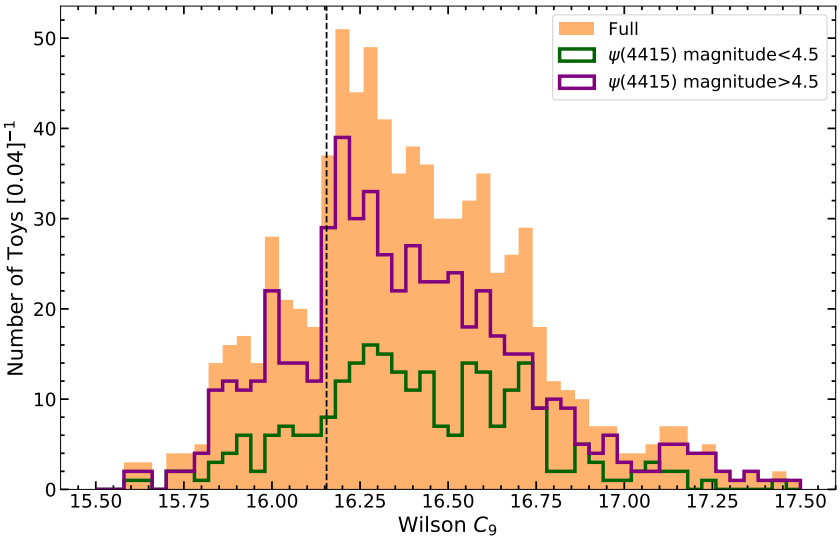

Figure 7.13: The distribution of the fitted values of Wilson C_9 to 1000 toy data sets generated using the PDF defined by the parameter values from the nominal result (black dashed). The effect of placing selections on $\eta_{\psi(4415)}$ is shown.

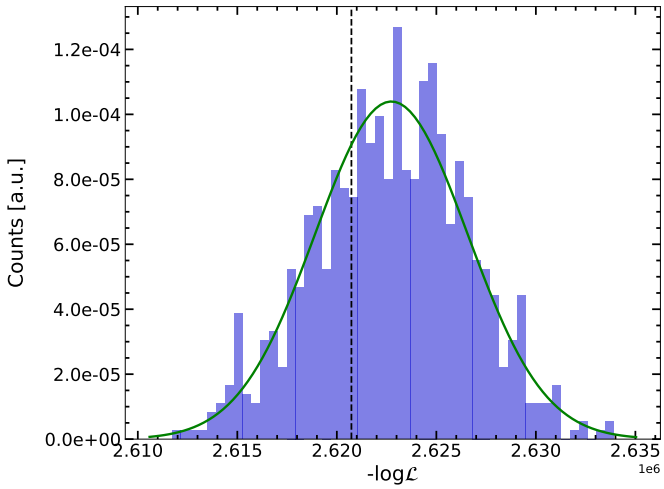

Figure 7.14: The distribution of likelihood values from toy data sets generated using the nominal result. The likelihood value of real data at the nominal result is shown by the black dashed line. For both real and toy data the likelihood values have been calculated by ignoring the data at the core of the J/ψ resonance in the range 3050-3110 MeV/c^2.

from this figure that the fit to data represents a good fit.

7.5 Systematic Uncertainties

Some of the largest sources of systematic uncertainties have been encoded directly in the likelihood. These are highlighted first, followed by a discussion on other potential sources of

Coefficient	Prior	Posterior
b_0^+	0.466 ± 0.014	0.448 ± 0.011
b_1^+	-0.89 ± 0.13	-1.12 ± 0.07
b_2^+	-0.21 ± 0.55	-1.00 ± 0.32

Table 7.5: The z-expansion coefficients for the form factor $f_+(q^2)$ with prior values from Ref. [44] and posterior values from the nominal fit to data.

systematic uncertainties.

7.5.1 Systematic effects encoded in the likelihood

One large source of systematic uncertainty is the prior knowledge of the form factor coefficients. The values for the form factor coefficients encode the local hadronic effects and have been extracted from Ref. [44] where LatticeQCD techniques have been used to estimate them. The shape of the non-resonant components are sensitive to the values of the form factor z-expansion coefficients (Section 2.3.1) and have a direct impact on the differential decay rate. The form factor with the largest contribution is $f_+(q^2)$ and its coefficients are treated as fit parameters with Gaussian constraints based on their uncertainties and correlations from Ref. [44]. Through this method, their systematic uncertainty is accounted for within the statistical uncertainty. The values of the prior and posterior form factor coefficients are shown in Table 7.5. The uncertainties associated to the other form factors ($f_0(q^2)$, $f_T(q^2)$) are negligible and therefore not included since their impact on the branching fraction is small.

The phases of the resonant contributions are particularly sensitive to the tail of the resolution model and the branching fraction calculations are sensitive to the width of the resolution model. In regions-2,3 due to the presence of the large narrow resonances (J/ψ and $\psi(2S)$), mismodelling of the resolution model will have a significant impact on the measured values. Therefore, the resolution model is validated using simulation (Section 5.3.3) with the tail and width parameters determined from the fit to data and thereby encoded in the likelihood.

The impact of the resolution in region-1 is small and the model estimated from simulation is used. Any residual uncertainty arising from the mismodelling of the resolution from simulation is assumed to be negligible.

The prior knowledge of the branching fraction of $\psi(3770)$ as well as the mass and width of $\psi(4040)$, $\psi(4160)$ and $\psi(4415)$ is not precise. For this reason, these parameters are treated as fit parameters with Gaussian constraints based on their prior uncertainties. Since the prior knowledge of the branching fractions of $\psi(4040)$, $\psi(4160)$ and $\psi(4415)$ is very imprecise, they are treated as fit parameters with no constraints.

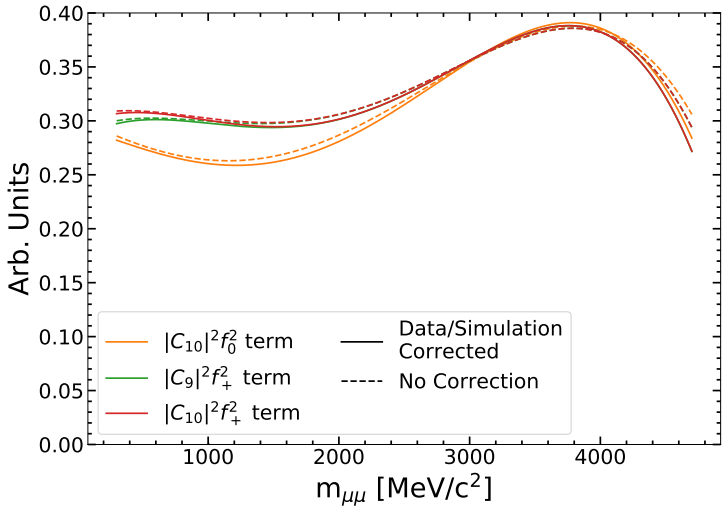

Figure 7.15: The evaluation of the efficiency of the detector from simulation with and without the application of corrections for data and simulation differences.

7.5.2 Other sources of systematic uncertainty

A potential source of systematic uncertainty arise from the modelling of the efficiency from the simulation. The efficiency affects the non-resonant components, similar to the form factors albeit to a smaller extent. Any residual mismodelling of the simulation could have a systematic effect on the efficiency calculation. The differences between data and simulation are corrected for through the method discussed in Sec. 5.1. The efficiency evaluation with and without this correction is shown in Figure 7.15. When the fit to data is repeated using the uncorrected efficiency, the value of Wilson C_9 and C_{10} changes by 0.05 and 0.03 units and the 2P magnitude and phase changes by 0.04 units and 0.1 rad.

The parameterisation of the combinatorial background model is estimated in the UMSB and the parameters linearly extrapolated to the signal region (Section 6.2.3). In order to estimate the uncertainties associated with this extrapolation, the UMSB data is bootstrapped and an ensemble of background models are generated. Repeating the fit to data using the alternative background models the associated systematic uncertainty can be assessed. The distribution of values estimated from the alternative fits to data of Wilson coefficients C_9 and C_{10} are shown in Figure 7.16. From this, the uncertainty arising from the background model extrapolation is estimated to be ~0.02 and is added to the result from the nominal fit.

A source of systematic uncertainty to Wilson coefficient C_9 is the estimated value of the charm loop effects at the subtraction point $q^2 = 0$ (Section 2.3.2). The once subtracted dispersion relation is used to model the non-local effects which requires the a priori knowledge of $Y_{c\bar{c}}^0$. For this, the estimate from Ref. [36] of $Y_{c\bar{c}}^0 \approx -0.10 \pm 0.05$ is used. The uncertainty associated with this number is taken as an external systematic uncertainty on Wilson C_9.

The impact of treating a symmetric tail parameterisation ($\alpha_R \equiv -\alpha_L$, Section 5.3.3) can

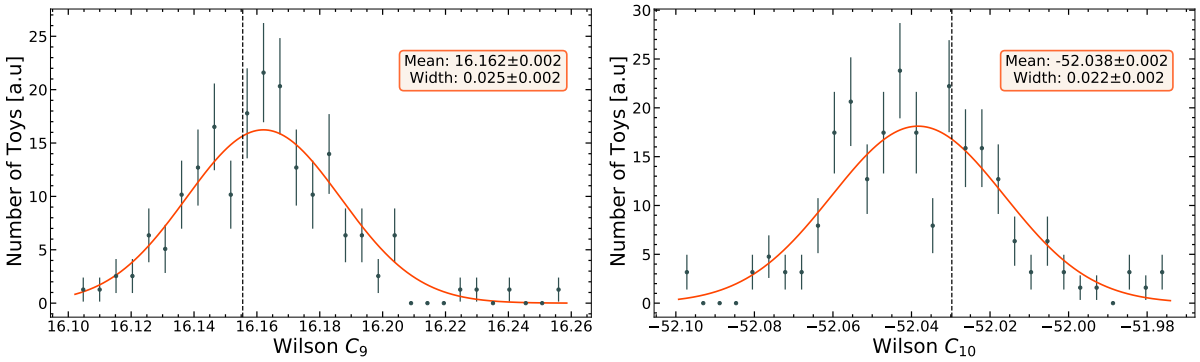

Figure 7.16: The distribution of Wilson coefficients C_9 and C_{10} from fits to data using alternative background model with parameters estimated and extrapolated from bootstrapped UMSB data.

Region	Parameter	Symmetric Model (Nominal)	Asymmetric Model
Region-2	α_L	1.13 ± 0.02	1.13 ± 0.01
	α_R	-1.13 ± 0.02	-1.12 ± 0.01
Region-3	α_L	0.90 ± 0.08	0.91 ± 0.05
	α_R	-0.90 ± 0.08	-0.90 ± 0.05

Table 7.6: The resolution tail parameter alpha estimated from the fit to data in region-2,3 using a symmetric parameterisation (α_R and α_L) and an asymmetric parameterisation where both α_R and α_L as independent fit parameter.

be studied by repeating the nominal fit by explicitly including α_R and α_L as independent fit parameter. The resulting values of these parameters in region-2,3 is shown alongside the nominal result in Table 7.6. The impact of this small difference is negligible and therefore no systematic uncertainty is assigned. The resolution tail parameters n_L and n_R have been fixed in the fit to data in region-2, -3. Due to the large correlation between n and α, the effect caused by this choice is expected to negligible.

7.6 A phenomenological study of the τ scattering contribution

So far the study of τ intermediate scattering states have purposefully been neglected by fixing $C_9^\tau = 0$ in order to understand the fit instability issues. The observed degeneracies of the 1P states, and especially of $\psi(4415)$ magnitude is expected to be resolved with more data. Including a description of the 2P state amplitude to the fit model has revealed some interesting effects. In this model, the 2P states are assumed to have an equal contribution (*i.e.* the same magnitude and phase). Nevertheless, it is clear that the 2P state contributions are larger than expected and destructively interfere with the local contribution.

As a reminder, the description of the τ scattering contribution from Eq 2.126 is written as,

$$Y_{\tau\bar{\tau}}(q^2) = -\frac{\alpha^2}{2\pi} C_9^\tau \left[h_S(m_\tau, q^2) - \frac{1}{3} h_P(m_\tau, q^2) \right], \tag{7.2}$$

where notations follow the definitions in Sec. 2.5. The negative sign in front of this equation indicates that the τ amplitude would destructively interfere with C_9 and given the nominal result, constructively interfere with the 2P states. Since the overall shape of the τ scattering amplitude is similar to that of the 2P states, it is possible that the measured magnitude of the 2P states is overestimated to include a small component of the τ scattering amplitude.

Assuming this scenario, the measured magnitude of the 2P states, η_{2P}^{meas}, is related to the assigned magnitude, η_{2P}^{assign} and C_9^τ through,

$$|\eta_{2P}^{meas}|^2 \mathbb{Y}_{2P}^2 = |C_9^\tau|^2 \mathbb{Y}_\tau^2 + |\eta_{2P}^{assign}|^2 \mathbb{Y}_{2P}^2 + |C_9^\tau| \cdot \eta_{2P}^{assign} \cdot 2Re(e^{i\delta_{2P}} \mathcal{I}), \tag{7.3}$$

where $\mathbb{Y}_{2P}^2$ and $\mathbb{Y}_{2P}^2$ are the square of the integrated contributions of the 2P and τ components with η_{2P} and C_9^τ factored out. The integrated interference between the 2P and τ components is given by $\mathcal{I}$ with the $e^{i\delta_{2P}}$, η_{2P}^{assign} and C_9^τ factored out. This quadratic equation can be solved for C_9^τ which is related to $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ through,

$$\mathcal{B}(B^+ \to K^+\tau^+\tau^-) \approx \begin{cases} 8.7 \times 10^{-9} \times |C_9^\tau|^2 & C_{10}^\tau = -C_9^\tau, \\ 2.7 \times 10^{-9} \times |C_9^\tau|^2 & C_{10}^\tau = 0, \end{cases} \tag{7.4}$$

for two cases of C_{10}^τ as shown in Ref. [50].

Using Eq. 7.3 and Eq. 7.4, the estimated $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ as a function of η_{2P}^{assign} is shown in Figure 7.17 with the SM value $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)_{SM} = (1.5 \pm 0.2) \times 10^{-7}$ highlighted [57]. To illustrate the effect arising from the interference between the 2P states and τ amplitude, different values for the phase of the 2P (δ_{2P}) states are also shown. The upper bounds on $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) < 2.25 \times 10^{-3}$ at 90% CL have been set by the BaBar experiment [56]. The three curves corresponding to different values for the phase are shown to illustrate the effect of constructive (δ_{2P}=-3.05 $\pm$ 0.13 rad), destructive (δ_{2P}=0 rad) and no interference (δ_{2P}=π/2 rad) with the τ scattering component. The three curves converge on the left because when all the measured magnitudes are assumed to be from the τ scattering contribution ($\eta_{2P}^{assign} = 0$), there is no interference term between 2P and τ scattering components. When $\delta_{2P} = \pi/2$ rad, the interference between these components is always zero and the relation is driven by a linear dependence of C_9^τ on η_{2P}^{assign}. At the measured value of $\delta_{2P} = -3.05 \pm 0.13$ rad the two components constructively interfere producing a higher $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ estimate compared to when $\delta_{2P} = \pi/2$. When $\delta_{2P} = 0$ rad there is a destructive interference between the two components allowing for a large τ scattering amplitude which is completely cancelled by the interference.

As discussed in Section 2.6, New Physics explanations to the B flavour anomalies predict a factor of 10^3 enhancement to the $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ bringing it to the level of 10^{-4}. At the

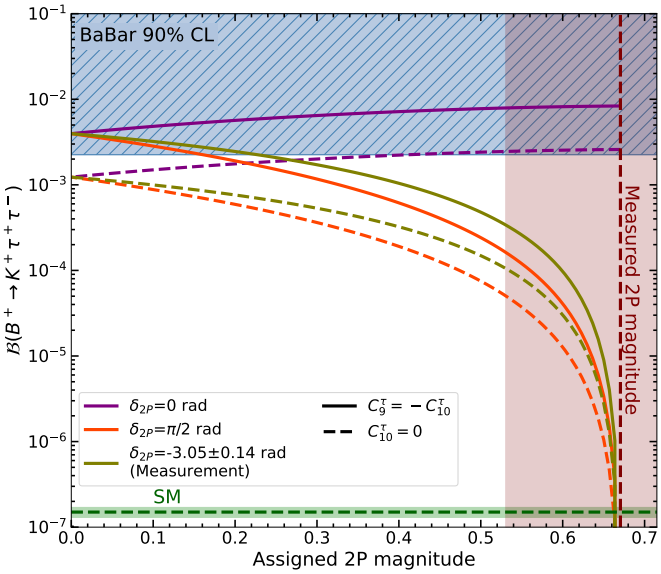

Figure 7.17: The estimated branching fraction of $B^+ \to K^+\tau^+\tau^-$ assuming a part of the measured 2P magnitude includes the τ scattering amplitude. This plot is obtained following Eq. 7.3 where for a given value of η_{2P}^{meas} and δ_{2P} a positive C_9^τ solution is found and converted into a branching fraction for $B^+ \to K^+\tau^+\tau^-$ following Eq. 7.4. Different values for the phase of the 2P states, δ_{2P}, are shown. Two different cases of $C_{10}^\tau = -C_9^\tau$ and $C_{10}^\tau = 0$ are considered. The horizontal blue band displays the experimental limits by the BaBar experiment [56]. The vertical brown band is the 1-σ error interval on the measured 2P states magnitude.

measured values for the δ_{2P} only a small portion of the magnitude is needed to compensate for the missing τ scattering amplitude. Since a large portion of the curves in Figure 7.17 are below the limits set by the BaBar collaboration, one would be able to set competitive limits on the $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$, inching closer to be able to detect a NP enhanced $B^+ \to K^+\tau^+\tau^-$.

CHAPTER

8 Conclusion and Future Prospects

The anomalies observed in decays of the B meson over the past decade have shown promising hints of physics Beyond Standard Model. However, there is no consensus on whether underestimated hadronic effects such as the non-local charm-loop contributions could potentially be the cause of the observed discrepancies. The analysis presented in this thesis uses $B^+ \to K^+\mu^+\mu^-$ decays collected by the LHCb collaboration between 2011 and 2018 to answer this question. Using a model which explicitly accounts for the local, non-local and their interferences, the existence of NP can be directly inferred.

The analysis is performed in the dimuon mass region $300 < m_{\mu\mu} < 4700$ MeV/c^2 ($0.09 < q^2 < 22.09$ GeV$^2/c^4$) accounting for the 1P and 2P intermediate non-local states. The contribution of intermediate 2P states DD, D^*D^* and DD^* are also included in the model.

The blinded values determined for the Wilson coefficients are,

$$C_9 = 16.16 \pm 0.27(\text{stat.}) \pm 0.06(\text{syst.}^*); \tag{8.1}$$

$$C_{10} = -52.03^{+0.35}_{-0.27}(\text{stat.}) \pm 0.03(\text{syst.}^*); \tag{8.2}$$

where the first uncertainties are statistical and the second uncertainties are systematic with the (*) indicating that this only includes the subset of systematics uncertainties evaluated in this thesis. The branching fraction measurements of the 1P and 2P states are presented in Table 7.1.

The interference between the local and non-local contributions was found to be large as visible in Figure 7.1. The largest contributor to this effect is from the larger than expected 2P states which show a fully destructive interference with magnitude, $\eta_{2P} = 0.67^{+0.11}_{-0.14}$ and phase, $\delta_{2P} = -3.05 \pm 0.13$ rad. In addition, excluding the 2P contribution from the model changes C_{10} by ~2 units while leaving C_9 unchanged. While anomalies in LFU tests would remain, unblinding the Wilson coefficients would reveal if some the observed B-anomalies will be explained by this non-local contribution. In addition including the τ scattering amplitude in the physics

model and treating C_9^τ as a fit parameter would reveal any NP enhanced $B^+ \to K^+\tau^+\tau^-$ decays, if present.

8.1 Future work

The decay of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ is chosen to be the normalisation mode and its branching fraction is fixed in the fit from prior knowledge (Section 6.3). The uncertainty on this value is 3%, and can be translated to an uncertainty on the branching fraction measurements. This will be done by repeating the fit to data at various normalisation branching fractions generated based on the prior uncertainty. The resulting distribution of residual values will be added as a systematic uncertainty to the nominal result. Compared to the form factor uncertainty that induces an overall uncertainty on the scale of the $B^+ \to K^+\mu^+\mu^-$ amplitude of the order of 10-15%, a 3% uncertainty on the scale of $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ is expected to be a sub-dominant systematic uncertainty.

Through the use of a 3D likelihood profile of Wilson C_9, C_{10} and $\eta_{\psi(4415)}$ the largest cause of the bias in C_9 and C_{10} will be removed. Any remaining bias for the other parameters at the nominal result will be be studied using toys and the deviations will be included as a systematic uncertainty.

The FFT method is chosen to perform the convolution due to the computational cost of calculating the full convolution integral. This method has the disadvantage that the resolution model cannot be parametrised as a function of dimuon mass and does not fully account for the variation shown in Figure 5.10 in Section 5.3.2. The systematic uncertainty arising from this choice can be estimated by generating toy data sets based on a model which fully account for the varying resolution model. Such a data set is generated using the unconvolved physics model and redistributing each point based on a dimuon mass dependent resolution model. By fitting to this data set using the fit model from Section 6.3, the distribution of the pull statistic can be assessed to estimate the uncertainty.

In the previous iteration of this measurement the potential existence of a broad resonance component at ~1890 MeV/c^2 was studied. Looking at Figure 7.1, a structure is visible which could originate from a resonances states such as $\rho(1700)$, $\omega(1650)$ and $\phi(1680)$ which have widths of 250 ± 100 MeV/c, 315 ± 35 MeV/c^2 and 150 ± 50 MeV/c^2 respectively [55]. Due to the closeness and broadness of these states, they will be investigated by including a single Breit-Wigner with its mass, width, magnitude and phase treated as fit parameter[1]. By including this component and repeating the fit to data, the significance of this state will be investigated.

[1] This additional component will be treated just as another light non-local component in Eq. 2.110

Abbreviations

ECAL Electromagnetic Calorimeter. 38, 42–44, 60

GEM Gas Electron Multiplier. 49

HCAL Hadronic Calorimeter. 38, 42

HPD Hybrid Photo Detectors. 44–47

IT Inner Tracker. 40, 41, 53, 55, 59, 60

LHC Large Hadron Collider. ix, 31, 37, 41, 46, 52, 55, 58, 63, 69

MWPC Multi-wire propotional chambers. 49

$-\log\mathcal{L}$ The negative log-likelihood of the given data set for a PDF defined by a set of parameters. . 123, 158

OT Outer Tracker. 40, 53, 55, 59, 60

PS Preshower Detector. 38, 42

p_{T} transverse momentum. 41, 49, 53–61, 76, 84, 85

RICH Ring Imaging Cherenkov. 38, 39, 43, 44, 46, 48, 50, 51, 60, 61, 63, 72

SPD Scinitillating Pad Detector. 38, 42, 53, 54, 70

`IsMuon` A binary muon identification variable. 50, 51, 70, 72

`L0DiMuon` The Level-0 muon trigger using the product of the two muons with the largest transverse momentum in each event. 54

L0Muon The Level-0 muon trigger using the muon with the largest transverse momentum. 53, 54

$\mathrm{DLL}_{\mu\pi}$ The difference in the log-likelihood of the muon and pion hypotheses.. 70, 72

muDLL The muon delta log-likelihood. Used as the discriminating variable between muon and non-muon hypotheses.. 51, 52

TT Tracker Turicensis. 38, 40, 41, 59

VELO VErtex LOcator. 38–41, 44, 53, 55, 59, 60

1P one-particle. 21–23, 25, 26, 28, 110–112, 121–123, 141, 145

2P two-particle. 21–26, 28, 29, 110–112, 116–119, 121, 122, 124, 126, 128, 130–132, 134, 135, 140–143, 145

BBDT bonsai boosted decision tree. 57

BDT boosted decision tree. 57, 73, 75–81, 84–88, 99

BF Branching Fraction. 110–112, 124, 132

BSM Beyond Standard Model. x, 1, 12, 14, 15, 18, 27, 31, 33, 34, 121, 145

CB Crystal Ball. 94, 99, 153

CCD Charge-coupled device. 63, 64

CFRP Carbon Fibre Reinforced Polymer. 45

CKM Cabbibo-Kobayashi-Maskawa. 10–14, 21, 26

CL Confidence Level. 30, 142

CP Charge Parity. 11, 18, 28, 42, 121

D0 The diameter of the circle containing 95% of the reflected and focused light from a point source. 63–66

DCB Double Crystal Ball. 94, 95, 99–101, 110, 153–156

DIRA Directional angle. The angle between the momentum vector of the track and the directional vector running between the primary vertex and the decay vertex. 70, 71

DOCA distance of closest approach. 56, 57

EFT Effective Field Theory. 1, 15, 16, 18, 21, 26, 32

FCNC Flavour Changing Neutral Current. 13, 14

FD Flight Distance. 40, 53, 57, 58, 70, 71, 73, 76

FFT Fast Fourier Transform. 94, 146

FPGA Field-Programmable Gate Arrays. 53

FSR Final State radiation. 62, 93, 94

GIM Glashow-Iliopolus-Maiani. 13, 14

HLT High Level Trigger. 53, 54, 84

HLT1 High Level Trigger 1. 54–56, 59–62

HLT2 High Level Trigger 2. 54, 56–62

HQET Heavy Quark Effective Theory. 15

IP Impact Parameter. 40, 41, 53, 55–58, 70, 71, 76, 85, 89

KDE Kernel Density Estimation. 108

L0 Level-0. 53, 54, 58, 59, 62

LCSR Light Cone Sum Rule. 20, 21, 26, 31

LEP Large Electron Positron. 37

LFU Lepton Flavour Universality. 30, 31, 34, 145

LQ LeptoQuark. 32

MC Monte-Carlo: Refers to the simulated particle decays whose behavior through the detector is also simulated. . 62

MVA multivariate analysis. 56, 57, 61, 62, 69, 73, 74, 76, 80

ndf number of degrees of freedom. 55, 56, 58

NN Neural Network. 48, 49

NP New Physics. x, 1, 2, 14, 16, 18, 19, 29–34, 67, 97, 131, 142, 143, 145, 146

OPE Operator Product Expansion. 16

PDF Probabilty Density Function. 103, 108, 111, 128, 131, 135–138, 147, 153, 154

PDG particle data group. 92, 103

PID Particle Identification. 2, 38, 39, 48, 49, 58–62, 69, 72, 80, 81, 83–88, 105

PMT Photomultiplier Tubes. 42, 43

pp A proton-proton collision.. 2, 37–41, 46, 52, 54, 59, 62

PV Primary Vertex. 40, 41, 55, 58, 59, 70, 71, 76

QCD Quantum Chromodynamics. xi, 14, 15, 17, 19–21, 139

QED Quantum Electrodynamics. 3, 8, 15, 31

QFT Quantum Field Theory. 3

RMS Root Mean Square. 93

ROC Receiver Operating Characteristic. 78

RoC Radius of Curvature. 63–66

RTA Real Time Analysis. 59

Run1 The LHC data taking period between 2011 and 2012 . 34, 37, 48, 53, 54, 57–62, 67, 69, 72, 79, 80, 91, 93, 95, 96, 99, 100, 102–106, 108, 109, 121–125, 132, 155, 156

Run2 The LHC data taking period between 2015 and 2018. xii, 34, 37, 46, 48, 52–54, 58–63, 67, 69, 72, 81, 84, 93, 95, 96, 99–105, 107–109, 121–125, 132, 155, 156

Run3 The next LHC data taking period scheduled to start in 2022. 59, 63

signal region 40 MeV/c^2 around the known B mass ((5239.33 < $m_{K\mu\mu}$ < 5319.33)). 2, 99, 102, 105, 108, 140

SM Standard Model. ix–xi, 1–7, 9–14, 16, 18, 26, 28, 30–33, 98, 111, 114–116, 118, 131, 134, 142

SV Secondary Vertex. 40

Topo-Track HLT2 tracks with additional requirements. 56, 57

Trigger Line A sequency of reconstruction algorithms and selections requirements used by the trigger. 54

UMSB Upper Mass Sideband. xix, 73, 75, 79, 102–104, 109, 140, 141

vev vacuum expectation value. 6, 7, 11

WLS wavelength-shifter. 42, 43

APPENDIX

A The Definition of a Crystal Ball Function

A Crystal Ball (CB) function is a double sided Gaussian with a power law tail on one side, the mathematical form of which is given by,

$$F_{cb}(m; m_0, \sigma, \alpha, n) = \frac{1}{N} \begin{cases} A\left(B - \frac{m-m_0}{\sigma}\right)^{-n}, & \text{for } \frac{m-m_0}{\sigma} < -|\alpha| \\ exp\left(-\frac{1}{2}\left[\frac{m-m_0}{\sigma}\right]^2\right) & \text{for } \frac{m-m_0}{\sigma} \geq -|\alpha| \end{cases}, \tag{A.1}$$

where,

$$A = \left(\frac{n}{|\alpha|}\right)^n exp\left(-\frac{|\alpha|^2}{2}\right), \tag{A.2}$$

$$B = \frac{n}{|\alpha|} - |\alpha|. \tag{A.3}$$

Here N is a normalisation constant, m_0 is the position of the peak and σ is the width. The tail parameter α defines the point at which the Gaussian transforms to a power law with power n.

In the case of a CB, the power law tail is only defined on one side. Another useful definition is the Double Crystal Ball (DCB) function which has a power law tail on both sides and written as,

$$F_{dcb}(m; m_0, \sigma, \alpha_L, \alpha_R, n_L, n_R) = \frac{1}{N} \begin{cases} A_L\left(B_L - \frac{m-m_0}{\sigma}\right)^{-n_L}, & \text{for } \frac{m-m_0}{\sigma} < -|\alpha_L| \\ exp\left(-\frac{1}{2}\left[\frac{m-m_0}{\sigma}\right]^2\right) & \text{for } -|\alpha_L| \leq \frac{m-m_0}{\sigma} < |\alpha_R| \\ A_R\left(B_R + \frac{m-m_0}{\sigma}\right)^{-n_R}, & \text{for } \frac{m-m_0}{\sigma} \geq |\alpha_R| \end{cases} \tag{A.4}$$

Here A_i, B_i, α_i and n_i are defined exactly as in Equation A.1, the subscripts L and R refer to the Left and Right CBs which make up the DCB. The DCB is used in Section 6.2.2 to model the misidentified pion background in the $m_{K\mu\mu}$ mass distribution.

In Section 5.3.3 a DCB is added with a Gaussian to model the detector's resolution. The resulting PDF is written as,

$$\mathcal{R}(m) = f_{dcb} F_{dcb}(m; m_0, \sigma_{dcb}, \alpha_L, \alpha_R, n_L, n_R) + (1 - f_{dcb}) G(m; m_0, \sigma_{gauss}), \tag{A.5}$$

where f_{dcb} is the fraction of the PDF between the DCB and Gaussian model G with mean at m_0 and width σ_{gauss}.

The model used to describe the shape of the signal decay in the $m_{K\mu\mu}$ mass fits in Section 6.2.2 is a DCB added with a narrow and wide Gaussian which is written as,

$$\begin{aligned}\mathcal{S}(m) = & f_{narrow} G_{narrow}(m; m_0, \sigma_{narrow}) + f_{wide} G_{wide}(m; m_0, \sigma_{wide}) \\ & + (1 - f_{narrow} - f_{wide}) F_{dcb}(m; m_0, \sigma_{dcb}, \alpha_L, \alpha_R, n_L, n_R).\end{aligned} \tag{A.6}$$

Here f_i are the fractions of the different PDFs which each have a different width parameter σ_i. The tail parameter α_i and n_i are estimated from fits to simulated samples while the mean m_0 and width σ_i are estimated directly from fits to data.

APPENDIX

B

Auxillary Mass Fits to Simulation

The shapes of the pion misidentified and signal candidates in the $m_{K\mu\mu}$ mass fits discussed in Section. 6.2.2 are estimated from fits to simulated samples. A DCB is used to model misidentified pion background and the fits to $m_{K\mu\mu}$ of Run1 and Run2 $B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$ simulated samples are shown in Figure B.1.

To model the signal candidates a DCB added with a narrow and wide Gaussian distribution is used. Figure B.2 shows fits to $B^+ \to K^+\mu^+\mu^-$, $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \to K^+\psi(2S)(\mu^+\mu^-)$ simulated samples for Run1 and Run2. From these fits the tail parameters of the DCB are estimated for the fit to $m_{K\mu\mu}$ data distribution.

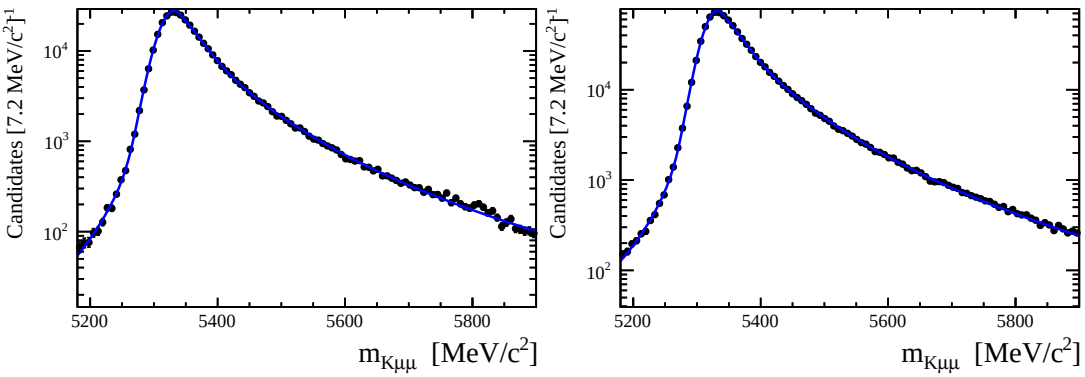

Figure B.1: The fit results to the reconstructed $m_{K\mu\mu}$ mass distribution from $B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$ simulation in Run1 (left) and Run2 (right). A DCB is used here to model the distribution.

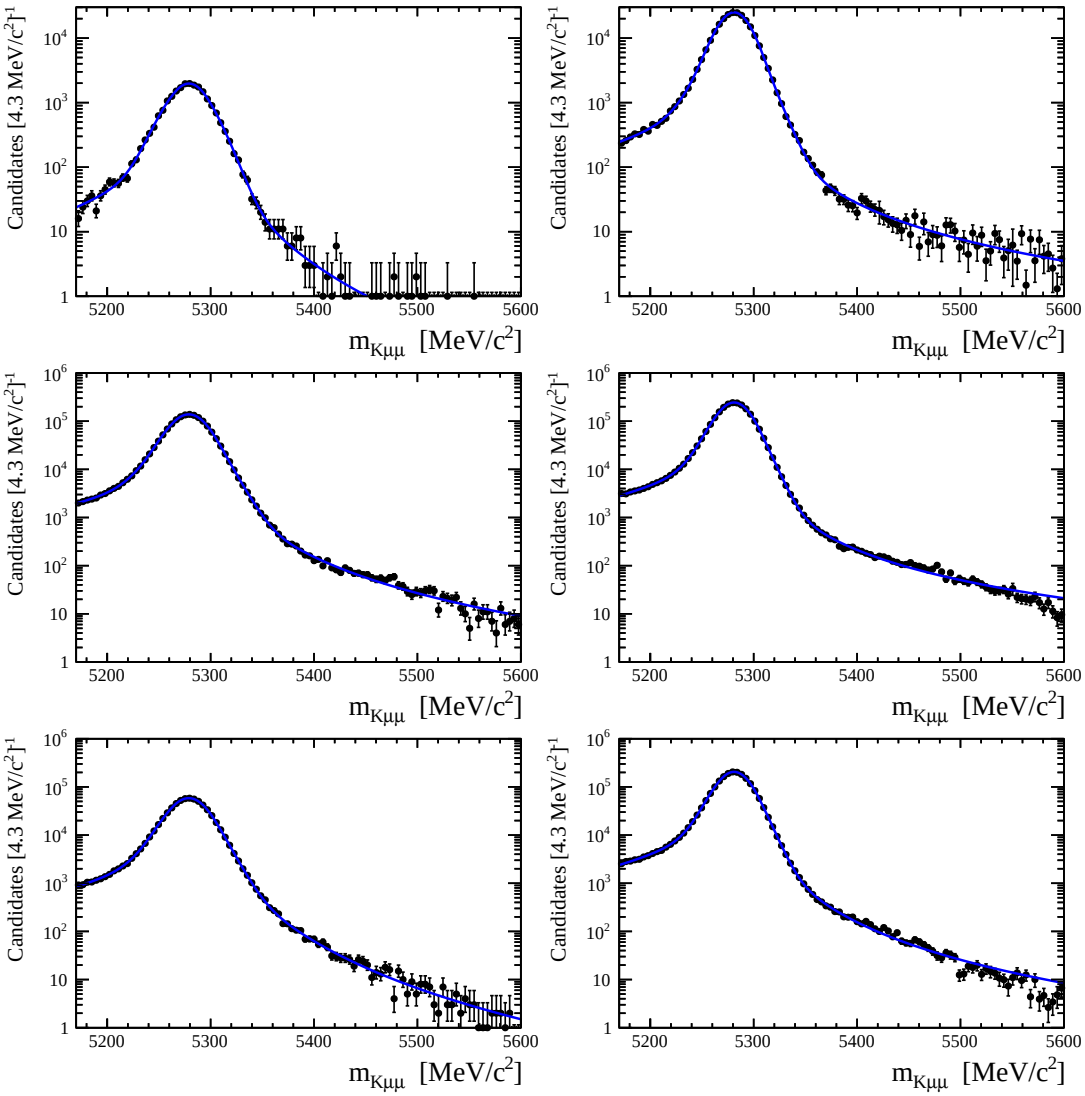

Figure B.2: The fit results to the reconstructed $m_{K\mu\mu}$ mass distribution from simulation in Run1 (left) and Run2 (right) with the rows showing fits in the regions 1,2 and 3 in order from the top. In region-1,2 and 3 the simulated samples used were $B^+ \to K^+\mu^+\mu^-$, $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ and $B^+ \to K^+\psi(2S)(\mu^+\mu^-)$ respectively. A DCB is used here to model the distribution.

APPENDIX

C Summary of the Fit Likelihood Function

The *CP* averaged differential decay rate which describe the dimuon spectrum is given by,

$$\frac{d\Gamma_\mu}{dq^2} = \frac{G_F^2\alpha^2|V_{tb}V_{ts}^*|^2}{2^7\pi^5}|\mathbf{k}|\beta_+\left\{\frac{2}{3}|\mathbf{k}|^2\beta_+^2\left|C_{10}f_+(q^2)\right|^2 + \frac{m_\mu^2(M_B^2-M_K^2)^2}{q^2M_B^2}\left|C_{10}f_0(q^2)\right|^2 + |\mathbf{k}|^2\left[1-\frac{1}{3}\beta_+^2\right]\left|C_9^{\text{eff}}f_+(q^2) + 2C_7^{\text{eff}}\frac{m_b+m_s}{M_B+M_K}f_T(q^2)\right|^2\right\}, \tag{C.1}$$

where all notations follow Sections 2.3-2.5.

The decay rate is corrected for the detector's efficiency ($\mathcal{E}_i(m_{\mu\mu}^{\text{true}})$) evaluated in Section 5.2 and convolved with the resolution models ($\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$), described in Section 5.3, in the three dimuon mass regions to give the physics model,

$$P(m_{\mu\mu}^{\text{rec}}) = \begin{cases} \mathcal{R}_1(m_{\mu\mu}^{\text{rec}}, m_{\mu\mu}^{\text{true}}) \circledast \left[\sum_i \mathcal{E}_i(m_{\mu\mu}^{\text{true}})\cdot 2m_{\mu\mu}^{\text{true}}\frac{d\Gamma_\mu^i}{dq^2}\right] & \text{for } m_{\mu\mu}^{\text{rec}} \in [300, 1800]\ \text{MeV}/c^2 \\ \mathcal{R}_2(m_{\mu\mu}^{\text{rec}}, m_{\mu\mu}^{\text{true}}) \circledast \left[\sum_i \mathcal{E}_i(m_{\mu\mu}^{\text{true}})\cdot 2m_{\mu\mu}^{\text{true}}\frac{d\Gamma_\mu^i}{dq^2}\right] & \text{for } m_{\mu\mu}^{\text{rec}} \in [1800, 3400]\ \text{MeV}/c^2 \\ \mathcal{R}_3(m_{\mu\mu}^{\text{rec}}, m_{\mu\mu}^{\text{true}}) \circledast \left[\sum_i \mathcal{E}_i(m_{\mu\mu}^{\text{true}})\cdot 2m_{\mu\mu}^{\text{true}}\frac{d\Gamma_\mu^i}{dq^2}\right] & \text{for } m_{\mu\mu}^{\text{rec}} \in [3400, 4700]\ \text{MeV}/c^2 \end{cases} \tag{C.2}$$

where the iterator i labels the the three double differential decay rate terms, shown in Eq. 5.1, each with a different $cos\theta_\ell$ dependence and their corresponding efficiency curves.

The combinatorial background model ($B_{\text{comb}}(m_{\mu\mu}^{\text{rec}})$) and misidentified pion background model ($B_\pi(m_{\mu\mu}^{\text{rec}})$) are estimated in Sections 6.2.3 and 6.2.4. These models are added to the physics model according to their respective fractions (f_{comb}, f_π) calculated from the B mass fits in Section 6.2.2 to give the fit model,

$$F(m_{\mu\mu}^{\text{rec}}) = (1 - f_{\text{comb}} - f_\pi)P(m_{\mu\mu}^{\text{rec}}) + f_{\text{comb}}B_{\text{comb}}(m_{\mu\mu}^{\text{rec}}) + f_\pi B_\pi(m_{\mu\mu}^{\text{rec}}). \tag{C.3}$$

An unbinned maximum likelihood estimation method is implemented using Minuit [136] and Tensorflow. This method estimates the parameter values which maximize the likelihood value. In practice however it is convenient to define the negative of the log of the likelihood ($-\log\mathcal{L}$), this is given by,

$$\begin{aligned} -\log\mathcal{L}(\mathbf{p}) &= -\log\prod_{i=1}^{N} F(m^{\text{rec}}_{\mu\mu};\mathbf{p}) \\ &= -\sum_{i=1}^{N}\log F(m^{\text{rec}}_{\mu\mu};\mathbf{p}), \end{aligned} \tag{C.4}$$

where N is the total number of candidates and $F(m^{\text{rec}}_{\mu\mu};\mathbf{p})$ is the value of the normalised fit model at $m^{\text{rec}}_{\mu\mu}$ given the set of parameters $\mathbf{p}$. The set of parameters $\hat{\mathbf{p}}$ which minimizes $-\log\mathcal{L}$,

$$-\frac{d\log\mathcal{L}(\mathbf{p})}{d\mathbf{p}}\bigg|_{\hat{\mathbf{p}}} = 0, \tag{C.5}$$

defines the most likely model which describes the data.

Bibliography

[1] **LHCb Collaboration**, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, arXiv:2103.11769.

[2] **LHCb Collaboration**, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, arXiv:1705.05802.

[3] **LHCb Collaboration**, R. Aaij *et al.*, *Test of lepton universality with $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays*, JHEP **05** (2020) 040, arXiv:1912.08139.

[4] **LHCb Collaboration**, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays*, JHEP **06** (2014) 133, arXiv:1403.8044.

[5] **HFLAV**, Y. S. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, Eur. Phys. J. **C81** (2021) 226, arXiv:1909.12524, updated results and plots available at https://hflav.web.cern.ch/.

[6] **LHCb Collaboration**, R. Aaij *et al.*, *Angular Analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ Decay*, Phys. Rev. Lett. **126** (2021) 161802, arXiv:2012.13241.

[7] **LHCb Collaboration**, R. Aaij *et al.*, *Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay*, Phys. Rev. Lett. **125** (2020) 011802, arXiv:2003.04831.

[8] G. Isidori, D. Lancierini, P. Owen, and N. Serra, *On the significance of new physics in $b \to s\ell^+\ell^-$ decays*, Phys. Lett. B **822** (2021) 136644, arXiv:2104.05631.

[9] M. Algueró *et al.*, *$b \to s\ell^+\ell^-$ Global Fits after R_{K_S} and $R_{K^{*+}}$*, arXiv:2104.08921.

[10] M. Algueró, B. Capdevila, A. Crivellin, and J. Matias, *Disentangling Lepton Flavour Universal and Lepton Flavour Universality Violating Effects in $b \to s\ell^+\ell^-$ Transitions*, arXiv:2205.15212.

[11] J. Lyon and R. Zwicky, *Resonances gone topsy turvy - the charm of QCD or new physics in $b \to s\ell^+\ell^-$?*, arXiv:1406.0566.

[12] **LHCb Collaboration**, R. Aaij *et al.*, *Measurement of the phase difference between short- and long-distance amplitudes in the $B^+ \to K^+\mu^+\mu^-$ decay*, Eur. Phys. J. C **77** (2017) 161, arXiv:1612.06764.

[13] A. J. Buras, *Gauge Theory of Weak Decays: The Standard Model and the Expedition to New Physics Summits*, Cambridge University Press, 2020.

[14] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8** (1964) 214.

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

[16] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[17] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531.

[18] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[19] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[20] **CKMfitter Group**, J. Charles *et al.*, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. C **41** (2005) 1, arXiv:hep-ph/0406184.

[21] **UTfit**, M. Bona *et al.*, *The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters from Lattice QCD: A Reappraisal after the Measurements of Delta m(s) and BR(B —> tau nu(tau))*, JHEP **10** (2006) 081, arXiv:hep-ph/0606167.

[22] **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. B **716** (2012) 1, arXiv:1207.7214.

[23] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (1970) 1285.

[24] H. D. Politzer, *Reliable perturbative results for strong interactions?*, Phys. Rev. Lett. **30** (1973) 1346.

[25] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, Phys. Rev. Lett. **30** (1973) 1343.

[26] A. V. Manohar and M. B. Wise, *Heavy quark physics*, vol. 10, 2000.

[27] T. Mannel, *Effective field theories in flavour physics - Introduction*, vol. 203, 2004.

[28] A. J. Buras, *Theoretical review of b-physics*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **368** (1995) 1, Proceedings of the Third International Workshop on B-Physics at Hadron Machines.

[29] R. P. Feynman and M. Gell-Mann, *Theory of the fermi interaction*, Phys. Rev. **109** (1958) 193.

[30] D. van Dyk, *The Decays $\bar{B} \to \bar{K}^{(*)}\ell^+\ell^-$ at Low Recoil and their Constraints on New Physics*, PhD thesis, Dortmund U., 2012.

[31] B. Grinstein, M. J. Savage, and M. B. Wise, *$B \to X_s e^+ e^-$ in the six-quark model*, Nuclear Physics B **319** (1989) 271.

[32] M. Misiak, *The $b \to s e^+ e^-$ and $b \to s\gamma$ decays with next-to-leading logarithmic QCD corrections*, Nucl. Phys. B **393** (1993) 23, [Erratum: Nucl.Phys.B 439, 461–465 (1995)].

[33] A. J. Buras and M. Munz, *Effective Hamiltonian for $B \to X_s e^+ e^-$ beyond leading logarithms in the NDR and HV schemes*, Phys. Rev. D **52** (1995) 186, arXiv:hep-ph/9501281.

[34] C. Bobeth, G. Hiller, and G. Piranishvili, *Angular distributions of $\bar{B} \to \bar{K}\ell^+\ell^-$ decays*, JHEP **12** (2007) 040, arXiv:0709.4174.

[35] C. Bobeth, M. Misiak, and J. Urban, *Photonic penguins at two loops and m_t dependence of $BR[B \to X_s l^+ l^-]$*, Nucl. Phys. B **574** (2000) 291, arXiv:hep-ph/9910220.

[36] A. Khodjamirian, T. Mannel, and Y. M. Wang, *$B \to K\ell^+\ell^-$ decay at large hadronic recoil*, JHEP **02** (2013) 010, arXiv:1211.0234.

[37] A. J. Buras, M. Misiak, M. Munz, and S. Pokorski, *Theoretical uncertainties and phenomenological aspects of $B \to X_s\gamma$ decay*, Nucl. Phys. B **424** (1994) 374, arXiv:hep-ph/9311345.

[38] C. Bobeth, G. Hiller, and D. van Dyk, *More Benefits of Semileptonic Rare B Decays at Low Recoil: CP Violation*, JHEP **07** (2011) 067, arXiv:1105.0376.

[39] W. Altmannshofer *et al.*, *Symmetries and Asymmetries of $B \to K^*\mu^+\mu^-$ Decays in the Standard Model and Beyond*, JHEP **01** (2009) 019, arXiv:0811.1214.

[40] R. Alonso, B. Grinstein, and J. Martin Camalich, *$SU(2) \times U(1)$ gauge invariance and the shape of new physics in rare B decays*, Phys. Rev. Lett. **113** (2014) 241802, arXiv:1407.7044.

[41] J. Aebischer, A. Crivellin, M. Fael, and C. Greub, *Matching of gauge invariant dimension-six operators for $b \to s$ and $b \to c$ transitions*, JHEP **05** (2016) 037, arXiv:1512.02830.

[42] T. Hurth and M. Nakao, *Radiative and Electroweak Penguin Decays of B Mesons*, Ann. Rev. Nucl. Part. Sci. **60** (2010) 645, arXiv:1005.1224.

[43] P. Ball and R. Zwicky, *New results on $b \to \pi, k, \eta$ decay form factors from light-cone sum rules*, Phys. Rev. D **71** (2005) 014015, arXiv:hep-ph/0406232.

[44] J. A. Bailey *et al.*, *$B \to K l^+ l^-$ Decay Form Factors from Three-Flavor Lattice QCD*, Phys. Rev. D **93** (2016) 025026, arXiv:1509.06235.

[45] N. Isgur and M. B. Wise, *Relationship Between Form-factors in Semileptonic $\bar{B}$ and D Decays and Exclusive Rare $\bar{B}$ Meson Decays*, Phys. Rev. D **42** (1990) 2388.

[46] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y.-M. Wang, *Charm-loop effect in $B \to K^{(*)} \ell^+ \ell^-$ and $B \to K^* \gamma$*, JHEP **09** (2010) 089, arXiv:1006.4945.

[47] C. G. Boyd, B. Grinstein, and R. F. Lebed, *Constraints on form-factors for exclusive semileptonic heavy to light meson decays*, Phys. Rev. Lett. **74** (1995) 4603, arXiv:hep-ph/9412324.

[48] C. Bourrely, I. Caprini, and L. Lellouch, *Model-independent description of $B \to \pi l \nu$ decays and a determination of $|V_{ub}|$*, Phys. Rev. D **79** (2009) 013008, arXiv:0807.2722, [Erratum: Phys.Rev.D 82, 099902 (2010)].

[49] **HPQCD Collaboration**, C. Bouchard *et al.*, *Rare decay $B \to K \ell^+ \ell^-$ form factors from lattice QCD*, Phys. Rev. D **88** (2013) 054509, arXiv:1306.2384, [Erratum: Phys.Rev.D 88, 079901 (2013)].

[50] C. Cornella *et al.*, *Hunting for $B^+ \to K^+ \tau^+ \tau^-$ imprints on the $B^+ \to K^+ \mu^+ \mu^-$ dimuon spectrum*, Eur. Phys. J. C **80** (2020) 1095, arXiv:2001.04470.

[51] S. M. Flatte, *Coupled - Channel Analysis of the pi eta and K anti-K Systems Near K anti-K Threshold*, Phys. Lett. B **63** (1976) 224.

[52] S. Okubo, *Phi meson and unitary symmetry model*, Phys. Lett. **5** (1963) 165.

[53] G. Zweig, in *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, D. B. Lichtenberg and S. P. Rosen, eds., pp. 22–101, 1964.

[54] J. Iizuka, *Systematics and phenomenology of meson family*, Prog. Theor. Phys. Suppl. **37** (1966) 21.

[55] **Particle Data Group**, P. A. Zyla *et al.*, *Review of Particle Physics*, PTEP **2020** (2020) 083C01.

[56] **BaBar Collaboration**, J. P. Lees *et al.*, *Search for $B^+ \to K^+\tau^+\tau^-$ at the BaBar experiment*, Phys. Rev. Lett. **118** (2017) 031802, arXiv:1605.09637.

[57] **HPQCD Collaboration**, C. Bouchard *et al.*, *Standard Model Predictions for $B \to K\ell^+\ell^-$ with Form Factors from Lattice QCD*, Phys. Rev. Lett. **111** (2013) 162002, arXiv:1306.0434, [Erratum: Phys.Rev.Lett. 112, 149902 (2014)].

[58] **LHCb Collaboration**, R. Aaij *et al.*, *Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction*, JHEP **11** (2016) 047, arXiv:1606.04731, [Erratum: JHEP 04, 142 (2017)].

[59] **LHCb Collaboration**, R. Aaij *et al.*, *Branching Fraction Measurements of the Rare $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$- Decays*, Phys. Rev. Lett. **127** (2021) 151801, arXiv:2105.14007.

[60] **LHCb Collaboration**, R. Aaij *et al.*, *Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays*, JHEP **06** (2015) 115, arXiv:1503.07138, [Erratum: JHEP 09, 145 (2018)].

[61] G. Hiller and F. Kruger, *More model-independent analysis of $b \to s$ processes*, Phys. Rev. D **69** (2004) 074020, arXiv:hep-ph/0310219.

[62] M. Bordone, G. Isidori, and A. Pattori, *On the Standard Model predictions for R_K and R_{K^*}*, Eur. Phys. J. C **76** (2016) 440, arXiv:1605.07633.

[63] **LHCb Collaboration**, R. Aaij *et al.*, *Analysis of Neutral B-Meson Decays into Two Muons*, Phys. Rev. Lett. **128** (2022) 041801, arXiv:2108.09284.

[64] **LHCb Collaboration**, R. Aaij *et al.*, *Measurement of the $B_s^0 \to \mu^+\mu^-$ decay properties and search for the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ decays*, Phys. Rev. D **105** (2022) 012010, arXiv:2108.09283.

[65] **ATLAS Collaboration**, M. Aaboud *et al.*, *Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector*, JHEP **04** (2019) 098, arXiv:1812.03017.

[66] **CMS Collaboration**, A. M. Sirunyan *et al.*, *Measurement of properties of $B_s^0 \to \mu^+\mu$ decays and search for $B^0 \to \mu^+\mu$ with the CMS experiment*, Journal of High Energy Physics **2020** (2020) 188, arXiv:1910.12127.

[67] **LHCb Collaboration**, R. Aaij *et al.*, *Search for the Rare Decays $B_s^0 \to e^+e^-$ and $B^0 \to e^+e^-$*, Phys. Rev. Lett. **124** (2020) 211802, arXiv:2003.03999.

[68] **LHCb Collaboration**, R. Aaij *et al.*, *Search for the decays $B_s^0 \to \tau^+\tau^-$ and $B^0 \to \tau^+\tau^-$*, Phys. Rev. Lett. **118** (2017) 251802, arXiv:1703.02508.

[69] M. Algueró *et al.*, *Emerging patterns of New Physics with and without Lepton Flavour Universal contributions*, Eur. Phys. J. C **79** (2019) 714, arXiv:1903.09578, [Addendum: Eur.Phys.J.C 80, 511 (2020)].

[70] L. Alasfar *et al.*, *B anomalies under the lens of electroweak precision*, JHEP **12** (2020) 016, arXiv:2007.04400.

[71] B. Capdevila *et al.*, *Searching for New Physics with $b \to s\tau^+\tau^-$ processes*, Phys. Rev. Lett. **120** (2018) 181802, arXiv:1712.01919.

[72] C. Cornella, J. Fuentes-Martin, and G. Isidori, *Revisiting the vector leptoquark explanation of the B-physics anomalies*, JHEP **07** (2019) 168, arXiv:1903.11517.

[73] **LHCb Collaboration**, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[74] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.

[75] S. Borghi *et al.*, *First spatial alignment of the LHCb VELO and analysis of beam absorber collision data*, Nucl. Instrum. Meth. A **618** (2010) 108.

[76] **LHCb Collaboration**, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30** (2015) 1530022, arXiv:1412.6352.

[77] **LHCb Collaboration**, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

[78] C. Abellán Beteta *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, arXiv:2008.11556.

[79] R. Forty and O. Ullaland, in *Particle Identification: Time-of-Flight, Cherenkov and Transition Radiation Detectors*, C. W. Fabjan and H. Schopper, eds., pp. 281–335, Springer, Cham, 2020.

[80] **LHCb RICH Group**, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2013) 2431, arXiv:1211.6759.

[81] C. Matteuzzi, *A RICH with aerogel for a hadron collider*, Nucl. Instrum. Meth. A **553** (2005) 356.

[82] R. Calabrese *et al.*, *Performance of the LHCb RICH detectors during LHC Run 2*, LHCb-DP-2021-004 (2022) arXiv:2205.13400.

[83] R. Forty, *Ring imaging Cherenkov detectors for LHC-B*, Nucl. Instrum. Meth. A **384** (1996) 167.

[84] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, EPJ Tech. Instrum. **6** (2019) 1, arXiv:1803.00824.

[85] **LHCb Collaboration**, *Global PID performance for charged particles*, , LHCB-FIGURE-2020-012.

[86] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[87] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.

[88] **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.

[89] **LHCb Collaboration**, R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

[90] G. Corti *et al.*, *Software for the LHCb experiment*, IEEE Trans. Nucl. Sci. **53** (2006) 1323.

[91] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking: Algorithm and Performance Studies*, LHCb-2007-015, CERN-LHCb-2007-015, CERN, Geneva, 2007.

[92] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Journal of Basic Engineering **82** (1960) 35.

[93] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. A **262** (1987) 444.

[94] E. Polycarpo and M. Gandelman, *The performance of the LHCb muon identification procedure*, CERN-LHCB-2007-145, CERN, Geneva, 2007.

[95] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth and Brooks, Monterey, CA, 1984.

[96] H.-J. Yang, B. P. Roe, and J. Zhu, *Studies of boosted decision trees for MiniBooNE particle identification*, Nucl. Instrum. Meth. A **555** (2005) 370, arXiv:physics/0508045.

[97] 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.

[98] M. Frank *et al.*, *Deferred High Level Trigger in LHCb: A Boost to CPU Resource Utilization*, J. Phys. Conf. Ser. **513** (2014) 012006.

[99] **LHCb Collaboration**, *RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector* , LHCB-FIGURE-2020-016, 2020.

[100] R. Aaij *et al.*, *Tesla : an application for real-time data analysis in High Energy Physics*, Comput. Phys. Commun. **208** (2016) 35, arXiv:1604.05596.

[101] **LHCb Computing Project**, S. Roiser and C. Bozzi, *The LHCb Software and Computing Upgrade towards LHC Run 3*, J. Phys. Conf. Ser. **1085** (2018) 032049.

[102] R. Aaij *et al.*, *A comprehensive real-time analysis model at the LHCb experiment*, JINST **14** (2019) P04006, arXiv:1903.01360.

[103] O. Callot, *FastVelo, a fast and efficient pattern recognition package for the Velo*, LHCb-PUB-2011-001, CERN, Geneva, 2011.

[104] E. E. Bowen, B. Storaci, and M. Tresch, *VeloTT tracking for LHCb Run II*, LHCb-PUB-2015-024, CERN, Geneva, 2016.

[105] **LHCb Collaboration**, R. Quagliani, *Novel real-time alignment and calibration of LHCb detector for Run II and tracking for the upgrade.*, J. Phys. Conf. Ser. **762** (2016) 012046.

[106] O. Callot and M. Schiller, *PatSeeding: A standalone track reconstruction algorithm*, LHCB-2008-042, CERN, Geneva, 2008.

[107] M. Needham and J. Van Tilburg, *Performance of the track matching*, LHCb-2007-020, 2007.

[108] A. Höcker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis* , CERN, Geneva, 2007. CERN-OPEN-2007-007.

[109] P. Speckmayer, A. Höcker, J. Stelzer, and H. Voss, *The toolkit for multivariate data analysis, TMVA 4*, IOP Publishing **219** (2010) 032057.

[110] M. De Cian, S. Farry, P. Seyfert, and S. Stahl, *Fast neural-net based fake track rejection in the LHCb reconstruction*, LHCb-PUB-2017-011, CERN, Geneva, 2017.

[111] T. Likhomanenko *et al.*, *LHCb Topological Trigger Reoptimization*, J. Phys. Conf. Ser. **664** (2015) 082025, arXiv:1510.00572.

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

[113] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

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

[115] 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.

[116] **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.

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

[118] **LHCb collaboration**, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007, 2012.

[119] M. Adinolfi *et al.*, *LHCb Upgraded RICH 1 Engineering Design Review Report*, LHCb-PUB-2016-014, CERN, Geneva, 2016.

[120] M. Needham, *Identification of Ghost Tracks using a Likelihood Method*, LHCb-2008-026, CERN, Geneva, 2008.

[121] N. Kazeev, *Machine Learning for particle identification in the LHCb detector.*, PhD thesis, University of Rome, 2020, Presented 21 Oct 2020.

[122] **LHCb Collaboration**, M. Hushchyn and V. Chekalina, *Particle-identification techniques and performance at LHCb in Run 2*, Nucl. Instrum. Meth. A **936** (2019) 568.

[123] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **543** (2005) 577.

[124] J. Zhu, S. Rosset, H. Zou, and T. Hastie, *Multi-class adaboost*, Statistics and its interface **2** (2006) .

[125] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, Journal of Computer and System Sciences **55** (1997) 119.

[126] F. Pedregosa *et al.*, *Scikit-learn: Machine learning in Python*, Journal of Machine Learning Research **12** (2011) 2825.

[127] L. Anderlini *et al.*, *The PIDCalib package*, LHCb-PUB-2016-021, CERN, Geneva, 2016.

[128] W. D. Hulsbergen, *Decay chain fitting with a kalman filter*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **552** (2005) 566.

[129] M. Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015. Software available from tensorflow.org.

[130] **ARGUS Collaboration**, H. Albrecht *et al.*, *Measurement of the polarization in the decay $B \to J/\psi K^*$*, Phys. Lett. B **340** (1994) 217.

[131] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, arXiv:physics/0306116.

[132] M. Jung, *Branching ratio measurements and isospin violation in B-meson decays*, Phys. Lett. B **753** (2016) 187, arXiv:1510.03423.

[133] **BES Collaboration**, M. Ablikim *et al.*, *Determination of the psi(3770), psi(4040), psi(4160) and psi(4415) resonance parameters*, eConf **C070805** (2007) 02, arXiv:0705.4500.

[134] T. M. Karbach and M. Schlupp, *Constraints on Yield Parameters in Extended Maximum Likelihood Fits*, arXiv:1210.7141.

[135] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. **9** (1938) 60.

[136] F. James and M. Winkler, *MINUIT User's Guide* , CERN, Geneva, 2004.