The scalar component of $B^0 \to K^+\pi^-\mu^+\mu^-$ decays

Samuel Thomas Cunliffe

High Energy Physics
Blackett Laboratory
Imperial College London

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

January 6, 2016

Abstract

The differential branching fraction of $B^0 \to K_1^*(892)^0\mu^+\mu^-$ decays is measured as a function of the squared invariant mass of the dimuon system. The data corresponds to $3\,\text{fb}^{-1}$ of integrated luminosity collected in 2011 and 2012 by the LHCb detector at the Large Hadron Collider at CERN. Integrated across squared dimuon invariant mass, and interpolated through excluded regions, the total branching fraction is found to be

$$\mathcal{B}\left[B^0 \to K_1^*(892)^0\mu^+\mu^-\right] = (1.058^{+0.017}_{-0.016} \pm 0.013 \pm 0.070) \times 10^{-6}.$$

In the theoretically favoured region of squared dimuon invariant mass, $(1.1 < q^2 < 6)\,\text{GeV}^2/c^4$, the differential branching fraction is found to be

$$\frac{\mathrm{d}\mathcal{B}\left[B^0 \to K_1^*(892)^0\mu^+\mu^-\right]}{\mathrm{d}q^2} = (0.402^{+0.020}_{-0.019} \pm 0.008 \pm 0.027) \times 10^{-7}.$$

In the two results above, the first uncertainty is statistical, the second systematic, and the third due to the uncertainty of the branching fraction of the normalisation channel. The differential branching fraction is in agreement with, although lower than, the Standard Model prediction of $(0.49 \pm 0.08) \times 10^{-7}$. As with the measurement of several decay modes including a quark-level $b \to s\ell^+\ell^-$ transition, the Standard Model predictions are consistently higher than the measured values.

In this analysis, for the first time, the fraction of S-wave in the $K\pi$ system is measured and explicitly accounted for. Previous analyses had measured the branching fraction of both P- and S-wave components and compared to predictions for pure P-wave. In the same theoretically favoured region of $(1.1 < q^2 < 6)\,\text{GeV}^2/c^4$, and for the $K\pi$ invariant mass range $(796 < m_{K\pi} < 996)\,\text{MeV}/c^2$, the fraction of S-wave is found to be

$$F_\text{S} = 0.097 \pm 0.016 \pm 0.008,$$

where the first uncertainty is statistical, and the second is systematic. This value is somewhat larger than expected, although no concrete theoretical predictions exist.

Declaration

The specific contributions to the work presented are as follows. Chapter 4, I adapted the selection against specific backgrounds from those presented in [1]. Section 4.3 describes an updated strategy very heavily based on the work in Ref. [1]. The work of Sections 4.3.4 and 4.3.5, in particular the development of the isolation variables, was performed by other collaborators. I performed all of the performance evaluation detailed in Section 4.4. Appendix D contains work of myself and other collaborators. In Chapters 5, 6, 7 and all other appendices, I performed the whole analysis.

– *Sam Cunliffe*
August 2015

Acknowledgments

This thesis would not have been possible without the support, encouragement, guidance and pressure (*'where's my plot?', 'I don't understand why this is taking so long'*) of my supervisor Mitesh. Thank you so much for everything.

I am deeply indebted to Kostas Petridis, Chris Parkinson, Mike McCann, and Ulrik Egede for starting me off on this project, explaining countless things, general guidance, advice, threaded code to calculate big matrices, sarcastic comments and debugging.

A big thank you to my non-physicist friends, too many to enumerate. We hung out in Reading or Southampton, went to gigs, ate dim sum together, climbed together, and you let me crash on your sofas. You're unlikely to read this, but thank you all the same. My friends Sam Hall and Robyn Lucas looked after me in Ferney-Voltaire and cooked some amazing dinners. Cheers neutrino guys: Pip Hamilton and Dom Brailsford for being an excellent housemate and for staring too much (respectively).

Thanks also to the so-far-unmentioned members of the '*LHCb* $B^0 \to K^*\mu^+\mu^-$ *analysis team*': Tom Blake, Marcin Chrząszcz, Samuel Coquereau, Nicola Serra, and Thomas Nikodem for all of your help, cross-checks, and scrutiny. Thanks Grieg Cowan and Stéphane T'Jampens for all of their helpful feedback and to Grieg for suggesting we look at the $Z(4430)$. Thank you Jocelyn Monroe and Oliver Buchmueller for, believe it or not, a very enjoyable viva exam.

Also thanks to current and past members of the Imperial LHCb group. In no particular order: Greg Ciezarek, Alex Shires, Patrick Owen, Paula Álvarez Cartelle, Dave Websdale, Federico Redi and Peter Dornan. For advice, ideas, feedback, practice talks and helpful comments on my slides.

Many thanks to David Straub for quickly providing additional Standard Model predictions in Table 7.2 and providing Fig. 8.1.

I gratefully acknowledge the Science and Technology Funding Council for their financial support throughout my studentship and their administrative help whilst I was on long-term attachment to CERN.

Many revisions of document were extensively (and heroically) proofread by Mitesh Patel. Additionally, various sections were checked by Ulrik Egede and Mike McCann. Any remaining errors are my own.

Finally thanks to my family: for everything else.

Contents

List of abbreviations

$\boldsymbol{A_{\mathrm{FB}}}$ the *CP*-averaged forward-backward asymmetry of the lepton (defined in Eqn. 2.138)

BDT boosted decision tree

BBDT 'Bonsai' boosted decision tree

BSM (Theories of physics) Beyond the Standard Model

CB Crystal Ball (function)

CKM Cabibbo–Kobayashi–Maskawa (matrix, shown in Eqns 2.69 and 2.70)

CP charge-parity

D-wave the $K\pi$ system in a tensor, $J = 2$ state

DLL Difference in Log-Likelihoods (between two particle hypotheses, for example)

$\mathbf{DLL}_{\boldsymbol{K\pi}}$ the DLL between a kaon and pion hypotheses (defined in Eqn. 3.2)

$\mathbf{DLL}_{\boldsymbol{p\pi}}$ the DLL between a proton and pion hypotheses (defined analogously to Eqn. 3.2)

$\mathbf{DLL}_{\boldsymbol{\mu\pi}}$ the DLL between a muon and pion hypotheses (defined in Eqn. 3.3)

DOCA distance of closest approach (between two tracks)

ECAL electromagnetic calorimeter

FCNC flavour-changing neutral current (decays which change quark flavour without changing the charge)

FD flight distance (of a particle before its decay)

$\mathbf{FD}\boldsymbol{\chi^2}$ flight distance χ^2 (the change in χ^2 when two vertices are combined into a single vertex fit)

$\boldsymbol{F_{\mathrm{L}}}$ the fraction of longitudinal polarised amplitudes, with respect to the total P-wave amplitudes (defined in Eqn. 2.136)

$1 - F_L$ the fraction of transverse polarised amplitudes, with respect to the total P-wave amplitudes (defined in Eqn. 2.137)

F_S the fractional size of the S-wave component (defined precisely in Section 6.3)

$F_S|_{644}^{1200}$ F_S within the window $m_{K\pi} \in [644, 1200]\,\text{MeV}/c^2$ (defined by Eqn. 6.36)

$F_S|_{796}^{996}$ F_S within the window $m_{K\pi} \in [796, 996]\,\text{MeV}/c^2$, *i.e.* $\pm 10\,\text{MeV}/c^2$ around the nominal mass of the $K_1^*(892)^0$ (defined by Eqn. 6.36)

GEMs gas electron multipliers

gen. generator-level

h.c. Hermitian conjugate

HCAL hadronic calorimeter

HLT1 the first software trigger

`Hlt1TrackAllL0` HLT1 track trigger path

`Hlt1TrackMuon` HLT1 muon track trigger path

HLT2 the second software trigger

`Hlt2SingleMuon` HLT2 algorithm for energetic single muon tracks (described in Table 3.2)

`Hlt2DiMuonDetatched` HLT2 algorithm for selecting pairs of muon tracks (described in Table 3.2)

`Hlt2TopoMuNBodyBBDT` HLT2 BBDT algorithm for selecting topological n-body objects with one or more muon candidates

`Hlt2TopoNBodyBBDT` HLT2 BBDT algorithm for selecting topological n-body objects without muon candidates

HQET heavy quark effective theory

IP impact parameter (the minimum distance of a track to a vertex)

IPχ^2 impact parameter χ^2 (the change in the χ^2 of a vertex fit when including a track)

`IsMuon` a boolean muon identification variable

IT (finely spaced) inner tracker sections

L0 low level hardware trigger 'level-zero'

`L0Muon` trigger path using information from the muon system

`L0DiMuon` a second trigger path using information from the muon system making a requirement on the product of the momenta of two muon candidates

LCSR light cone sum rules

LHC Large Hadron Collider

LHCb Large Hadron Collider beauty

M1–5 muon stations

MoM method of moments

MVA multivariate analysis

MWPCs multi-wire proportional chambers

$\boldsymbol{m_{\mu\mu}}$ dimuon invariant mass

$\boldsymbol{m_{K\pi}}$ $K\pi$ invariant mass

$\boldsymbol{m_{K\pi\mu\mu}}$ full B^0 candidate invariant mass

NP new physics (particles, effects) from BSM theories

NR non-resonant

$\boldsymbol{N_{\mathrm{tracks}}}$ the number of tracks in an event

OPE operator-product expansion

OT (more coarsely spaced) outer tracker sections

P-wave the $K\pi$ system in a vector, $J = 1$ state

pdf probability density function

phsp phase-space

PID particle identification

$\boldsymbol{pp}$ proton–proton

PS preshower detector

$\boldsymbol{p_{\mathrm{T}}}$ component of the momentum transverse to the beam line

PV primary vertex

q^2 squared dilepton invariant mass ($q^2 = m_{\mu\mu}{}^2$)

QCD quantum chromodynamics

QCDF QCD factorisation

QED quantum electrodynamics

R1–4 regions of the muon stations

reco. selected and reconstructed

RICH ring-imaging Cherenkov subdetector

ROC reciever operating characteristic (a plot of true-positive rate versus false-positive rate)

R_K the ratio of branching fractions of the $B^+ \to K^+\mu^+\mu^-$ to $B^+ \to K^+e^+e^-$ decays (defined in Eqn. 2.93)

$\sqrt{s}$ centre-of-mass energy

S-wave the $K\pi$ system in a scalar, $J = 0$ state

SM Standard Model of particle physics

SPD scintillating pad detector

SUSY supersymmetry

SV secondary vertex

T1–3 downstream tracking stations

TIS trigger independently of signal, the signal candidate was required to have not been selected by the trigger

TOS trigger on signal, the signal candidate was required to have been selected by the trigger

TT upstream tracking station

TV tertiary vertex

VELO vertex locator

vev vacuum expectation value

η pseudorapidity, defined as $\eta = -\ln(\tan(\theta/2))$

$\boldsymbol{\vartheta}$ pointing angle, between the reconstructed direction of the four-body candidate and the FD direction

$\boldsymbol{\theta_K}$ the angle between the kaon in the rest frame of the $K\pi$ system and the direction of the $K\pi$ object in the rest frame of the B meson

$\mathbf{cos}\,\boldsymbol{\theta_K}$ the cosine of θ_K (defined in Eqn. 2.94)

$\boldsymbol{\theta_\ell}$ the angle between the positively (negatively) charged lepton in the rest frame of the dilepton system and the direction of the dilepton object in the rest frame of the B ($\overline{B}$) meson

$\mathbf{cos}\,\boldsymbol{\theta_\ell}$ the cosine of θ_ℓ (defined in Eqns 2.95 and 2.96)

$\boldsymbol{\Lambda}_{\mathbf{QCD}}$ the scale of the theory of QCD

$\boldsymbol{\phi}$ the angle between the planes of the $K\pi$ and dilepton systems in the rest frame of the B meson (defined in Eqns 2.97–2.100)

$\boldsymbol{\chi^2}/\mathbf{ndf}$ χ^2 divided by the number of degrees of freedom

List of figures

List of tables

Chapter 1

Introduction

It has been said that the overall goal of the study of fundamental physics is to answer the question

$$\mathcal{L} = ? \tag{1.1}$$

where $\mathcal{L}$ is the Lagrangian density [2]. This statement was presumably written somewhat in jest, but is in reality, rather close to the truth.

The working theory of fundamental particle physics is the so-called Standard Model (SM). In this theory, matter and three of the four known fundamental forces are described as interacting quantum fields. Aspects of the SM have been verified to exceptional accuracy therefore

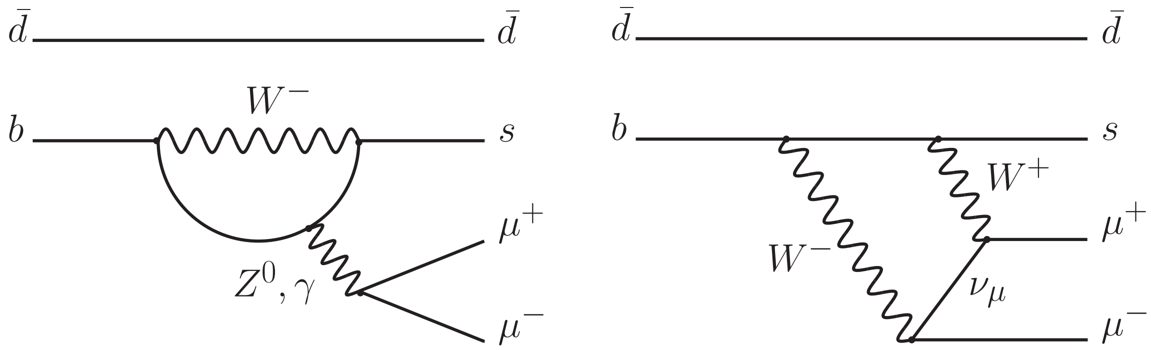
$$\mathcal{L} \approx \mathcal{L}_{\mathrm{SM}}. \tag{1.2}$$

There are at least three problems with the SM which necessitate the approximation in Eqn. 1.2. Firstly, gravity is not successfully described at the quantum scale and is omitted from the SM. Secondly, there is evidence for matter which interacts through gravity but does not emit electromagnetic radiation [3, 4][1]. There is no description of this *dark matter* in the SM. The third problem is the *naturalness* and *fine tuning* of the theory. In the SM, many parameters are set to values provided by experiment. The seemingly arbitrary input parameters mean that the SM is unsatisfactory as a *fundamental* theory. So whilst the SM provides an excellent description of the results of subatomic particle physics experiments, one hopes that there will be some theory of physics Beyond the Standard Model (BSM) which will alleviate these problems.

It is natural to ask: what does one require of a theory of physics? Occam's razor dictates that the theory be the simplest thing which describes all experimental results. In this context, 'simple' means that the theory should have few input parameters and that the minimal version is best. The theory should make new predictions contrary to the previous theory which are verified by experiment. The theory should unify the forces, provide some candidate for dark matter, and describe the force of gravity.

[1] Review 24 in [4]

Figure 1.1: The two leading order Standard Model (SM) diagrams for the process $\overline{B}^0 \to \overline{K}^{(*)0}\mu^+\mu^-$. The electroweak penguin diagram (left) and box diagram (right) are examples of flavour-changing neutral current (FCNC) $b \to s\ell^+\ell^-$ transitions which, in the SM, which must occur with intermediary, virtual particles.

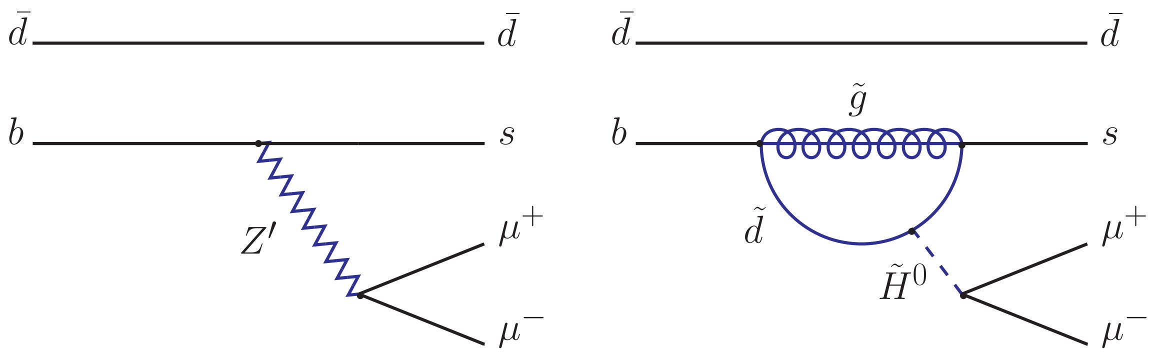


The class of supersymmetry (SUSY) theories, which postulate one extra symmetry in the quantum field theory, is one of the simplest plausible extensions to the SM. The additional symmetry implies a duplicate of all known fundamental particles. Unfortunately, at the time of writing, no statistically significant direct experimental evidence for SUSY has been observed [5].

This thesis presents a measurement of a physical observable quantity as an indirect probe for physics BSM. The decay process $B^0 \to K^{*0}\mu^+\mu^-$ is one example of a general class of decays mediated by the quark-level $b \to s\ell^+\ell^-$ transition. The transition of a quark into another of the same electric charge requires a flavour-changing neutral current (FCNC) and is then sensitive to contributions from virtual BSM particles. The leading order Feynman diagrams for the $B^0 \to K^{(*)0}\mu^+\mu^-$ process in the SM are shown in Fig. 1.1. The left hand diagram in Fig. 1.1 is known as an *electroweak penguin*. 'Electroweak' since the dilepton pair are propagated by a virtual electroweak Z boson, 'penguin' since, with some imagination, the diagram could be seen to resemble a penguin. In the SM, all FCNCs must proceed via higher order diagrams (*loop-level* processes) like those shown in Fig. 1.1. This is because it is not possible to directly connect two neutral quarks of differing flavours via a single vertex (a *tree-level* transition). This fact is a consequence of the structure of the SM and is described in Section 2.1.4. This may not be the case in BSM theories, which suffer no such restriction. Examples of Feynman diagrams for this process mediated by possible BSM particles are given in Fig. 1.2. Since the SM diagrams are suppressed, BSM diagrams may occur at a comparable level to the SM diagrams.

The branching fraction, $\mathcal{B}$, is the fraction of B^0 mesons which decay to some final state with respect to all possible final states. The branching fraction for the decay process $B^0 \to K^*(892)^0\mu^+\mu^-$ is a physically observable quantity which is sensitive to contributions from these BSM particles. It is also useful to express the branching fraction as a differential

Figure 1.2: Two example Beyond the Standard Model (BSM) diagrams for the process $\overline{B}^0 \to \overline{K}^{(*)0}\mu^+\mu^-$. The process could be mediated by a virtual Z' vector boson at tree-level (left) or by supersymmetric sparticles (right). Whilst FCNCs do not occur at tree-level in the SM, there is no such restriction in BSM theories. The BSM diagrams may therefore occur at a comparable level to the SM diagrams.



branching fraction in squared dilepton invariant mass (q^2) which allows for a more rigorous comparison to theoretical predictions in the SM. As will be explained in more detail in the following chapter, observables related to the decay $B^0 \to K^{*0}\mu^+\mu^-$ suffer from a largely undesirable pollution when the K^{*0}; which, for the analysis presented in this thesis, is reconstructed from the decay $K^{*0} \to K\pi$; is in a scalar $J = 0$, or *S-wave*, state. This *S-wave component* (or *S-wave pollution*) has the effect of diluting the sensitivity to possible BSM effects. This thesis presents the first measurement of the size of this S-wave component. Having extracted the S-wave component, a measurement of the (differential) branching fraction of the decay with a pure vector K^* (P-wave) is performed for the first time. The analysis is performed on data collected by the Large Hadron Collider beauty (LHCb) detector at the Large Hadron Collider (LHC) during 2011 and 2012.

This thesis is organised as follows. A summary of the flavour-changing structure of the SM, followed by a motivation for the study of the $b \to s\ell^+\ell^-$ process, is presented in Chapter 2. In the final sections of Chapter 2, some angular conventions, and the theoretical angular distribution are introduced. The LHCb detector is introduced in Chapter 3 and the selection of $B^0 \to K^+\pi^-\mu^+\mu^-$ events in data is described in Chapter 4. The fractional size of the S-wave component ($F_{\rm S}$) is extracted with a partial angular analysis comprising a fit to $K\pi$ invariant mass ($m_{K\pi}$). The fit is performed over the range $m_{K\pi} \in [644, 1200]\,\text{MeV}/c^2$. This range is chosen such that the effect of tensor D-wave resonances (such as the $K_2^*(1430)^0$) are vanishingly small [6]. This fit procedure requires a description of the angular efficiency and the efficiency in regions of q^2. The efficiency function is extracted from simulation, the procedure for which is described in Chapter 5. The model and fit procedure used to extract $F_{\rm S}$ and the differential branching fraction is presented in Chapter 6. Finally, results are presented in Chapter 7 followed by the conclusions in Chapter 8.

Throughout most of this thesis charge conjugation is implied. The exceptions are: Section 2.5, where angles are defined separately for the charge-parity (CP)-conjugate process; and Section 2.6, where the angular distribution is derived but subsequently the CP-averaged observables are used. Specific kaon resonances are always specified with subscript J in addition to the full name allocated by the PDG, *e.g.* $K^*_1(892)^0$, $K^*_0(1430)^0$. The notation K^* denotes any exited kaon of any angular momentum state[2].

[2]Elsewhere in literature the specific process in the case of the vector $B^0 \to K^*_1(892)^0\ell^+\ell^-$ is often referred to as simply $B^0 \to K^*\ell^+\ell^-$ for shorthand.

Chapter 2

Theory of flavour-changing neutral currents and $b \rightarrow s\ell^+\ell^-$ processes

The decay $B^0 \rightarrow K^{(*)0}\ell^+\ell^-$ is a FCNC process. Such processes serve to illustrate the hierarchical flavour structure of the SM. This structure within the SM, and why it arises, is described in Section 2.1. Extensions to the SM often predict some differing flavour structure, and/or extra particles, which must contribute to the loop-level processes virtually. With many possible models, a largely model-independent effective field theory framework known as the operator-product expansion (OPE) is employed in order to assist the search for physics BSM. The effective field theory approach is motivated and described in Section 2.2. The specific example of the OPE for $b \rightarrow s\ell^+\ell^-$ processes such as $B^0 \rightarrow K^{*0}\mu^+\mu^-$ is described in Section 2.3. The current status, and phenomenological interpretations of $b \rightarrow s\ell^+\ell^-$ measurements is presented in Section 2.4. This serves to motivate the interest in $B^0 \rightarrow K_1^*(892)^0\ell^+\ell^-$ observables and gives context to the measurement presented in this thesis. The differential branching fraction is corrected for the contribution of the scalar component. This is achieved by a partial angular analysis, therefore angular conventions, and the angular differential decay rate for $B^0 \rightarrow K_J^*\ell^+\ell^-$ are described in Sections 2.5 and 2.6.

2.1 Flavour structure of the Standard Model

The argument presented in this section follows Ref. [7] with additional reference to [8, 9].

The SM is a quantum field theory with local gauge symmetry under the group

$$\mathsf{SU}(3) \times \mathsf{SU}(2) \times \mathsf{U}(1). \tag{2.1}$$

The group $\mathsf{SU}(3)$ describes the strong interaction through the theory of quantum chromodynamics (QCD) [10], with eight gluon fields, A_μ^g $(g = 1...8)$. The groups $\mathsf{SU}(2)\times\mathsf{U}(1)$ describe the unified electroweak interaction [11] with four associated gauge fields, $W_\mu^a(a = 1, 2, 3)$

and B_μ. In addition, there is a minimal procedure for spontaneous symmetry breaking through the Higgs mechanism [12]. The matter content of the theory consists of three flavours of leptons and three flavours of quarks described by Dirac spinor fields. Of importance to the present discussion, the left- and right-handed chiral fields are given by

$$\psi_R = P_R\psi = \frac{(1+\gamma^5)}{2}\psi, \tag{2.2}$$

$$\psi_L = P_L\psi = \frac{(1-\gamma^5)}{2}\psi, \tag{2.3}$$

where $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$ is the product of all four Dirac gamma matrices (γ^μ).

The lepton fields can be arranged into three doublets

$$L^i = \begin{bmatrix} \nu_L^i \\ \ell_L^i \end{bmatrix} \in \left\{ \begin{bmatrix} \nu_{eL} \\ e_L \end{bmatrix}, \begin{bmatrix} \nu_{\mu L} \\ \mu_L \end{bmatrix}, \begin{bmatrix} \nu_{\tau L} \\ \tau_L \end{bmatrix} \right\} \tag{2.4}$$

and three singlets

$$\ell^i = (\ell_R^i) \in \{e_R, \mu_R, \tau_R\}\,. \tag{2.5}$$

Right-handed neutrinos are not included in the theory. The quark fields may be written in a similar way,

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

with right-handed singlets of both up- and down-type quarks,

$$u^i = (u_R^i) \in \{u_R, c_R, t_R\} \tag{2.7}$$

and

$$d^i = (d_R^i) \in \{d_R, s_R, b_R\}\,. \tag{2.8}$$

The flavour indices $i = 1, 2, 3$ select one of the flavour doublets or singlets. Summation is implicit over pairs of flavour indices.

The SM Lagrangian density is the most general renormalizable Lagrangian density consistent with the observed particle fields and the gauge symmetry of the theory. It may be written in three parts,

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Kinetic}} + \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Higgs}}. \tag{2.9}$$

The first part of this equation contains terms of the form[1]

$$\mathcal{L}_{\text{Kinetic}} \sim i\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi, \tag{2.10}$$

describing the kinetic energy of the fields, ψ. In these kinetic terms, the gauge covariant derivative is

$$\mathcal{D}_\mu = \left(\partial_\mu + igA_\mu^gT^g + ig_2W_\mu^a\tau^a + ig_1B_\mu Y\right). \tag{2.11}$$

[1]Spacetime contraction is implicit over pairs of spacetime indices μ, ν.

The second, third, and fourth terms contain the generators of $\mathsf{SU}(3)$, the T^g; of $\mathsf{SU}(2)$, the τ^a; and the $\mathsf{U}(1)$ hypercharge generator, Y. The g, g_1, and g_2 are coupling constants. The second term in Eqn. 2.9 describes the gauge fields with terms like

$$\mathcal{L}_{\text{Gauge}} \sim -iF^{\mu\nu}F_{\mu\nu}, \tag{2.12}$$

where $F^{\mu\nu}$ is a Gauge field strength tensor. These terms do not distinguish between the generations of the quarks and leptons. They are sometimes called *flavour blind* or *flavour diagonal* in the literature.

2.1.1 The complex scalar doublet field

The last term in Eqn. 2.9 is the source of flavour-changing structure in the SM. It is

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}^\mu H)^\dagger(\mathcal{D}_\mu H) - V(H) + \mathcal{L}^\ell_{\text{Yukawa}} + \mathcal{L}^q_{\text{Yukawa}}, \tag{2.13}$$

where the terms, in order, are: the Higgs field's kinetic term; its potential; and its Yukawa interaction terms with the lepton and quark fields. Flavour structure arises due to these Yukawa terms. The Higgs doublet is formed of two complex scalar fields,

$$H = \begin{bmatrix} H_+ \\ H_0 \end{bmatrix}, \tag{2.14}$$

and the simplest renormalizable potential is of the form[2]

$$V(H) = \lambda(H^\dagger H)^2 - \mu^2 H^\dagger H \tag{2.15}$$

$$= \lambda\left((H^\dagger H) - \frac{\mu^2}{2\lambda}\right)^2 - \frac{\mu^4}{4\lambda}. \tag{2.16}$$

This has a non-zero minimum for $\mu^2 > 0$. The minimum is

$$H^\dagger H = \frac{\mu^2}{2\lambda}, \tag{2.17}$$

so the vacuum expectation value (vev) of the Higgs field is

$$\sqrt{\frac{\mu^2}{2\lambda}} \equiv \frac{1}{\sqrt{2}}\nu. \tag{2.18}$$

Without loss of generality one can perform a global gauge transformation so that the vacuum state of the field is,

$$\langle H\rangle = \begin{bmatrix} 0 \\ \frac{\nu}{\sqrt{2}} \end{bmatrix}. \tag{2.19}$$

An example of such a potential is shown in Fig 2.1, the vev corresponds to the trough in this figure. A global gauge choice is equivalent to a choice of position in the trough of Fig 2.1.

[2]Some references define the potential without the relative sign between the λ and μ^2 terms. If one does this then one must flip the inequality of μ^2 later.

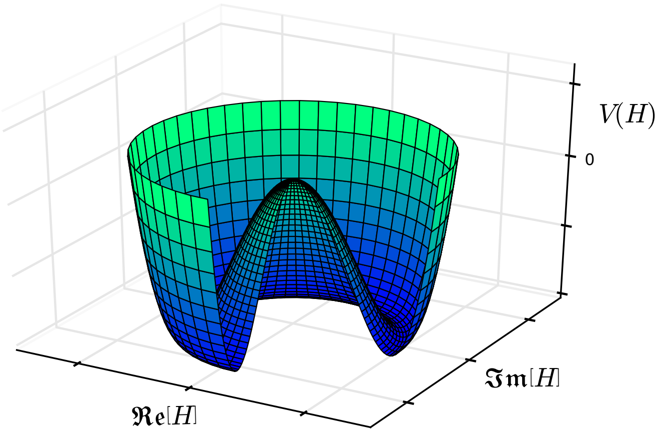


Figure 2.1: A potential of the form of Eqn. 2.15 with $\mu^2 > 0$. A sector is omitted to display the inside. The central minimum of the potential is positive, spontaneously breaking the symmetry corresponds to choosing a position in the trough.

The field can be expanded about this expectation value to obtain

$$H = \begin{bmatrix} h_+ + i\chi_+ \\ \frac{\nu}{\sqrt{2}} + h_0 + i\chi_0 \end{bmatrix}, \tag{2.20}$$

where h_+, $\chi_{+,0}$ are real scalar fields. Reinserting this into Eqn. 2.15 and expanding, yields only one mass term[3]

$$2\nu^2 h_0^2 \equiv m_H^2 h_0^2 \implies m_H = \sqrt{\frac{\lambda}{2}}\nu. \tag{2.21}$$

The fields h_+, χ_+, and χ_0 represent the massless Goldstone bosons[4]. Local gauge transformations further allow one to remove the Goldstone boson fields. In this *unitary gauge*, the form simplifies to

$$H = \begin{bmatrix} 0 \\ \frac{\nu}{\sqrt{2}} + h_0 \end{bmatrix}. \tag{2.22}$$

2.1.2 Aside: mass terms for the gauge bosons

Although not directly related to the flavour structure of the SM, the Higgs mechanism is introduced in order to provide mass terms for the gauge bosons in a gauge invariant way. This important result will now be highlighted as an aside. The gauge covariant derivative for the Higgs doublet field is,

$$\mathcal{D}_\mu H = \left(\partial_\mu + ig_2 W_\mu^a \frac{\sigma^a}{2} + i\frac{g_1}{2} B_\mu\right) H. \tag{2.23}$$

The Pauli sigma matrices, are the generators of $\mathsf{SU}(2)$ $\tau^a = \sigma^a/2$ $(a = 1, 2, 3)$ for doublets and $Y = 1/2$ is the hypercharge assignment for the Higgs field.

[3]From dimensionality arguments, the Klein-Gordon mass term for a boson field must be quadratic in the field (*i.e.* $\sim m^2\phi^2$).

[4]In fact this is a more general result than for a complex doublet. For any n scalar fields introduced, breaking the symmetry in this way will result in $n-1$ massless fields and one with a mass term such as h_0.

Substituting Eqn. 2.22 into the terms $(\mathcal{D}^\mu H)^\dagger(\mathcal{D}_\mu H)$ of Eqn. 2.13 yields the terms

$$\frac{g_2^2\nu^2}{8}(W^{1\mu}W^1_\mu + W^{2\mu}W^2_\mu) \tag{2.24}$$

and

$$\frac{\nu^2}{8}(g_2W^{3\mu} - g_1B^\mu)(g_2W^3_\mu - g_1B_\mu) \tag{2.25}$$

amongst others. Substituting

$$W^\pm_\mu \equiv \frac{W^1_\mu \mp iW^2_\mu}{\sqrt{2}} \tag{2.26}$$

into Eqn. 2.24 gives a mass term for a charged vector boson field $W^\pm$,

$$\frac{g_2^2\nu^2}{4}W^{+\mu}W^-_\mu \equiv m_W^2W^{+\mu}W^-_\mu \implies m_W = \frac{g_2\nu}{2}. \tag{2.27}$$

One can redefine the couplings g_1 and g_2 in terms of a weak mixing angle, θ_W,

$$g_1 = \sin\theta_W \cdot \sqrt{g_1^2 + g_2^2}, \tag{2.28}$$

$$g_2 = \cos\theta_W \cdot \sqrt{g_1^2 + g_2^2}, \tag{2.29}$$

and substitute these formulae into Eqn. 2.25 obtaining

$$\frac{\nu^2\,(g_1^2 + g_2^2)}{8}(W^{3\mu}\cos\theta_W - B^\mu\sin\theta_W)(W^3_\mu\cos\theta_W - B_\mu\sin\theta_W). \tag{2.30}$$

Substituting

$$Z_\mu \equiv W^3_\mu\cos\theta_W - B_\mu\sin\theta_W, \tag{2.31}$$

into Eqn. 2.30 gives a mass term for a neutral vector boson field Z^0,

$$\frac{\nu^2\,(g_1^2 + g_2^2)}{8}Z^\mu Z_\mu \equiv m_Z^2 Z^\mu Z_\mu \implies m_Z = \frac{m_W}{\cos\theta_W}. \tag{2.32}$$

Finally, the massless field

$$A_\mu \equiv W^3_\mu\cos\theta_W + B_\mu\sin\theta_W, \tag{2.33}$$

is the photon described by quantum electrodynamics (QED).

To summarise this important result, the minimal Higgs mechanism (a complex scalar doublet) is introduced into the theory. The Higgs field acquires a vev which breaks the electroweak symmetry $\mathsf{SU}(2) \times \mathsf{U}(1)$ in such a way as to introduce massive Higgs, $W^\pm$, and Z^0 bosons and a massless photon. The parameters introduced are related by simple expressions,

$$m_H = \sqrt{\frac{\lambda}{2}}\nu; \quad m_W = \frac{g_2\nu}{2}; \quad m_Z = \frac{m_W}{\cos\theta_W}. \tag{2.34}$$

2.1.3 The Yukawa terms

A Yukawa interaction is an interaction between any Dirac field and a scalar field. The interaction of the lepton and quark fields with the Higgs doublet is given by

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = -y^{\ell}_{ij}\overline{L^i}H\ell^j + \text{h.c.}, \tag{2.35}$$

and

$$\mathcal{L}^{q}_{\text{Yukawa}} = -y^{d}_{ij}\overline{Q^i}Hd^j - y^{u}_{ij}\overline{Q^i}\epsilon Hu^j + \text{h.c.}, \tag{2.36}$$

where h.c. denotes Hermitian Conjugate and the y_{ij} are the couplings. The 2×2 antisymmetric matrix,

$$\epsilon = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{2.37}$$

serves to select the up-type fields. There is no equivalent term for the neutrinos. Following the principle of including all renormalizable terms consistent with gauge invariance and the matter content of the theory, one is obliged include Eqns. 2.35 and 2.36 in the Lagrangian density of Eqn. 2.13.

After the Higgs field acquires a vev the Yukawa terms simplify. Substituting Eqn. 2.22 into Eqn. 2.35 the lepton Yukawa term becomes:

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = -\, y^{\ell}_{ij}\begin{bmatrix}\overline{\nu^i_L} & \overline{\ell^i_L}\end{bmatrix}\begin{bmatrix} 0 \\ \nu/\sqrt{2} + h_0 \end{bmatrix}\ell^j_R - y^{\ell}_{ij}\overline{\ell^i_R}\begin{bmatrix} 0 & \frac{\nu}{\sqrt{2}} + h_0 \end{bmatrix}\begin{bmatrix}\nu^j_L \\ \ell^j_L\end{bmatrix} \tag{2.38}$$

$$= -\,\frac{y^{\ell}_{ij}\nu}{\sqrt{2}}\overline{\ell^i_L}\ell^j_R - \frac{y^{\ell}_{ij}\nu}{\sqrt{2}}\overline{\ell^i_R}\ell^j_L + \text{terms with } h_0. \tag{2.39}$$

The first two terms are Dirac mass terms for the lepton field, except that the Yukawa couplings, y^{ℓ}_{ij} have two flavour indices ij. They may be re-written as a matrix product in a *flavour space* with,

$$\boldsymbol{\ell}_L \equiv \begin{bmatrix} e_L \\ \mu_L \\ \tau_L \end{bmatrix}; \quad \boldsymbol{\ell}_R \equiv \begin{bmatrix} e_R \\ \mu_R \\ \tau_R \end{bmatrix}; \quad \boldsymbol{Y}^{\ell} \equiv \begin{bmatrix} y^{\ell}_{11} & y^{\ell}_{12} & y^{\ell}_{13} \\ y^{\ell}_{21} & y^{\ell}_{22} & y^{\ell}_{23} \\ y^{\ell}_{31} & y^{\ell}_{32} & y^{\ell}_{33} \end{bmatrix}. \tag{2.40}$$

Equation 2.39 becomes

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = -\frac{\nu}{\sqrt{2}}\left(\overline{\boldsymbol{\ell}_L}\boldsymbol{Y}\boldsymbol{\ell}_R + \overline{\boldsymbol{\ell}_R}\boldsymbol{Y}\boldsymbol{\ell}_L\right) + \text{terms with } h_0. \tag{2.41}$$

Similarly for the quark Yukawa terms, Eqn. 2.22 is substituted into Eqn. 2.36, which yields

$$\begin{aligned}\mathcal{L}^{q}_{\text{Yukawa}} = &-\, y^{d}_{ij}\begin{bmatrix}\overline{u^i_L} & \overline{d^i_L}\end{bmatrix}\begin{bmatrix} 0 \\ \nu/\sqrt{2} + h_0 \end{bmatrix}d^j_R - y^{d}_{ij}\overline{d^i_R}\begin{bmatrix} 0 & \nu/\sqrt{2} + h_0 \end{bmatrix}\begin{bmatrix}u^j_L \\ d^j_L\end{bmatrix} \\ &- y^{u}_{ij}\begin{bmatrix}\overline{u^i_L} & \overline{d^i_L}\end{bmatrix}\begin{bmatrix} \frac{\nu}{\sqrt{2}} + h_0 \\ 0 \end{bmatrix}u^j_R - y^{u}_{ij}\overline{u^i_R}\begin{bmatrix} \frac{\nu}{\sqrt{2}} + h_0 & 0 \end{bmatrix}\begin{bmatrix}u^j_L \\ d^j_L\end{bmatrix}\end{aligned} \tag{2.42}$$

$$\begin{aligned}= &-\,\frac{y^{d}_{ij}\nu}{\sqrt{2}}\overline{d^i_L}d^j_R - \frac{y^{d}_{ij}\nu}{\sqrt{2}}\overline{d^i_R}d^j_L - \frac{y^{u}_{ij}\nu}{\sqrt{2}}\overline{u^i_L}u^j_R - \frac{y^{u}_{ij}\nu}{\sqrt{2}}\overline{u^i_R}u^j_L \\ &+ \text{terms with } h_0.\end{aligned} \tag{2.43}$$

Defining

$$\boldsymbol{d}_L \equiv \begin{bmatrix} d_L \\ s_L \\ b_L \end{bmatrix}; \quad \boldsymbol{d}_R \equiv \begin{bmatrix} d_R \\ s_R \\ b_R \end{bmatrix}; \quad \boldsymbol{Y}^d \equiv \begin{bmatrix} y^d_{11} & y^d_{12} & y^d_{13} \\ y^d_{21} & y^d_{22} & y^d_{23} \\ y^d_{31} & y^d_{32} & y^d_{33} \end{bmatrix}; \tag{2.44}$$

and

$$\boldsymbol{u}_L \equiv \begin{bmatrix} u_L \\ c_L \\ t_L \end{bmatrix}; \quad \boldsymbol{u}_R \equiv \begin{bmatrix} u_R \\ c_R \\ t_R \end{bmatrix}; \quad \boldsymbol{Y}^u \equiv \begin{bmatrix} y^u_{11} & y^u_{12} & y^u_{13} \\ y^u_{21} & y^u_{22} & y^u_{23} \\ y^u_{31} & y^u_{32} & y^u_{33} \end{bmatrix}; \tag{2.45}$$

one may be similarly re-write,

$$\begin{aligned} \mathcal{L}^q_{\text{Yukawa}} = &- \frac{\nu}{\sqrt{2}} \left(\overline{\boldsymbol{d}_L} \boldsymbol{Y}^d \boldsymbol{d}_R + \overline{\boldsymbol{d}_R} \boldsymbol{Y}^d \boldsymbol{d}_L \right) - \frac{\nu}{\sqrt{2}} \left(\overline{\boldsymbol{u}_L} \boldsymbol{Y}^u \boldsymbol{u}_R + \overline{\boldsymbol{u}_R} \boldsymbol{Y}^u \boldsymbol{u}_L \right) \\ &+ \text{terms with } h_0. \end{aligned} \tag{2.46}$$

2.1.4 Diagonalising the Yukawa matrix in flavour space

Any matrix may be diagonalised with a bi-unitary transformation. Therefore the following transformations

$$\begin{array}{llllll} \boldsymbol{\ell}_L & \longmapsto & U^{\ell}_{(L)} \boldsymbol{\ell}'_L; & \boldsymbol{\ell}_R & \longmapsto & U^{\ell}_{(R)} \boldsymbol{\ell}'_R; \\ \boldsymbol{d}_L & \longmapsto & U^{d}_{(L)} \boldsymbol{d}'_L; & \boldsymbol{d}_R & \longmapsto & U^{d}_{(R)} \boldsymbol{d}'_R; \\ \boldsymbol{u}_L & \longmapsto & U^{u}_{(L)} \boldsymbol{u}'_L; & \boldsymbol{u}_R & \longmapsto & U^{u}_{(R)} \boldsymbol{u}'_R; \end{array} \tag{2.47}$$

are defined. In these expressions, the $U^{\ell,d,u}_{(R/L)}$ are six independent 3×3 unitary matrices required to bring the matrices of Yukawa couplings into diagonal form. By design, they yield

$$\boldsymbol{Y}^{\ell}_{\text{diag.}} = U^{\ell\dagger}_{(R)} \boldsymbol{Y}^{\ell} U^{\ell}_{(L)} = \frac{\sqrt{2}}{\nu} \begin{bmatrix} m_e & 0 & 0 \\ 0 & m_\mu & 0 \\ 0 & 0 & m_\tau \end{bmatrix}; \tag{2.48}$$

$$\boldsymbol{Y}^{d}_{\text{diag.}} = U^{d\dagger}_{(R)} \boldsymbol{Y}^{d} U^{d}_{(L)} = \frac{\sqrt{2}}{\nu} \begin{bmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{bmatrix}; \tag{2.49}$$

$$\boldsymbol{Y}^{u}_{\text{diag.}} = U^{u\dagger}_{(R)} \boldsymbol{Y}^{u} U^{u}_{(L)} = \frac{\sqrt{2}}{\nu} \begin{bmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{bmatrix}. \tag{2.50}$$

Equation 2.41 then becomes

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = - \frac{\nu}{\sqrt{2}} \left(\overline{\boldsymbol{\ell}'_L} \boldsymbol{Y}_{\text{diag.}} \boldsymbol{\ell}'_R + \overline{\boldsymbol{\ell}'_R} \boldsymbol{Y}_{\text{diag.}} \boldsymbol{\ell}'_L \right) + \text{terms with } h_0 \tag{2.51}$$

$$\begin{aligned} = &- m_e \left(\overline{e'_L} e'_R + \overline{e'_R} e'_L \right) - m_\mu \left(\overline{\mu'_L} \mu'_R + \overline{\mu'_R} \mu'_L \right) - m_\tau \left(\overline{\tau'_L} \tau'_R + \overline{\tau'_R} \tau'_L \right) \\ &+ \text{terms with } h_0, \end{aligned} \tag{2.52}$$

which are the lepton mass terms. Similarly Eqn. 2.46 becomes

$$\begin{aligned}\mathcal{L}^{q}_{\text{Yukawa}} &= -\frac{\nu}{\sqrt{2}}\left(\overline{\boldsymbol{d}'_L}\boldsymbol{Y}^d_{\text{diag.}}\boldsymbol{d}'_R + \overline{\boldsymbol{d}'_R}\boldsymbol{Y}^d_{\text{diag.}}\boldsymbol{d}'_L\right) - \frac{\nu}{\sqrt{2}}\left(\overline{\boldsymbol{u}'_L}\boldsymbol{Y}^u_{\text{diag.}}\boldsymbol{u}'_R + \overline{\boldsymbol{u}'_R}\boldsymbol{Y}^u_{\text{diag.}}\boldsymbol{u}'_L\right) \\ &\quad + \text{terms with } h_0 \end{aligned} \tag{2.53}$$

$$\begin{aligned} &= - m_d\left(\overline{d'_R}d'_L + \overline{d'_L}d'_R\right) - m_s\left(\overline{s'_R}s'_L + \overline{s'_L}s'_R\right) - m_b\left(\overline{b'_R}b'_L + \overline{b'_L}b'_R\right) \\ &\quad - m_u\left(\overline{u'_R}u'_L + \overline{u'_L}u'_R\right) - m_c\left(\overline{c'_R}c'_L + \overline{c'_L}c'_R\right) - m_t\left(\overline{t'_R}t'_L + \overline{t'_L}t'_R\right) \\ &\quad + \text{terms with } h_0 \end{aligned} \tag{2.54}$$

which are the quark mass terms. The redefinition of the fermion fields in Eqn. 2.47 to the eigenstates of the Yukawa coupling matrix provides a *mass basis*, denoted with primes. Reverting to the element notation (with flavour indices i, j, k, m, p, q), the original $\mathsf{SU}(2)$ doublets transform into the mass basis as

$$L^i = \begin{bmatrix} \nu^i_L \\ \ell^i_L \end{bmatrix} \longmapsto \begin{bmatrix} \nu^i_L \\ (U^\ell_{(L)})_{ij}\ell'_{Lj} \end{bmatrix} \tag{2.55}$$

$$Q^i = \begin{bmatrix} u^i_L \\ d^i_L \end{bmatrix} \longmapsto \begin{bmatrix} (U^u_{(L)})_{ij}u'_{Lj} \\ (U^d_{(L)})_{ij}d'_{Lj} \end{bmatrix} = (U^u_{(L)})_{ij}\begin{bmatrix} u'_{Lj} \\ (U^{u\dagger}_{(L)})_{jk}(U^d_{(L)})_{km}d'_{Lm} \end{bmatrix}. \tag{2.56}$$

The doublet Q^i combines up- and down-type quark fields which have *different* unitary transformations.

The gauge covariant derivatives for the doublets are,

$$\mathcal{D}_\mu L^i = \left(\partial_\mu + ig_2 W^a_\mu \frac{\sigma^a}{2} - i\frac{g_1}{2}B_\mu\right)L^i, \tag{2.57}$$

$$\mathcal{D}_\mu Q^i = \left(\partial_\mu + igA^g_\mu T^g + ig_2 W^a_\mu \frac{\sigma^a}{2} + i\frac{g_1}{6}B_\mu\right)Q^i. \tag{2.58}$$

The hypercharge assignments are $Y = -1/2$ for L^i and $Y = 1/6$ for Q^i.

After the introduction of the Higgs doublet and the spontaneous symmetry breaking, the covariant derivatives may be re-expressed in terms of the physical $W^\pm$, Z^0, and photon fields

$$\begin{aligned}\mathcal{D}_\mu L^i = \Bigg(\partial_\mu &+ i\frac{g_2}{\sqrt{2}}\left(W^+_\mu\frac{\sigma^+}{2} + W^-_\mu\frac{\sigma^-}{2}\right) \\ &+ i\sqrt{g_1^2+g_2^2}\left(\frac{\sigma^3}{2} - \sin^2\theta_W\mathcal{Q}\right)Z_\mu + ie\mathcal{Q}A_\mu\Bigg)L^i,\end{aligned} \tag{2.59}$$

$$\begin{aligned}\mathcal{D}_\mu Q^i = \Bigg(\partial_\mu &+ igA^g_\mu T^g + i\frac{g_2}{\sqrt{2}}\left(W^+_\mu\frac{\sigma^+}{2} + W^-_\mu\frac{\sigma^-}{2}\right) \\ &+ i\sqrt{g_1^2+g_2^2}\left(\frac{\sigma^3}{2} - \sin^2\theta_W\mathcal{Q}\right)Z_\mu + ie\mathcal{Q}A_\mu\Bigg)Q^i.\end{aligned} \tag{2.60}$$

Where the electric charge generator

$$\mathcal{Q} = \frac{\sigma^3}{2} + Y = \begin{bmatrix} 1+Y & 0 \\ 0 & Y-1 \end{bmatrix} \tag{2.61}$$

the electric coupling strength

$$e \equiv g_2 \sin\theta_W, \tag{2.62}$$

and

$$\sigma^{\pm} = \sigma^1 \pm i\sigma^2 = \begin{bmatrix} 0 & 1 \pm 1 \\ 1 \mp 1 & 0 \end{bmatrix} \tag{2.63}$$

are all introduced.

The kinetic terms, $i\overline{\psi}\gamma^\mu \mathcal{D}_\mu \psi$, give interaction terms for the quark and lepton fields with the physical boson fields. Expanding $i\overline{Q^i}\gamma^\mu \mathcal{D}_\mu Q^i$ yields the coupling term for the left-handed quark states with the charged bosons,

$$i\overline{Q^i}\gamma^\mu \frac{ig_2}{\sqrt{2}} \left(W_\mu^+ \frac{\sigma^+}{2} + W_\mu^- \frac{\sigma^-}{2} \right) Q^i, \tag{2.64}$$

and with the neutral bosons,

$$i\overline{Q^i}\gamma^\mu i\sqrt{g_1^2 + g_2^2} \left(\frac{\sigma^3}{2} - s_W^2 \mathcal{Q} \right) Z_\mu Q^i + i\overline{Q^i}\gamma^\mu ie\mathcal{Q}A_\mu Q^i, \tag{2.65}$$

amongst others. The shorthand notation $s_W \equiv \sin\theta_W$ is introduced for brevity. The coupling terms with the charged bosons (Eqn 2.64), in the mass basis in element notation (Eqn. 2.56) is,

$$\begin{aligned} & i\overline{Q^i}\gamma^\mu \frac{ig_2}{\sqrt{2}} \left(W_\mu^+ \frac{\sigma^+}{2} \right) Q^i + \text{h.c.} \\ &= \frac{-g_2}{\sqrt{2}} W_\mu^+ \left[\overline{u'_{Lp}} \quad \overline{(U^{u\dagger}_{(L)} U^d_{(L)})_{pq} d'_{Lq}} \right] (U^u_{(L)})^*_{pi} (U^u_{(L)})_{ij} \begin{bmatrix} 0 & \gamma^\mu \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u'_{Lj} \\ (U^{u\dagger}_{(L)} U^d_{(L)})_{jk} d'_{Lk} \end{bmatrix} + \text{h.c.} \\ &= \frac{-g_2}{\sqrt{2}} W_\mu^+ (U^{u\dagger}_{(L)} U^d_{(L)})_{jk} \overline{u'_{Lj}} \gamma^\mu d'_{Lk} + \text{h.c.} \end{aligned} \tag{2.66, 2.67}$$

The factor

$$(U^{u\dagger}_{(L)} U^d_{(L)})_{jk} \equiv V_{jk} \tag{2.68}$$

is the jk^{th} element of the Cabibbo–Kobayashi–Maskawa (CKM) matrix [13]. In matrix notation,

$$U^{u\dagger}_{(L)} U^d_{(L)} \equiv \boldsymbol{V}_{\text{CKM}} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix}. \tag{2.69}$$

This matrix is where flavour structure enters into the theory. The CKM matrix is unitary (by construction) in the SM and has four degrees of freedom which are independent of the choice of phase of the quark fields[5]. The matrix has a hierarchical structure which

[5]A complex 3×3 unitary matrix has 9 degrees of freedom, however 5 phases may be absorbed by redefinition of the lepton fields. Supposing that there were more than 3 quark families, the 3×3 unitary matrix would not be unitary, therefore testing the unitarity is an indirect way to search for physics BSM.

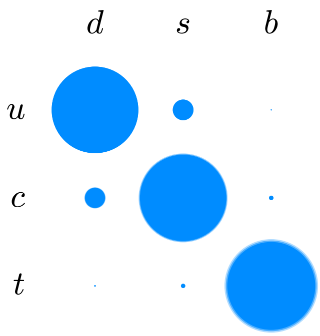


Figure 2.2: Values from [4]. A visualisation of the hierarchical structure of the CKM matrix. The blob radius is proportional to the matrix element size.

is visualised in Fig. 2.2. Mathematically, the structure is illustrated by the Wolfenstein parameterisation [14],

$$\boldsymbol{V}_{\text{CKM}} = \begin{bmatrix} 1-\frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{bmatrix} + O(\lambda^4). \tag{2.70}$$

The parameter of expansion,

$$\lambda = V_{us} \approx 0.22. \tag{2.71}$$

The complex part of the CKM matrix is the source of violation of *CP* symmetry in the weak interaction. Significant experimental effort is devoted to measuring elements of the CKM matrix and checking that it is unitary. The latest combinations of such measurements are available in Refs [15].

On the other hand the neutral coupling terms (Eqn. 2.65),

$$i\overline{Q^i}\gamma^\mu i\sqrt{g_1^2+g_2^2}\left(\frac{\sigma^3}{2} - s_W^2\mathcal{Q}\right)Z^\mu Q^i + i\overline{Q^i}\gamma^\mu ie\mathcal{Q}A_\mu Q^i \tag{2.72}$$

$$\begin{aligned} &= -\sqrt{g_1^2+g_2^2}\,Z_\mu \begin{bmatrix}\overline{u'_{Lp}} & \overline{V_{pq}d'_{Lq}}\end{bmatrix} (U^u_{(L)})^*_{pi}(U^u_{(L)})_{ij} \begin{bmatrix} (\frac{1}{2}-\frac{2}{3}s_W^2)\gamma^\mu & 0 \\ 0 & (-\frac{1}{2}+\frac{1}{3}s_W^2)\gamma^\mu \end{bmatrix} \begin{bmatrix} u'_{Lj} \\ V_{jk}d'_{Lk} \end{bmatrix} \\ &\quad - eA_\mu \begin{bmatrix}\overline{u'_{Lp}} & \overline{V_{pq}d'_{Lq}}\end{bmatrix} (U^u_{(L)})^*_{pi}(U^u_{(L)})_{ij} \begin{bmatrix} \frac{2}{3}\gamma^\mu & 0 \\ 0 & -\frac{1}{3}\gamma^\mu \end{bmatrix} \begin{bmatrix} u'_{Lj} \\ V_{jk}d'_{Lk} \end{bmatrix} \\ &= -\sqrt{g_1^2+g_2^2}\,Z_\mu \left(\left(\frac{1}{2}-\frac{2}{3}s_W^2\right)\overline{u'_{Lj}}\gamma^\mu u'_{Lj} + \left(-\frac{1}{2}+\frac{1}{3}s_W^2\right)\overline{d'_{Lk}}\gamma^\mu d'_{Lk}\right) \\ &\quad - eA_\mu\left(\frac{2}{3}\overline{u'_{Lj}}\gamma^\mu u'_{Lj} - \frac{1}{3}\overline{d'_{Lk}}\gamma^\mu d'_{Lk}\right), \end{aligned} \tag{2.73}$$

do not mix up- and down-type quarks[6]. Therefore *FCNCs do not occur at tree level in the SM*. All flavour transitions must proceed via the $W^\pm$ bosons whose flavour-changing

[6]The CKM factors in the down-type neutral coupling terms ($\overline{V}\delta V\overline{d}\gamma d$) dissappear in the last line of Eqn. 2.73 since

$$\overline{V_{pq}}\delta_{pj}V_{jk} = (U^d_{(L)})^*_{qm}(U^{u\dagger}_{(L)})^*_{mp}\delta_{pj}(U^{u\dagger}_{(L)})_{jn}(U^d_{(L)})_{nk} = \delta_{qk}$$

quark couplings arise from the non-diagonal Yukawa couplings y^q_{ij}. In other words, the mass eigenstates and the weak eigenstates are *misaligned in flavour space*.

Since FCNC processes must proceed via loop-level diagrams in the SM, they are an excellent place to search for possible BSM effects due to the small SM contribution. Often, BSM theories require additional particles and interactions that have not yet been observed. The contribution from diagrams with virtual BSM particles can be at a visible level or can interfere with the SM contributions to give measurable effects. Much theoretical and experimental effort is therefore devoted to improving and testing predictions for FCNC processes.

2.1.5 Qualitative description of quantum chromodynamics

The strong interaction, described by the theory of quantum chromodynamics (QCD), does not contribute to the flavour-changing structure of the SM. However the $b \to s\ell^+\ell^-$ quark-level process receives higher order corrections from diagrams with gluons. In addition the quark-level process is contained within a physical meson system which is described by QCD. It is therefore relevant to qualitatively describe QCD, which will assist in motivating the effective theory approach in Section 2.2 and the application to $B^0 \to K_1^*(892)^0\ell^+\ell^-$, described in Section 2.3.

The theory of QCD describes the gluon field interaction with the quarks. In QCD, the running coupling of the theory becomes stronger at lower energies (long distances). The converse of this, that the coupling becomes weaker at high-energy scales is known as the *asymptotic freedom* of QCD. Analogous to the fine structure constant of electrodynamics, the definition

$$\alpha_s = \frac{g^2}{4\pi}, \tag{2.74}$$

where g is the coupling coefficient for the fields (introduced in Eqn 2.58), is usually made. The solution to the renormalisation group equation is

$$\alpha_s(\varrho) = \frac{6\pi}{(33 - 2n_f)\log(\varrho/\Lambda_{\text{QCD}})}, \tag{2.75}$$

where n_f is the number of quark flavours. This expression governs the behaviour of α_s at some scale ϱ. The scale of the theory, Λ_{QCD}, is defined as the scale at which α_s becomes large as ϱ^2 is decreased. Experimental measurements find a value of $\Lambda_{\text{QCD}} \approx 200\,\text{MeV}/c^2$. This behaviour means that perturbative calculations become difficult at low energies (long distances).

2.2 Motivation for an effective field theory approach

The majority of particle physics is concerned with testing the SM, and searching for new physics (NP). Often, searches for NP are performed within the context of a given BSM theory. Such an approach is 'top-down' in nature: the measurement of a physical quantity is motivated by a given theory which predicts some measurable difference with respect to the SM prediction. For the $b \to s\ell^+\ell^-$ transition, and more generally, FCNC processes, this approach is somewhat reversed. The 'bottom-up' approach used here is an effective field theory description. This approach has two main advantages: model-independence and the separation of scales. The effective field theory framework makes no model-dependent assumptions about the structure of NP and the general form of the couplings can be included explicitly. The second advantage that low-energy effects, which are difficult to reliably predict due to the nonperturbative nature of QCD, are separated from the high-energy effects that are sensitive to contributions from NP. Since b-quarks cannot be created at energy scales smaller than their mass, effects below this scale are not normally relevant to the system of the energetic weak decay [8]. Therefore the b-quark mass provides the scale of the system:

$$\varrho = m_b \approx 5\,\mathrm{GeV}/c^2 \gg \Lambda_{\mathrm{QCD}}. \tag{2.76}$$

Such an approach is not without disadvantages. Whilst the low-energy effects in the *quark-level* process are separated out, the description of the physical hadronic system introduces additional low-energy QCD effects. The hadronic system is parameterised by so-called *hadronic form factors* which are difficult to calculate. Another issue arises in FCNC decays of the b-quark to light quarks ($b \to s$ or $b \to d$). This issue is regarding treatment of corrections due to loops involving a c-quark. Both of these disadvantages will be discussed in the following sections.

2.2.1 Operator-product expansions

The effective field theory approach used to describe $b \to s\ell^+\ell^-$ processes is a specific example of an operator-product expansion (OPE). The OPE technique has often been used during the development of the quantum field theory description of particle physics. The canonical example, is in the description of nuclear beta decay before the discovery of the $W^\pm$ bosons where a four-point operator was used with a coupling strength G_F. This section provides a brief description of the technique before the details of the OPE for $b \to s\ell^+\ell^-$ processes is presented in Section 2.3.

The Lagrangian density of the SM and any BSM theories can be written

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \Delta\mathcal{L}_{\mathrm{NP}} \tag{2.77}$$

where $\Delta\mathcal{L}_{\mathrm{NP}}$ refers to extra terms in the theory due to NP contributions.

An OPE replaces this with the effective Lagrangian,

$$\mathcal{L}_{\mathrm{eff}} = \sum_i C_i(\varrho)\mathcal{O}_i. \tag{2.78}$$

In performing this replacement, effects of interactions above the scale ϱ (high-energy, short-distance) are contained within the *Wilson coefficients*, $C_i = C_i(\varrho)$. These interactions are often referred to as the 'heavy degrees of freedom'. In the case where $\varrho = m_b$, the SM $W^{\pm}$, Z^0, and t-quark all fall under this definition of 'heavy' and are absorbed into the C_i.

The low-energy (long-distance) effects below the scale ϱ are absorbed into the local operators $\mathcal{O}_i = \mathcal{O}_i(\varrho)$. The effective operators act on the low-energy fields with some generic Lorentz structure. For example one operator will describe a scalar interaction. The corresponding scalar Wilson coefficient represents the size of the scalar contribution which is predicted in the SM and within BSM theory frameworks. A deviation from the SM prediction for the corresponding Wilson coefficient would imply a NP scalar interaction. Testing theoretical predictions for the Wilson coefficients with experiment therefore provides sensitivity to broad classes of NP particles, independent of any model. The Wilson coefficients are often separated into predictions within the SM, and some additive modification due to NP contributions,

$$C_i \equiv C_i^{\mathrm{SM}} + C_i^{\mathrm{NP}}. \tag{2.79}$$

More convenient for the calculation of decay amplitudes, is the equivalent effective Hamiltonian

$$\langle f|\mathcal{H}_{\mathrm{eff}}|i\rangle = \sum_i C_i(\varrho)\langle f|\mathcal{O}_i|i\rangle\Big|_{\varrho}, \tag{2.80}$$

for some transition from initial state $|i\rangle$ to a final state $|f\rangle$, evaluated at the scale, ϱ. The decay amplitudes contain the Wilson coefficients which would ideally be compared directly to data. However the quark-level decay occurs within a hadronic system.

Unlike the b-quark mass, the length scale of the *hadronic system* is on the order of Λ_{QCD} and therefore described by low-energy QCD which is theoretically challenging. Hadronic form factors are introduced to parameterise the decay amplitudes for the meson-level transition. Typically, form factors are calculated with nonperturbative techniques. The hadronic uncertainty associated with these form factors is typically the dominant source of theoretical uncertainty in predictions for observables. There is therefore a preference to measure observables, which are typically ratios or differences of quantities, where the form-factor uncertainty cancels, at least to leading order. In $b\to s\ell^+\ell^-$ decay phenomenology, so-called 'theoretically clean' observables are constructed in such a way as to maximise sensitivity to NP effects through the Wilson coefficients and minimise dependence on the form factors.

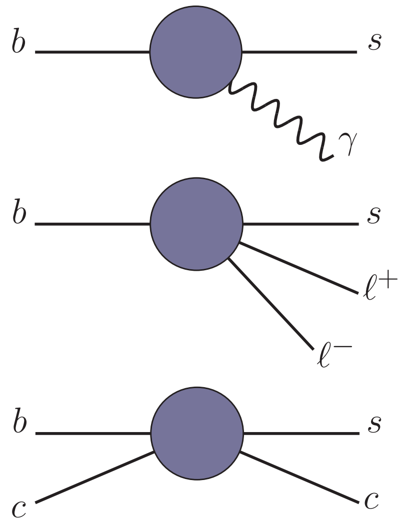


Figure 2.3: Schematic diagrams for operators $\mathcal{O}_7$ (top), $\mathcal{O}_{9,10}$ (centre), and $\mathcal{O}_{1c,2c}$ (bottom). The effective operator $\mathcal{O}_7$ relates to diagrams mediated by a radiated photon, the operator $\mathcal{O}_9$ relates to diagrams mediated by a vector current, and the operator $\mathcal{O}_{10}$ to diagrams mediated by an axial current. The operators are defined in Equations 2.82, 2.83, and 2.84 respectively. The effective operators $\mathcal{O}_{1c,2c}$ relate to diagrams with internal c-quarks.

2.3 Application of the effective field theory approach to $b \to s\ell^+\ell^-$ and the hadronic description of $B^0 \to K_1^*(892)^0\ell^+\ell^-$ decay amplitudes

The $b \to s\ell^+\ell^-$ processes can be described by the following effective Hamiltonian [8, 16],

$$\mathcal{H}_{\text{eff}} = -\frac{4G_{\text{F}}}{\sqrt{2}}\frac{e^2}{16\pi^2}V_{tb}V_{ts}^* \sum_{i=1,\ldots 10,\text{S,P}} (C_i\mathcal{O}_i + C_i'\mathcal{O}_i'). \tag{2.81}$$

In this expression, G_{F} and e are Fermi's constant and the electromagnetic coupling strength respectively. The operators expected to be dominant and their chiral partners are [16–18],

$$\mathcal{O}_7 = \frac{e}{g^2}m_b(\bar{s}\sigma_{\mu\nu}P_R b)F^{\mu\nu}; \qquad \mathcal{O}_7' = \frac{e}{g^2}m_b(\bar{s}\sigma_{\mu\nu}P_L b)F^{\mu\nu}; \tag{2.82}$$

$$\mathcal{O}_9 = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu\ell); \qquad \mathcal{O}_9' = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_R b)(\bar{\ell}\gamma^\mu\ell); \tag{2.83}$$

$$\mathcal{O}_{10} = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu\gamma_5\ell); \qquad \mathcal{O}_{10}' = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_R b)(\bar{\ell}\gamma^\mu\gamma_5\ell); \tag{2.84}$$

where $P_{L,R}$ are the chrial projection operators (introduced in Eqns 2.2 and 2.3). The chiral partner operators receiving the primes are such that $\mathcal{O}_i'$ is zero or suppressed in the SM. The factor $F^{\mu\nu}$ is the electromagnetic field strength tensor, $\sigma_{\mu\nu} \equiv i\,[\gamma_\mu, \gamma_\nu]$, and m_b is the running mass of the b-quark.

Equation 2.82 is the radiative photon operator and relates to diagrams mediated via a photon which decays into the dilepton pair. Equations 2.83 and 2.84 are the vector and axial vector operators respectively. Schematic diagrams of $\mathcal{O}_7$ and $\mathcal{O}_{9,10}$ are shown in Fig. 2.3.

The remaining operators ($\mathcal{O}_{1-6,8,\text{S,P}}$) are either heavily suppressed, or the corresponding coefficient is already well constrained by experiment [16]. For example the scalar coefficient

C_S is constrained by the measurement of the $B^0_s \to \mu^+\mu^-$ decay process [19]. Two further operators (which will be mentioned in Section 2.4) are those where the diagrams have internal c-quarks

$$\mathcal{O}_{1c} = (\bar{s}\gamma_\mu T^g P_L c)(\bar{c}\gamma^\mu T^g P_L b); \qquad \mathcal{O}'_{1c} = (\bar{s}\gamma_\mu T^g P_R c)(\bar{c}\gamma^\mu T^g P_R b); \tag{2.85}$$

$$\mathcal{O}_{2c} = (\bar{s}\gamma_\mu P_L c)(\bar{c}\gamma^\mu P_L b); \qquad \mathcal{O}'_{2c} = (\bar{s}\gamma_\mu P_R c)(\bar{c}\gamma^\mu P_R b). \tag{2.86}$$

These are the so-called 'four-quark' operators and are also shown schematically in Fig 2.3. There are also $\mathcal{O}_{1u,2u}$ which are obtained making the replacement $c \to u$ in the above. The remaining $\mathcal{O}_{2-6,8,\mathrm{S,P}}$ are given full in, for example Ref. [18], and will be mentioned only briefly. The operators $\mathcal{O}_{3-6}$ are QCD penguin operators; the operator $\mathcal{O}_8$ is the radiative gluon operator; and $\mathcal{O}_{\mathrm{S,P}}$ correspond to diagrams with an internal scalar or pseudo-scalar particle and are extremely suppressed.

In the SM there are no right-handed interactions for the vector and axial vector currents, so $C'^{\mathrm{SM}}_{9,10} = 0$, and the SM contributions to C'_7 are helicity-suppressed: $C'^{\mathrm{SM}}_7 = (m_s/m_b)C^{\mathrm{SM}}_7$.

There is an additional technicality where the Wilson coefficients are more conveniently expressed as *effective Wilson coefficients*,

$$C^{\mathrm{eff}}_7 = \frac{4\pi}{\alpha_s} C_7 - \frac{1}{3} C_3 - \frac{4}{9} C_4 - \frac{20}{3} C_5 - \frac{80}{9} C_6 \tag{2.87}$$

$$C^{\mathrm{eff}}_9 = \frac{4\pi}{\alpha_s} C_9 + Y(q^2) \tag{2.88}$$

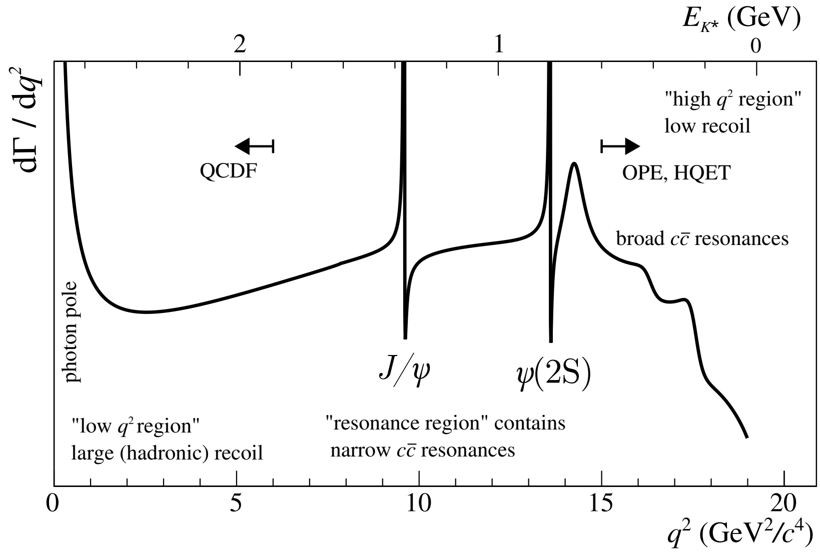
$$C^{\mathrm{eff}}_{10} = \frac{4\pi}{\alpha_s} C_{10} \tag{2.89}$$

$$C'^{\mathrm{eff}}_{7,9,10} = \frac{4\pi}{\alpha_s} C'_{7,9,10} \tag{2.90}$$

since they frequently appear in amplitudes with such combinations of the other coefficients [16][7]. The function $Y(q^2)$ is a suppressed combination of further Wilson coefficients that is given in Refs [16, 18], for example. Given present measurements, that will be discussed further below, it is important to note that $Y(q^2)$ contains the coefficients $C_{1c,2c}$. An equivalent statement is that C^{eff}_9 receives corrections from *charm-loop* diagrams, with $\mathcal{O}_{1c,2c}$. Reference [16] gives predictions according to the SM of $C^{\mathrm{eff,\,SM}}_7 \approx -0.3$, $C'^{\mathrm{eff,\,SM}}_7 \approx -0.006$, $C^{\mathrm{eff}}_9 - Y(q^2) \approx 4.2$, and $C^{\mathrm{eff,\,SM}}_{10} \approx -4.1$ which are calculated at the scale $\varrho = m_b$. No uncertainties are provided for these SM predictions. However, the Wilson coefficients are not observable quantities, and the dominant uncertainty contributing to observables comes from form factors.

[7] In fact, the C_i are evaluated at the scale $\varrho \approx M_W$ and evolved down to the scale of the b-quark ($\varrho = m_b$). This procedure introduces a mixing between the C_i which is why the C^{eff}_i are written as above and why one can consider effect of the reduced set of operators $C^{\mathrm{eff}}_{7,9,10}$. This technical point is described in, for example, Section 2.2 of Ref. [18]. For the purposes of this discussion the simple statement given in Ref. [16] is sufficient.

Figure 2.4: An illustrative sketch of the squared dilepton invariant mass, q^2, (lower scale) or hadron energy, E_{K^*}, (upper scale) distribution for $B^0 \to K^{(*)}\ell^+\ell^-$ decays, including the charmonia resonances. The various regions depicted are, from left-to-right: the photon pole region containing the divergence as $q^2 \to 0$; the low q^2 region where the $K^{(*)}$ has relatively high-energy ($E_{K^*} \gg \Lambda_{\rm QCD}$); the resonance region where $c\bar{c}$ effects dominate; and finally, the high q^2 region where the $K^{(*)}$ object has relatively low-energy ($E_{K^*} \lesssim \Lambda_{\rm QCD}$) and where $q^2 = O(m_b^2)$. The figure is adapted from Ref. [22].



2.3.1 The $B \to K^*$ form factors and regions of squared dilepton invariant mass

The meson-level decay amplitudes depend on the effective Wilson coefficients and the $B \to K^*$ form factors. The latter parameterise the hadronic system and depend on the energy shared between the dilepton object and the K^*. It is convention to use q^2 as the variable describing this energy share. The amplitudes are parameterised in terms of a set of seven form factors [16] which are often referred to in the literature as the *full form factors* and are typically denoted $\{V(q^2),\, A_{0,1,2}(q^2),\, T_{1,2,3}(q^2)\}$. As discussed below, under certain conditions it is possible to reduce this set to two *soft form factors* [20, 21], typically denoted $\{\xi_\parallel(q^2),\, \xi_\perp(q^2)\}$.

The treatment of the form factors differs in regions of q^2. An illustrative sketch of the q^2 distribution for $B^0 \to K^{(*)}\ell^+\ell^-$ and the various regions is given in Fig. 2.4.

Below the threshold $q^2 \approx (2m_c)^2 \approx 7\,\text{GeV}^2/c^4$ is the *low-q^2* region. This region is often referred to as the region of 'large hadronic recoil' since the energy of the K^* is relatively large, $E_{K^*} \gg \Lambda_{\rm QCD}$. In the low-$q^2$ region the assumption of QCD factorisation (QCDF) applies. This assumption is that gluons exchanged between the quarks in the meson do not interact with the decay process and these two effects are separable. An expansion in $\Lambda_{\rm QCD}/E_{K^*}$ may be employed to parameterise the form factors. The soft form factors are derived from the leading order term in this expansion and are given in terms of the full form factors in, for example, Ref. [21]. So called 'power corrections' to the expansion are the leading source of systematic uncertainty for predictions at low q^2. There are also so-called 'non-factorisable corrections' which correct the assumption of QCDF. In the low-q^2 region, form factors may be calculated using light cone sum rules (LCSR) techniques [22]. At the time of writing, the most recent calculation of the full form factors using LCSR is given in Ref. [23] which is an update to Ref. [24].

The threshold $q^2 \approx (2m_c)^2 \approx 7\,\mathrm{GeV}^2/c^4$ arises where $c\bar{c}$ pair production effects are conventionally expected to become large. Below $q^2 \approx 1\,\mathrm{GeV}^2/c^4$, the effects of light resonances become large and a divergence appears in the theory as $q^2 \to 0$. This divergence is the so-called *photon pole*. In the vicinity of the photon pole, $\mathcal{O}_7$ dominates [22]. There is therefore a region $1 \lesssim q^2 \lesssim 6\,\mathrm{GeV}^2/c^4$ which is far from the photon pole and the $c\bar{c}$ threshold and is 'theoretically clean'. For $q^2 \in [1,6]\,\mathrm{GeV}^2/c^4$, the interference between $\mathcal{O}_7$ and $\mathcal{O}_9$ is large and provides good sensitivity to C_9 [22].

The charmonia resonances of the J/ψ and $\psi(2S)$ dominate above the $c\bar{c}$ threshold until $q^2 \approx 15\,\mathrm{GeV}^2/c^4$. The $c\bar{c}$ loop diagrams and the associated long-distance QCD effects dominate in this region by around two orders of magnitude [16]. This *resonance region* is therefore very challenging and no precise theoretical predictions are available.

The *high-q^2* region is above $q^2 \approx 15\,\mathrm{GeV}^2/c^4$. Analogously with low q^2, this region is often referred to in terms of the K^* energy as 'low hadronic recoil'. In this region, $q^2 = O(m_b^2)$ and $E_{K^*} < \Lambda_{\mathrm{QCD}}$. High q^2 is treated with heavy quark effective theory (HQET) expanding in powers of $\Lambda_{\mathrm{QCD}}/m_b$. Additionally, QCD calculations may be performed on a lattice here, with the assumption that the K^* is a narrow resonance and is at rest in the rest frame of the B meson. Lattice QCD calculations use the same seven full form factors that describe low q^2. The most recent lattice predictions are available in Ref. [25].

2.3.2 The $B^0 \to K_1^*(892)^0\ell^+\ell^-$ decay amplitudes

To give an explicit example of a meson-level decay amplitude, the leading order expression for the longitudinally polarised $B^0 \to K_1^*(892)^0\ell^+\ell^-$ amplitude from Ref. [16],

$$
\begin{aligned}
A_{1L(R)0} = -\frac{N}{2m_{K^*}\sqrt{q^2}}\Bigg\{ &\left[(C_9^{\mathrm{eff}} - C_9^{\prime\mathrm{eff}}) \mp (C_{10}^{\mathrm{eff}} - C_{10}^{\prime\mathrm{eff}})\right] \\
&\times \left[(m_B^2 - m_{K^*}^2 - q^2)(m_B + m_{K^*})A_1(q^2) - \lambda\frac{A_2(q^2)}{m_B + m_{K^*}}\right] \\
&+ 2m_b(C_7^{\mathrm{eff}} - C_7^{\prime\mathrm{eff}})\left[(m_B^2 + 3m_{K^*}^2 - q^2)T_2(q^2) - \frac{\lambda}{m_B^2 - m_{K^*}^2}T_3(q^2)\right]\Bigg\},
\end{aligned} \tag{2.91}
$$

contains the Wilson coefficients and the full form factors. The right-handed amplitude is formed by taking the sum where $\mp$ appears. Using the soft form factors, Ref. [21] shows that this simplifies to

$$A_{1L(R)0} = -\frac{N(m_B^2 - q^2)^2}{2m_{K^*} m_B \sqrt{q^2}} \left[(C_9^{\text{eff}} - C_9^{\prime\text{eff}}) \mp (C_{10} - C_{10}^{\prime}) + 2\frac{m_b}{m_B}(C_7^{\text{eff}} - C_7^{\prime\text{eff}}) \right] \xi_{\parallel}(q^2). \quad (2.92)$$

The normalisation, $N \propto V_{tb}V_{ts}^{*}$, and the factor λ are consistently defined between the two expressions. As will be shown, these expressions (and similar) appear in the angular differential decay rate as bilinear combinations. Various bilinear combinations of these expressions are observable quantities which give constraints on the size of the Wilson coefficients. Relevant examples of such observables are the total and differential branching fractions, of interest in this thesis.

2.3.3 The effect of a scalar component

Sensitivity to $C_{7,9,10}^{\text{eff}}$ is enhanced in the ideal case of the P-wave decay $B^0 \to K_1^*(892)^0\ell^+\ell^-$. Experimentally, the K^* is selected with a window of $m_{K\pi}$. In previous experimental measurements [26–29] a requirement was made that $m_{K\pi} \in [796, 996]\,\text{MeV}/c^2$. This window is equivalent to $\pm 100\,\text{MeV}/c^2$ around the nominal mass of the $K_1^*(892)^0$, $m_{892} = (895.81 \pm 0.19)\,\text{MeV}/c^2$ [4].

However such an approach does not exclude contributions from other K^* states. In particular, the contribution from the $K\pi$ system in a scalar, $J = 0$ state (S-wave) is known to be problematic, in particular at low q^2 where it may be sizeable [30]. Assuming massless leptons, the S-wave component adds two complex decay amplitudes to the six P-wave amplitudes. The extra amplitudes interfere and add extra nuisance terms to the angular distribution. This will be shown in more detail in Section 2.6. All of the P-wave angular terms effectively obtain a factor of $(1 - F_{\text{S}})$, where F_{S} is the fractional size of the S-wave component. This is also true for the differential branching fraction, as will be shown in Section 2.6. This is often referred to *S-wave pollution* in the literature since it is difficult to estimate. The S-wave contribution is the leading source of systematic uncertainty in the experimental analyses of Refs [26–28].

The analysis described in this thesis makes the first determination of the size of this S-wave contribution and the first measurement of the branching fraction of the P-wave decay where this contamination is not treated as a systematic uncertainty.

2.3.4 Summary

To summarise this section, the quark-level $b \to s\ell^+\ell^-$ transition is described in a model-independent way in terms of effective operators and corresponding Wilson coefficients. This OPE relies on the separation of scales $m_b \gg \Lambda_{\text{QCD}}$. The dominant coefficients are $C_{7,9,10}^{\text{eff}}$ which receive corrections from other coefficients. The hadronic structure is parameterised in terms of seven (two) full (soft) form factors. The decay amplitudes contain the Wilson coefficients and the form factors. The form factors are determined using different methods for different regions of q^2 (or E_{K^*}). At low q^2 it is possible to expand in $\Lambda_{\rm QCD}/E_{K^*}$ and use the soft (or the full) form factors. At high q^2 the expansion is performed in powers of $\Lambda_{\rm QCD}/m_b$. At low (high) q^2, LCSR (lattice QCD) techniques are used to calculate the full form factors. Above the $c\bar{c}$ threshold and below $q^2 \approx 15\,{\rm GeV}^2/c^4$, the narrow charmonia resonances are overwhelmingly dominant and low-energy QCD effectively prohibits predictions in this region. The dominant uncertainties in predictions of the amplitudes are the corrections to the hadronic form factors, $1/m_b$ power corrections (at low q^2), and the effect of $c\bar{c}$ contributions (at high q^2). There are six complex decay amplitudes in the case of $B^0 \to K_1^*(892)^0\ell^+\ell^-$. The presence of a S-wave component to the K^* adds two additional amplitudes which dilute the NP sensitivity to the Wilson coefficients.

2.4 The current status of experimental measurements and the phenomenology of $b \to s\ell^+\ell^-$

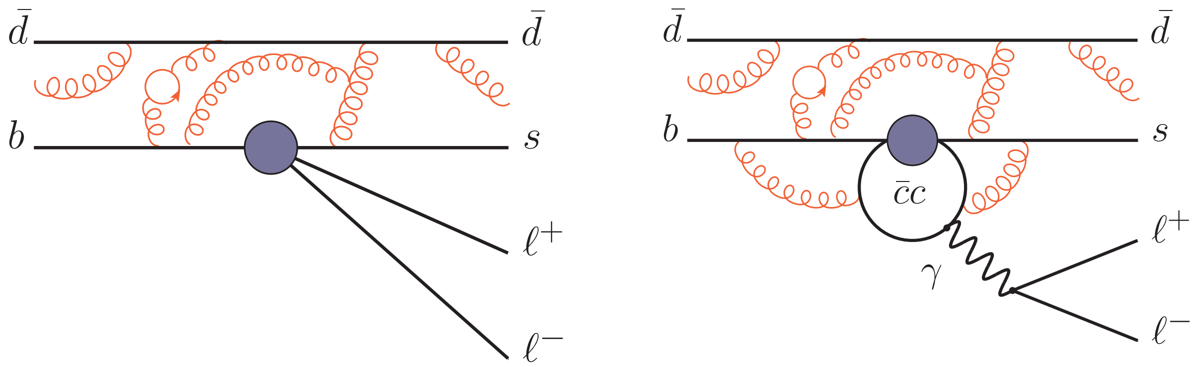
With data corresponding to $1\,{\rm fb}^{-1}$, measurements of two sets of angular observables for $B^0 \to K_1^*(892)^0\mu^+\mu^-$ have been performed at LHCb [26, 27]. The measurement of a particular set of observables, which will be defined in Section 2.6 (Eqn. 2.140), found a local deviation of 3.7σ from the SM prediction in Ref. [31]. During the preparation of this thesis this deviation has been confirmed with data corresponding to $3\,{\rm fb}^{-1}$ [29]. Additionally, the branching fractions of the channels $B_s^0 \to \phi(1020)\mu^+\mu^-$, $B \to K\mu^+\mu^-$, and $B^+ \to K^{*+}\mu^+\mu^-$, which are all $b \to s\ell^+\ell^-$ transitions, are measured at LHCb. The former is found to be 3.3σ below the SM prediction for $q^2 \in [1, 6]\,{\rm GeV}^2/c^4$ [32][8], the latter two are found to be slightly lower than theoretical predictions, although consistent within uncertainties [28]. The process $B_s^0 \to \phi(1020)\mu^+\mu^-$ has the same leading order SM diagrams as shown in Fig. 1.1 but with the spectator d-quark replaced by an s-quark. These decays are sensitive to analogous form factor effects as those described above in Section 2.3.1, although the exact form factor parameterisations may be different.

Global fits to the above experimental data, which extract the Wilson coefficients, indicate a tension from the SM in C_9 [33–35]. This tension could be explained by an additional vector boson Z' contributing to these decays [35]. There has been some debate [36–38] about the size of the uncertainty of the soft form factors from Ref. [31], however recent calculations using the full form factors [23] have shown that this uncertainty was only slightly underestimated and the tension remains.

It has also been suggested [39] that the size of the $c\bar{c}$ contribution (equivalently, the size of the $\mathcal{O}_{1c,2c}$ contribution to the function $Y(q^2)$) has been underestimated. Predictions of $C_9^{\rm eff}$ should then receive larger corrections from gluons introduced by the $c\bar{c}$ loop. This

[8]Reference [32] states 3.5σ. However a minor problem was discovered, after submission for publication, with the theory uncertainty which reduces this.

Figure 2.5: A schematic picture of potentially underestimated contributions from operators $\mathcal{O}_{1c,2c}$. The left-hand figure shows $\mathcal{O}_9$. The gluon lines in the left-hand figure represent the long-distance effects that are accounted for by form factors. The right-hand figure shows an example of a diagram with operators $\mathcal{O}_{1c,2s}$. The gluon lines from the quark loop represent the part of the correction to C_9^{eff} which may be underestimated.



is shown schematically in Fig. 2.5. Such an effect is postulated based on a disagreement between descriptions of the resonances in the dimuon spectrum of $B^+ \to K^+\mu^+\mu^-$ decays [40].

If the phenomenological interpretation of $C_9^{\text{NP}} \neq 0$ and a new vector Z' particle are to be believed, there is a natural question of whether the Z' couples universally to leptons [41]. The ratio of branching fractions of the $B^+ \to K^+\mu^+\mu^-$ to $B^+ \to K^+e^+e^-$ decays,

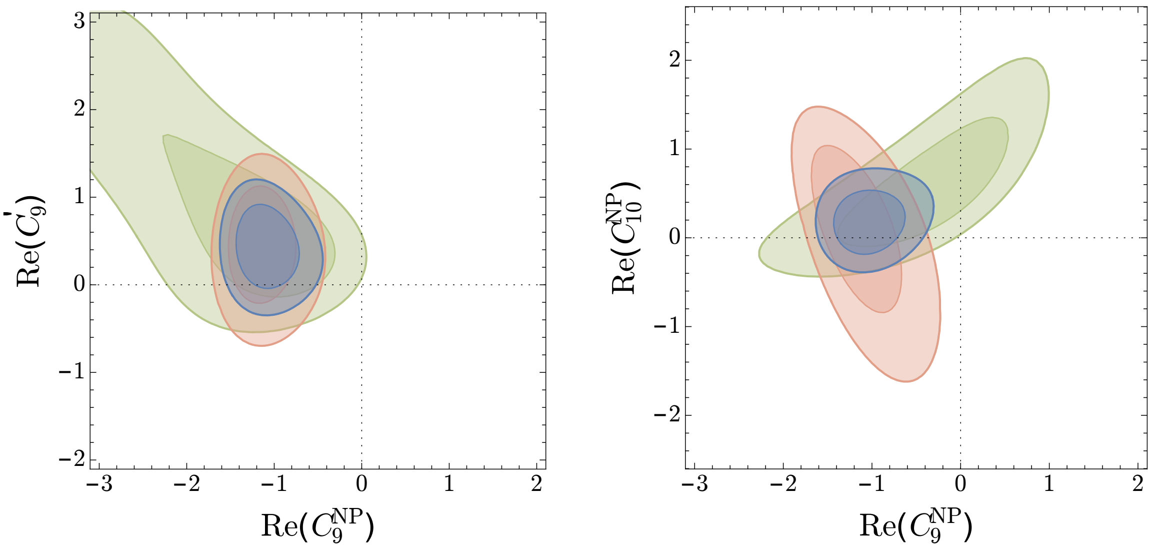
$$R_K \equiv \frac{\mathcal{B}\left[B^\pm \to K^\pm\mu^+\mu^-\right]}{\mathcal{B}\left[B^\pm \to K^\pm e^+e^-\right]}, \tag{2.93}$$

is predicted to be $R_K = 1 + O(m_\mu^2/m_b^2)$ in the SM for $q^2 > 1\,\text{GeV}^2/c^4$ [42]. This ratio was measured by the LHCb collaboration, and found to lie 2.6σ from unity in the region $q^2 \in [1,6]\,\text{GeV}^2/c^4$ [43]. This deviation is not statistically significant, however long-distance QCD uncertainty is not an issue for this observable and a consistent explanation is a Z' with non-universal couplings to leptons. This measurement will be improved with additional data and analyses of R_ϕ and R_{K^*} defined analogously to Eqn. 2.93 are in preparation. If lepton couplings to the proposed Z' are non-universal then different operators and Wilson coefficients should be included for the different leptons. This possibility is explored in Refs [44, 45].

The addition of the modest deviation in R_K contributes to the tension in the global fits [44,46,47]. An alternative NP interpretation is that the deviation is caused by the scalar operator, $\mathcal{O}_S$. A scalar NP contribution is less favoured once other FCNC constraints, particularly from the observation of the decay $B_s^0 \to \mu^+\mu^-$ [19], are taken into account [44].

Figure 2.6 shows some of the most recent global fit results from Ref. [48] which is an update to the analysis in Refs [34, 47]. This fit is performed with the full form factors, and takes 88 measurements of 76 observables as inputs. The inputs are described in

Figure 2.6: The result of a global fit to the Wilson coefficients shown as the $\mathfrak{Re}\left(C_9^{\mathrm{NP}}\right) - \mathfrak{Re}\left(C_9'\right)$ plane (left) and the $\mathfrak{Re}\left(C_9^{\mathrm{NP}}\right) - \mathfrak{Re}\left(C_{10}^{\mathrm{NP}}\right)$ plane (right). The darker and lighter contours correspond to 1σ and 2σ best fit regions respectively. The blue shaded area shows the best fit point which is 3.7σ from the $C_9^{\mathrm{NP}} = 0$ (the SM point by construction). The red (green) area shows the constraint from angular observables (branching fractions). Figures taken from Ref. [48].

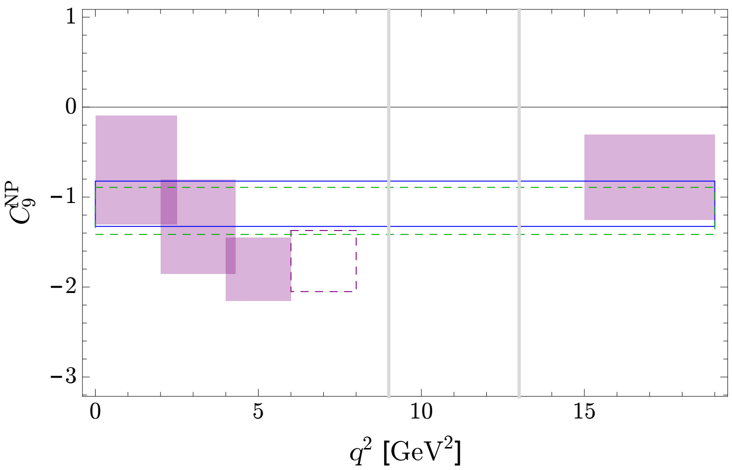


the above references and are not listed here. Of particular note, angular observables from $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ are included. The branching fractions of $B_s^0 \to \phi(1020)\mu^+\mu^-$, $B \to K\mu^+\mu^-$, $B^+ \to K^{*+}\mu^+\mu^-$, and $B_s^0 \to \mu^+\mu^-$ are also inputs[9]. A non-zero C_9^{NP} is preferred at the 3.7σ level. This increases to 4.1σ when the R_K deviation is included. The possibility of an underestimated contribution from the four-quark operators $\mathcal{O}_{1c,2c}$ is not excluded. Figure 2.7 illustrates a hint of a q^2-dependence of the fitted C_9^{NP} that might indicate such an effect. Underestimating the $c\bar{c}$ loop diagrams (*i.e.* $\mathcal{O}_{1c,2c}$ suggested by Ref. [39]) could lead to such an effect however the required shift in the amplitudes is thought to be unfeasibly large [48]. Furthermore an underestimated hadronic effect should be lepton flavour universal, and would not explain the deviation seen in R_K. By contrast, a mistreatment of the *form factors* would be manifest at the edges of the q^2 where the different methods (LCSR and lattice QCD) dominate; this kind of dependence is not seen.

The differential branching fraction which is relevant to this thesis, will add additional information to the global fits.

[9] Although in the case of $B_s^0 \to \phi(1020)\mu^+\mu^-$, the analysis in Ref. [48] predates the publication of the 3.3σ discrepancy in Ref. [32].

Figure 2.7: The possible effect of $c\bar{c}$ loop interference. The solid blue (dashed green) box represents the best fit point of the global fit (a fit only with $B^0 \to K^{*0}\mu^+\mu^-$ observables). Red shaded boxes represent global fits for several bins of q^2 and the resulting value of $C_9^{\rm NP}$. Whilst all are consistent with no q^2 dependence (a flat line), the central values appear to show a trend approaching the J/ψ resonance at $q^2 = 9.58\,\text{GeV}^2/c^4$. The dashed red box is the global fit in the region $q^2 \in [6, 8]\,\text{GeV}^2/c^4$. This region is excluded from the global fit specifically because $c\bar{c}$ effects are expected there. If this were genuinely a q^2 dependence of the fitted $C_9^{\rm NP}$, it would indicate mistreatment of $c\bar{c}$ effects as there should be no q^2 dependence for any $C_i^{\rm NP}$. The size of the $c\bar{c}$ contribution necessary to create such an effect is very large. Figure taken from Ref. [48].



2.5 Definitions of helicity angles, the angular basis

The remainder of this chapter is devoted to a description of the theoretical angular distributions of the $B^0 \to K_J^*\ell^+\ell^-$ decay. Before describing the angular distribution, it is necessary to define the angular basis used. The angular basis is not directly relevant for the differential branching fraction, however the size of the S-wave contribution is extracted using a partial angular analysis. Furthermore the description of the angular efficiency introduces a slight dependence on the full set of angles.

The process $B^0 \to K_1^*(892)^0\ell^+\ell^-$, is a pseudo-scalar to vector, vector transition. It has four degrees of freedom. These are conventionally chosen as three angles, (θ_ℓ, θ_K, ϕ), and q^2. For a general K^* state, that is: $B^0 \to K_J^*\ell^+\ell^-$, the $K\pi$ invariant mass ($m_{K\pi}$) provides information about the resonance spectrum.

The angle θ_K is defined as the angle between the kaon in the rest frame of the $K\pi$ system and the direction of the $K\pi$ object in the rest frame of the B meson,

$$\cos\theta_K = \frac{\vec{p}_K^{\,K\pi} \cdot \vec{p}_{K\pi}^{\,B}}{|\vec{p}_K^{\,K\pi}||\vec{p}_{K\pi}^{\,B}|} = \vec{u}_K^{\,K\pi} \cdot \vec{u}_{K\pi}^{\,B}, \tag{2.94}$$

for both $B^0 \to K^+\pi^-\ell^+\ell^-$ and the CP-conjugated decay ($\overline{B}^0 \to K^-\pi^+\ell^+\ell^-$). The notation $\vec{p}_a^{\,b}$ denotes the vector momentum of particle a in the rest frame of particle b and the notation $\vec{u}_a^{\,b}$ denotes the corresponding unit vector. Figure 2.8 shows this angle diagrammatically.

Figure 2.8: Diagrams showing the angle θ_K. The decay $B^0 \to K^+\pi^-\ell^+\ell^-$ (left) and the decay $\overline{B}^0 \to K^-\pi^+\ell^+\ell^-$ (right) have the same definition, via Eqn. 2.94. Both diagrams are shown in the rest frame of the $K\pi$ system.

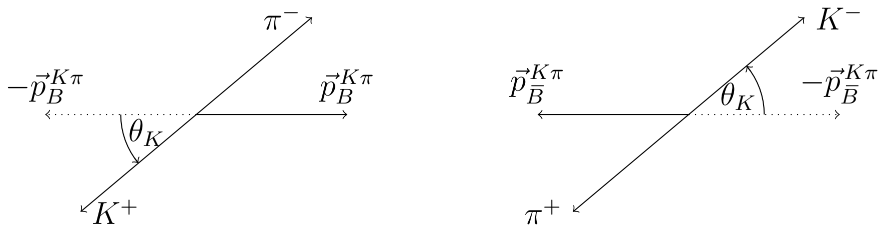
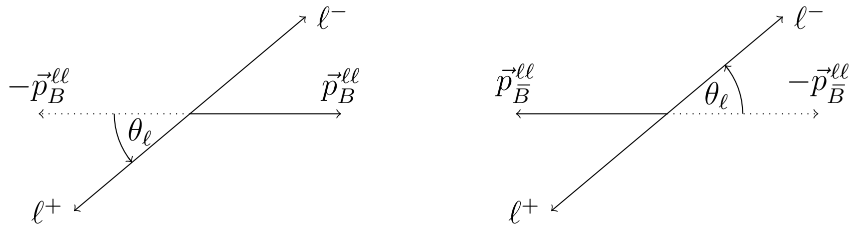


Figure 2.9: Diagrams showing the angle θ_ℓ, for the decay $B^0 \to K^+\pi^-\ell^+\ell^-$ (left) and the decay $\overline{B}^0 \to K^-\pi^+\ell^+\ell^-$ (right). For the former θ_ℓ is defined via Eqn. 2.95, for the latter, via Eqn. 2.96. Both diagrams are shown in the rest frame of the dilepton system.



The angle θ_ℓ is defined as the angle between the positively (negatively) charged lepton in the rest frame of the dilepton system and the direction of the dilepton object in the rest frame of the B ($\overline{B}$) meson. For the decay $B^0 \to K^+\pi^-\ell^+\ell^-$,

$$\cos\theta_\ell = \frac{\vec{p}_{\ell^+}^{\,\ell\ell} \cdot \vec{p}_{\ell\ell}^{\,B}}{|\vec{p}_{\ell^+}^{\,\ell\ell}||\vec{p}_{\ell\ell}^{\,B}|} = \vec{u}_{\ell^+}^{\,\ell\ell} \cdot \vec{u}_{\ell\ell}^{\,B}, \tag{2.95}$$

and for the decay $\overline{B}^0 \to K^-\pi^+\ell^+\ell^-$,

$$\cos\theta_\ell = \frac{\vec{p}_{\ell^-}^{\,\ell\ell} \cdot \vec{p}_{\ell\ell}^{\,\overline{B}}}{|\vec{p}_{\ell^-}^{\,\ell\ell}||\vec{p}_{\ell\ell}^{\,\overline{B}}|} = \vec{u}_{\ell^-}^{\,\ell\ell} \cdot \vec{u}_{\ell\ell}^{\,\overline{B}}, \tag{2.96}$$

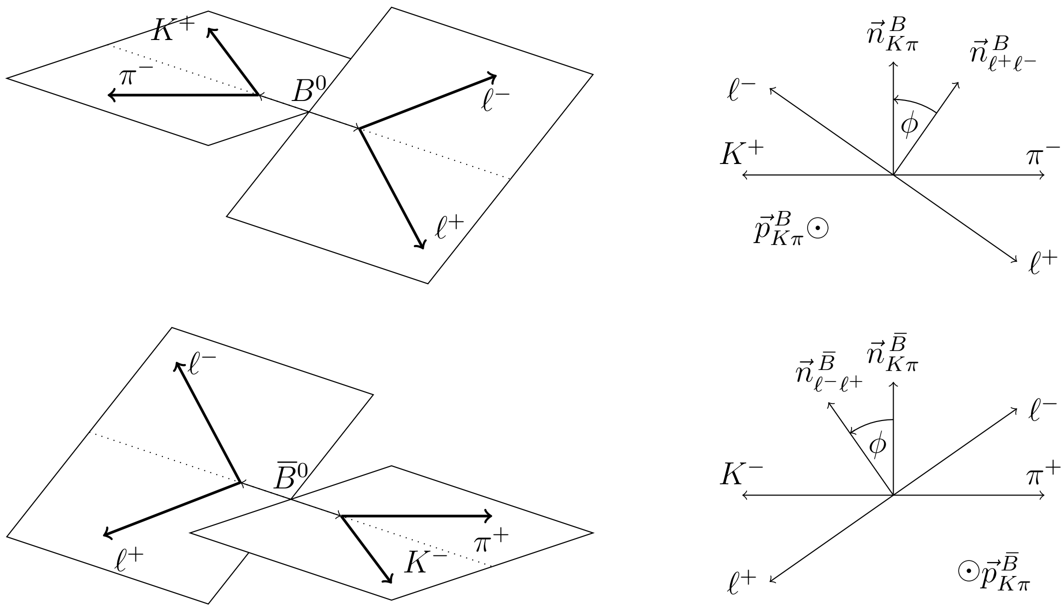
These are shown in the left- and right-hand diagrams of Fig. 2.9 respectively.

Finally, the angle ϕ is defined as the angle between the planes of the $K\pi$ and dilepton systems in the rest frame of the B meson. For the decay $B^0 \to K^+\pi^-\ell^+\ell^-$,

$$\cos\phi = \frac{\vec{p}_{\ell^+}^{\,B} \times \vec{p}_{\ell^-}^{\,B}}{|\vec{p}_{\ell^+}^{\,B} \times \vec{p}_{\ell^-}^{\,B}|} \cdot \frac{\vec{p}_{K}^{\,B} \times \vec{p}_{\pi}^{\,B}}{|\vec{p}_{K}^{\,B} \times \vec{p}_{\pi}^{\,B}|} = \vec{n}_{\ell^+\ell^-}^{\,B} \cdot \vec{n}_{K\pi}^{\,B}; \tag{2.97}$$

$$\sin\phi = \left(\frac{\vec{p}_{\ell^+}^{\,B} \times \vec{p}_{\ell^-}^{\,B}}{|\vec{p}_{\ell^+}^{\,B} \times \vec{p}_{\ell^-}^{\,B}|} \times \frac{\vec{p}_{K}^{\,B} \times \vec{p}_{\pi}^{\,B}}{|\vec{p}_{K}^{\,B} \times \vec{p}_{\pi}^{\,B}|}\right) \cdot \frac{\vec{p}_{K\pi}^{\,B}}{|\vec{p}_{K\pi}^{\,B}|} = \left(\vec{n}_{\ell^+\ell^-}^{\,B} \times \vec{n}_{K\pi}^{\,B}\right) \cdot \vec{u}_{K\pi}^{\,B} \tag{2.98}$$

Figure 2.10: Diagrams showing the angle ϕ. For the decay $B^0 \to K^+\pi^-\ell^+\ell^-$ (top), ϕ is defined via Eqns 2.97 and 2.98. For the decay $\overline{B}^0 \to K^-\pi^+\ell^+\ell^-$ (bottom), it is defined via Eqns 2.99 and 2.100. All diagrams are shown in the B meson rest frame.



and for the decay $\overline{B}^0 \to K^-\pi^+\ell^+\ell^-$,

$$\cos\phi = \frac{\vec{p}^{\overline{B}}_{\ell^-} \times \vec{p}^{\overline{B}}_{\ell^+}}{\left|\vec{p}^{\overline{B}}_{\ell^-} \times \vec{p}^{\overline{B}}_{\ell^+}\right|} \cdot \frac{\vec{p}^{\overline{B}}_{K} \times \vec{p}^{\overline{B}}_{\pi}}{\left|\vec{p}^{\overline{B}}_{K} \times \vec{p}^{\overline{B}}_{\pi}\right|} \qquad = \vec{n}^{\overline{B}}_{\ell^-\ell^+} \cdot \vec{n}^{\overline{B}}_{K\pi}; \tag{2.99}$$

$$\sin\phi = \left(\frac{\vec{p}^{\overline{B}}_{\ell^-} \times \vec{p}^{\overline{B}}_{\ell^+}}{\left|\vec{p}^{\overline{B}}_{\ell^-} \times \vec{p}^{\overline{B}}_{\ell^+}\right|} \times \frac{\vec{p}^{\overline{B}}_{K} \times \vec{p}^{\overline{B}}_{\pi}}{\left|\vec{p}^{\overline{B}}_{K} \times \vec{p}^{\overline{B}}_{\pi}\right|}\right) \cdot \frac{\vec{p}^{\overline{B}}_{K\pi}}{\left|\vec{p}^{\overline{B}}_{K\pi}\right|} \qquad = \left(\vec{n}^{\overline{B}}_{\ell^-\ell^+} \times \vec{n}^{\overline{B}}_{K\pi}\right) \cdot \vec{u}^{\overline{B}}_{K\pi}. \tag{2.100}$$

The notation $\vec{n}^{c}_{ab}$ denotes the unit normal to the decay plane of particles a and b in the rest frame of particle c. The ordering of indices ab picks a direction for the normal vector, $\vec{n}^{c}_{ab} = -\vec{n}^{c}_{ba}$. The angle ϕ is shown diagrammatically in Fig. 2.10.

2.6 Angular differential decay rate

The angular decay rate for $B^0 \to K_J^* \ell^+ \ell^-$ [6][10] is,

$$
\begin{aligned}
\frac{\mathrm{d}^5\Gamma}{\mathrm{d}q^2 \mathrm{d}m_{K\pi} \mathrm{d}\cos\theta_K \mathrm{d}\cos\theta_\ell \mathrm{d}\phi} =& \frac{3}{8}\Big(\mathcal{J}_1^c + 2\mathcal{J}_1^s + (\mathcal{J}_2^c + 2\mathcal{J}_2^s)\cos 2\theta_\ell + 2\mathcal{J}_3 \sin^2\theta_\ell \cos 2\phi \\
&+ 2\sqrt{2}\mathcal{J}_4 \sin 2\theta_\ell \cos\phi + 2\sqrt{2}\mathcal{J}_5 \sin\theta_\ell \cos\phi \\
&+ 2\mathcal{J}_6 \cos\theta_\ell + 2\sqrt{2}\mathcal{J}_7 \sin\theta_\ell \sin\phi \\
&+ 2\sqrt{2}\mathcal{J}_8 \sin 2\theta_\ell \sin\phi + 2\mathcal{J}_9 \sin^2\theta_\ell \sin 2\phi \Big).
\end{aligned}
\tag{2.101}
$$

The angular coefficients, $\mathcal{J}_i$, are given by,

$$
\begin{aligned}
\mathcal{J}_1^c =& |\mathcal{A}_{L0}|^2 + |\mathcal{A}_{R0}|^2 + 8\frac{m_\ell^2}{q^2}\mathfrak{Re}\left(\mathcal{A}_{L0}\mathcal{A}_{R0}^*\right) + 4\frac{m_\ell^2}{q^2}|\mathcal{A}_t|^2 && (2.102) \\
\mathcal{J}_1^s =& \frac{3}{4}\left(|\mathcal{A}_{L\parallel}|^2 + |\mathcal{A}_{L\perp}|^2 + |\mathcal{A}_{R\parallel}|^2 + |\mathcal{A}_{R\perp}|^2\right)\left(1 - \frac{4m_\ell^2}{3q^2}\right) \\
&+ \frac{4m_\ell^2}{q^2}\mathfrak{Re}\left(\mathcal{A}_{L\parallel}\mathcal{A}_{R\parallel}^* + \mathcal{A}_{L\perp}\mathcal{A}_{R\perp}^*\right) && (2.103) \\
\mathcal{J}_2^c =& -\beta_\ell^2\left(|\mathcal{A}_{L0}|^2 + |\mathcal{A}_{R0}|^2\right) && (2.104) \\
\mathcal{J}_2^s =& \frac{1}{4}\beta_\ell^2\left(|\mathcal{A}_{L\parallel}|^2 + |\mathcal{A}_{L\perp}|^2 + |\mathcal{A}_{R\parallel}|^2 + |\mathcal{A}_{R\perp}|^2\right) && (2.105) \\
\mathcal{J}_3 =& \frac{1}{2}\beta_\ell^2\left(|\mathcal{A}_{L\perp}|^2 - |\mathcal{A}_{L\parallel}|^2 + |\mathcal{A}_{R\perp}|^2 - |\mathcal{A}_{R\parallel}|^2\right) && (2.106) \\
\mathcal{J}_4 =& \frac{1}{\sqrt{2}}\beta_\ell^2\left(\mathfrak{Re}\left(\mathcal{A}_{L0}\mathcal{A}_{L\parallel}^*\right) + \mathfrak{Re}\left(\mathcal{A}_{R0}\mathcal{A}_{R\parallel}^*\right)\right) && (2.107) \\
\mathcal{J}_5 =& \sqrt{2}\beta_\ell\left(\mathfrak{Re}\left(\mathcal{A}_{L0}\mathcal{A}_{L\perp}^*\right) - \mathfrak{Re}\left(\mathcal{A}_{R0}\mathcal{A}_{R\perp}^*\right)\right) && (2.108) \\
\mathcal{J}_6 =& 2\beta_\ell\left(\mathfrak{Re}\left(\mathcal{A}_{L\parallel}\mathcal{A}_{L\perp}^*\right) - \mathfrak{Re}\left(\mathcal{A}_{R\parallel}\mathcal{A}_{R\perp}^*\right)\right) && (2.109) \\
\mathcal{J}_7 =& \sqrt{2}\beta_\ell\left(\mathfrak{Im}\left(\mathcal{A}_{L0}\mathcal{A}_{L\parallel}^*\right) - \mathfrak{Im}\left(\mathcal{A}_{R0}\mathcal{A}_{R\parallel}^*\right)\right) && (2.110) \\
\mathcal{J}_8 =& \frac{1}{\sqrt{2}}\beta_\ell^2\left(\mathfrak{Im}\left(\mathcal{A}_{L0}\mathcal{A}_{L\perp}^*\right) + \mathfrak{Im}\left(\mathcal{A}_{R0}\mathcal{A}_{R\perp}^*\right)\right) && (2.111) \\
\mathcal{J}_9 =& \beta_\ell^2\mathfrak{Im}\left(\mathcal{A}_{L\parallel}\mathcal{A}_{L\perp}^*\right) + \mathfrak{Im}\left(\mathcal{A}_{R\parallel}\mathcal{A}_{R\perp}^*\right), && (2.112)
\end{aligned}
$$

where $\beta_\ell = \sqrt{1 - 4m_\ell^2/q^2}$. The functions $\mathcal{A}$ are defined as

$$
\begin{aligned}
\mathcal{A}_{H\,0/t} &= \sum_{J=0,1,2\ldots} Y_J^0(\theta_K, 0)\, A_{JH\,0/t}\, g_{K_J^*}(m_{K\pi}) && (2.113) \\
\mathcal{A}_{H\,\parallel/\perp} &= \sum_{J=0,1,2\ldots} Y_J^{-1}(\theta_K, 0)\, A_{JH\,\parallel/\perp}\, g_{K_J^*}(m_{K\pi}) && (2.114)
\end{aligned}
$$

[10]The notation of Ref. [6] is adapted to avoid confusion with conventional notation for the pure P-wave coefficients (namely that the I_i and A in Ref. [6] are written $\mathcal{J}_i$ and $\mathcal{A}$ here). The normalisation of Eqn (13) in Ref. [6] is slightly altered for consistency with the convention in Refs [16, 21].

where $Y_l^m(\theta, 0)$ are the spherical harmonics, the function $g_{K_J^*}$ encodes the $m_{K\pi}$ dependence, and A_{JHp} are the complex decay amplitudes. The subscript order JHp will be preserved and will be: orbital angular momentum J; chirality $H = L, R$; and polarisation $p = 0, t, \perp, \parallel$ with symbols corresponding to longitudinal, time-like, perpendicular and parallel polarisations. Note that the P-wave longitudinally polarised amplitude, $A_{1L(R)0}$, was introduced in Eqns. 2.91 and 2.92.

For $q^2 \gg m_\ell^2$, one may neglect lepton mass terms. In this approximation, $\beta_\ell \to 1$ and $m_\ell^2/q^2 \to 0$. The angular coefficients, particularly $\mathcal{J}_{1,2}$ are simplified and the time-like polarised amplitudes A_{JHt} disappear. Such a simplification holds for most of the q^2 spectrum of $B^0 \to K_1^*(892)^0\mu^+\mu^-$ and $B^0 \to K_1^*(892)^0 e^+e^-$ decays. For simplicity, the approximation of vanishing lepton mass is made in this thesis.

2.6.1 The angular differential decay rate for pure P-wave $B^0 \to K_1^*(892)^0\ell^+\ell^-$ decays

The angular distribution in the ideal case of P-wave $B^0 \to K_1^*(892)^0\ell^+\ell^-$ has been extensively studied [16–18, 21, 31, 36, 49, 50]. This distribution is obtained from Eqns 2.113 and 2.114 by specifying $J = 1$.

In this case, the relevant bilinear combinations of Eqns 2.113 and 2.114 become:

$$|\mathcal{A}_{H0}|^2 = \frac{3}{4\pi} |A_{1H0}|^2 \left|g_{K_1^*}(m_{K\pi})\right|^2 \cos^2\theta_K \tag{2.115}$$

$$\left|\mathcal{A}_{H\,\parallel/\perp}\right|^2 = \frac{3}{8\pi} \left|A_{1H\,\parallel/\perp}\right|^2 \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K \tag{2.116}$$

$$\mathcal{A}_{H0}\mathcal{A}^*_{H,\parallel/\perp} = \frac{3}{8\pi}\frac{1}{\sqrt{2}} A_{1H0}A^*_{1H\,\parallel/\perp} \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin 2\theta_K \tag{2.117}$$

The angular coefficients (Eqns 2.102–2.112) become:

$$\begin{aligned}
\mathcal{J}_1^c &= \frac{3}{4\pi}\left(|A_{1L0}|^2 + |A_{1R0}|^2\right)\left|g_{K_1^*}(m_{K\pi})\right|^2 \cos^2\theta_K \\
&= \frac{3}{4\pi} I_1^c \left|g_{K_1^*}(m_{K\pi})\right|^2 \cos^2\theta_K && (2.118)\\
\mathcal{J}_1^s &= \frac{3}{4}\frac{3}{8\pi}\left(\left|A_{1L\parallel}\right|^2 + |A_{1L\perp}|^2 + \left|A_{1R\parallel}\right|^2 + |A_{1R\perp}|^2\right)\left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K \\
&= \frac{3}{8\pi} I_1^s \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K && (2.119)\\
\mathcal{J}_2^c &= -\mathcal{J}_1^c = \frac{3}{4\pi} I_2^c \left|g_{K_1^*}(m_{K\pi})\right|^2 \cos^2\theta_K && (2.120)\\
\mathcal{J}_2^s &= \frac{1}{3}\mathcal{J}_1^s = \frac{3}{8\pi} I_2^s \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K && (2.121)\\
\mathcal{J}_3 &= \frac{1}{4}\frac{3}{8\pi}\left(|A_{1L\perp}|^2 - \left|A_{1L\parallel}\right|^2 + |A_{1R\perp}|^2 - \left|A_{1R\parallel}\right|^2\right)\left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K \\
&= \frac{3}{8\pi} I_3 \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K && (2.122)
\end{aligned}$$

$$
\begin{aligned}
\mathcal{J}_4 &= \frac{3}{8\pi}\frac{1}{(\sqrt{2})^2}\left(\mathfrak{Re}\left(A_{1L0}A^*_{1L\parallel}\right) + \mathfrak{Re}\left(A_{1R0}A^*_{1R\parallel}\right)\right)\left|g_{K_1^*}(m_{K\pi})\right|^2 \sin 2\theta_K \\
&= \frac{3}{8\pi}\frac{1}{\sqrt{2}} I_4 \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin 2\theta_K && (2.123)\\
\mathcal{J}_5 &= \frac{3}{8\pi}\frac{\sqrt{2}}{\sqrt{2}}\left(\mathfrak{Re}\left(A_{1L0}A^*_{1L\perp}\right) - \mathfrak{Re}\left(A_{1R0}A^*_{1R\perp}\right)\right)\left|g_{K_1^*}(m_{K\pi})\right|^2 \sin 2\theta_K \\
&= \frac{3}{8\pi}\frac{1}{\sqrt{2}} I_5 \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin 2\theta_K && (2.124)\\
\mathcal{J}_6 &= 2\frac{3}{8\pi}\left(\mathfrak{Re}\left(A_{1L\parallel}A^*_{1L\perp}\right) - \mathfrak{Re}\left(A_{1R\parallel}A^*_{1R\perp}\right)\right)\left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K \\
&= \frac{3}{8\pi} I_6 \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K && (2.125)\\
\mathcal{J}_7 &= \frac{3}{8\pi}\frac{\sqrt{2}}{\sqrt{2}}\left(\mathfrak{Im}\left(A_{1L0}A^*_{1L\parallel}\right) - \mathfrak{Im}\left(A_{1R0}A^*_{1R\parallel}\right)\right)\left|g_{K_1^*}(m_{K\pi})\right|^2 \sin 2\theta_K \\
&= \frac{3}{8\pi}\frac{1}{\sqrt{2}} I_7 \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin 2\theta_K && (2.126)\\
\mathcal{J}_8 &= \frac{3}{8\pi}\frac{1}{(\sqrt{2})^2}\left(\mathfrak{Im}\left(A_{1L0}A^*_{1L\perp}\right) + \mathfrak{Im}\left(A_{1R0}A^*_{1R\perp}\right)\right)\left|g_{K_1^*}(m_{K\pi})\right|^2 \sin 2\theta_K \\
&= \frac{3}{8\pi}\frac{1}{\sqrt{2}} I_8 \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin 2\theta_K && (2.127)\\
\mathcal{J}_9 &= \left(\mathfrak{Im}\left(A_{1L\parallel}A^*_{1L\perp}\right) + \mathfrak{Im}\left(A_{1R\parallel}A^*_{1R\perp}\right)\right)\left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K \\
&= \frac{3}{8\pi} I_9 \left|g_{K_1^*}(m_{K\pi})\right|^2 \sin^2\theta_K. && (2.128)
\end{aligned}
$$

The I_i which have been introduced will become the angular coefficients of the pure P-wave angular distribution.[11] The $m_{K\pi}$ dependence is the same for all A_{1Hp}. Integrating over $m_{K\pi}$, all bilinear combinations of the amplitudes obtain a constant factor

$$\zeta \equiv \int \left|g_{K_1^*}(m_{K\pi})\right|^2 \mathrm{d}m_{K\pi}. \tag{2.129}$$

This factor is often absorbed into the normalisation of the angular distribution, however here it is left explicit since it will have a different form for P- and S-wave contributions in the following section. Substituting Eqns 2.119–2.128 into Eqn. 2.101 and integrating over

[11] The I_i are defined in, for example, Refs [16, 21] analogously to the $\mathcal{J}_i$ (Eqns 2.102–2.112) but for the specific pure P-wave case. Note that some Refs, *e.g.* [21] use J_i.

$m_{K\pi}$,

$$\begin{aligned}\frac{1}{\zeta}\frac{\mathrm{d}^4\Gamma}{\mathrm{d}q^2\mathrm{d}\cos\theta_K\mathrm{d}\cos\theta_\ell\mathrm{d}\phi} =& \frac{9}{32\pi}\bigg(I_1^c\cos^2\theta_K + I_1^s\sin^2\theta_K \\ &+ (I_2^c\cos^2\theta_K + I_2^s\sin^2\theta_K)\cos 2\theta_\ell \\ &+ I_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi \\ &+ I_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi \\ &+ I_5\sin 2\theta_K\sin\theta_\ell\cos\phi \\ &+ I_6\sin^2\theta_K\cos\theta_\ell \\ &+ I_7\sin 2\theta_K\sin\theta_\ell\sin\phi \\ &+ I_8\sin 2\theta_K\sin 2\theta_\ell\sin\phi \\ &+ I_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi\bigg).\end{aligned} \tag{2.130}$$

Integrating over $\cos\theta_K$, $\cos\theta_\ell$, and ϕ one obtains

$$\frac{1}{\zeta}\frac{\mathrm{d}\Gamma}{\mathrm{d}q^2} = \frac{1}{4}\left(3I_1^c + 6I_1^s - I_2^c - 2I_2^s\right) \tag{2.131}$$

$$= \sum_{L,R}\sum_{0,\perp,\parallel}\left|A_{1Hp}\right|^2. \tag{2.132}$$

Equations 2.130 and 2.132 were derived for $B^0 \to K_1^*(892)^0\ell^+\ell^-$. Following the same procedure for $\overline{B}^0 \to \overline{K_1^*}(892)^0\ell^+\ell^-$, yields a distribution $\bar{\Gamma}$ with angular terms $\bar{I}_i$ in terms of bilinear combinations of the amplitudes $\bar{A}$. A set of *CP*-averaged observables, S_i, are conventionally defined,

$$S_i \equiv \left(I_i + \bar{I}_i\right)\bigg/\frac{1}{\zeta}\frac{\mathrm{d}(\Gamma+\bar{\Gamma})}{\mathrm{d}q^2}. \tag{2.133}$$

Note that the *CP*-averaged rate is the sum of the squares of longitudinal and transverse P-wave amplitudes,

$$\gamma \equiv \frac{1}{\zeta}\frac{\mathrm{d}(\Gamma+\bar{\Gamma})}{\mathrm{d}q^2} = \sum_{L,R}\sum_{0,\perp,\parallel}\left(\left|A_{1Hp}\right|^2 + \left|\bar{A}_{1Hp}\right|^2\right), \tag{2.134}$$

which is proportional to the P-wave differential branching fraction,

$$\frac{\mathrm{d}\mathcal{B}\left[B^0 \to K_1^*(892)^0\ell^+\ell^-\right]}{\mathrm{d}q^2} = \frac{\mathrm{d}}{\mathrm{d}q^2}\frac{\Gamma+\bar{\Gamma}}{\mathbf{\Gamma}} = \frac{\gamma}{\mathbf{\Gamma}}, \tag{2.135}$$

where $\mathbf{\Gamma}$ is the total rate of decay of B^0 and $\overline{B}^0$ mesons. The differential branching fraction to the dimuon final state, $\mathrm{d}\mathcal{B}\left[B^0 \to K_1^*(892)^0\mu^+\mu^-\right]/\mathrm{d}q^2$, is the observable of interest for this thesis.

The definition in Eqn 2.133 implies that the fraction of longitudinal polarised amplitudes

$$F_{\mathrm{L}} \equiv \frac{\left|A_{1L0}\right|^2 + \left|A_{1R0}\right|^2 + \left|\bar{A}_{1L0}\right|^2 + \left|\bar{A}_{1R0}\right|^2}{\gamma} = \frac{I_1^c + \bar{I}_1^c}{\gamma} = S_1^c, \tag{2.136}$$

and that the fraction of transverse polarised amplitudes

$$1 - F_{\mathrm{L}} = \frac{4}{3}\frac{I_1^s + \bar{I}_1^s}{\gamma} = \frac{4}{3}S_1^s. \tag{2.137}$$

Conventionally, the *CP*-averaged forward-backward asymmetry of the lepton[12] is also introduced, as

$$A_{\mathrm{FB}} = \frac{3}{4}S_6. \tag{2.138}$$

Thus the *CP*-averaged angular distribution may be written

$$\begin{aligned}
\frac{1}{\zeta\gamma}\frac{\mathrm{d}^4(\Gamma+\bar{\Gamma})}{\mathrm{d}q^2\mathrm{d}\cos\theta_K\mathrm{d}\cos\theta_\ell\mathrm{d}\phi} =& \frac{9}{32\pi}\Bigg(\frac{3}{4}(1-F_{\mathrm{L}})\sin^2\theta_K + F_{\mathrm{L}}\cos^2\theta_K \\
&+ \frac{1}{4}(1-F_{\mathrm{L}})\sin^2\theta_K\cos 2\theta_\ell - F_{\mathrm{L}}\cos^2\theta_K\cos 2\theta_\ell \\
&+ S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi \\
&+ S_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi \\
&+ S_5\sin 2\theta_K\sin\theta_\ell\cos\phi \\
&+ \frac{4}{3}A_{\mathrm{FB}}\sin^2\theta_K\cos\theta_\ell \\
&+ S_7\sin 2\theta_K\sin\theta_\ell\sin\phi \\
&+ S_8\sin^2\theta_K\sin 2\theta_\ell\sin\phi \\
&+ S_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi\Bigg).
\end{aligned} \tag{2.139}$$

Several parameterisations of the angular coefficients have been proposed [16,31,49,50]. These parameterisations are designed to maximise experimental sensitivity to the Wilson Coefficients whilst minimising the effect of form-factor uncertainties. One particular reparameterisation, presented in Ref. [31], makes the definition

$$P_i \equiv \frac{S_i}{\sqrt{F_{\mathrm{L}}(1-F_{\mathrm{L}})}}, \tag{2.140}$$

such that the soft form factor dependence is minimal. The authors of Ref. [31] refer to the P_i as 'form factor independent' which should, arguably more accurately, be called 'soft form factor independent'.

The angular distribution presented in this section is for the ideal case of purely P-wave $B^0 \to K_1^*(892)^0\ell^+\ell^-$ decays. Experimentally, the $K_1^*(892)^0$ is selected by requiring the data lie within $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$. This approach is unable to distinguish the P-wave decays from other decays $B^0 \to K_J^*\ell^+\ell^-$. In particular scalar, S-wave processes such as $B^0 \to K_0^*(1430)^0\ell^+\ell^-$ will, at some level, be selected by such an invariant mass window.

[12]The asymmetry of the positive (negative) lepton in the rest frame of the dilepton with respect to the B ($\overline{B}$) direction. Equivalently this is the difference of integrals of the *CP*-averaged distribution for $\cos\theta_\ell \in [-1, 0]$ and $\cos\theta_\ell \in [0, 1]$.

2.6.2 The angular differential decay rate for $B^0 \to K_1^*(892)^0\ell^+\ell^-$ decays with an additional S-wave component

In many studies and analyses, the S-wave component is added as a corrective term to the pure P-wave distribution,

$$\frac{\mathrm{d}^4(\Gamma+\bar{\Gamma})}{\mathrm{d}q^2\mathrm{d}\cos\theta_K\mathrm{d}\cos\theta_\ell\mathrm{d}\phi} \longmapsto (1-F_\mathrm{S})\left.\frac{\mathrm{d}^4(\Gamma+\bar{\Gamma})}{\mathrm{d}q^2\mathrm{d}\cos\theta_K\mathrm{d}\cos\theta_\ell\mathrm{d}\phi}\right|_\mathrm{P} + F_\mathrm{S}\cos 2\theta_\ell + \text{P-S interference terms} \tag{2.141}$$

This approach is not sufficient for the analysis presented in this thesis which models the $m_{K\pi}$ dependence of an explicit S-wave component. Instead, Eqns 2.113 and 2.114 are evaluated for the two-component[13] case $J = 0, 1$:

$$\mathcal{A}_{H,0} = \frac{1}{\sqrt{4\pi}} A_{0H0}\, g_{K_0^*} + \sqrt{\frac{3}{4\pi}} A_{1,H,0}\, g_{K_1^*} \cos\theta_K \tag{2.142}$$

$$\mathcal{A}_{H,||/\perp} = \sqrt{\frac{3}{8\pi}} A_{1H,||/\perp}\, g_{K_1^*} \sin\theta_K; \tag{2.143}$$

In order to further simplify the distribution, and to increase the statistics of the data, the so-called 'folding' transformation in ϕ:

$$\phi \longmapsto \phi' = \begin{cases} \phi+\pi; & \phi < 0 \\ \phi; & \phi \geq 0 \end{cases} \tag{2.144}$$

is employed to make terms $\cos\phi$ and $\sin\phi$ vanish from Eqn. 2.101. This is the same strategy as used in Refs [26, 27].

The terms $\mathcal{J}_1^{c,s}$ are of immediate relevance since they do not disappear when integrating over $\cos\theta_K$ and ϕ. Additionally $\mathcal{J}_1^c$ is the only term containing the S-wave amplitude.

[13] Specialising the formalism of Ref. [6] in this case has previously been performed in Ref. [51]. However the present section uses a slightly different convention and accounts for minor errors in Ref [51].

Making the substitution above of Eqns 2.142 and 2.143:

$$\begin{aligned}\mathcal{J}_1^c &= \frac{1}{4\pi}\left|g_{K_0^*}\right|^2\left(|A_{0L0}|^2+|A_{0R0}|^2\right)\\
&\quad+\frac{3}{4\pi}\left|g_{K_1^*}\right|^2\left(|A_{1R0}|^2+|A_{1R0}|^2\right)\cos^2\theta_K\\
&\quad+\frac{\sqrt{3}}{4\pi}2\mathfrak{Re}\left((g_{K_0^*}g_{K_1^*}^*)\cdot(A_{0L0}A_{1L0}^*+A_{0R0}A_{1R0}^*)\right)\cos\theta_K \qquad (2.145)\\
&\equiv \frac{1}{4\pi}\left|g_{K_0^*}\right|^2\cdot\gamma_{\mathrm{S}}\\
&\quad+\frac{3}{4\pi}\left|g_{K_1^*}\right|^2\cdot\gamma_{\mathrm{P0}}\cos^2\theta_K\\
&\quad+\frac{\sqrt{3}}{2\pi}\mathfrak{Re}\left((g_{K_0^*}g_{K_1^*}^*)\cdot(x_1+ix_2)\right)\cos\theta_K, \qquad (2.146)\end{aligned}$$

$$\begin{aligned}\mathcal{J}_1^s &= \frac{3}{4}\frac{3}{8\pi}\left|g_{K_1^*}\right|^2\left(|A_{1L\perp}|^2+\left|A_{1L\parallel}\right|^2+|A_{1R\perp}|^2+\left|A_{1R\parallel}\right|^2\right)\sin^2\theta_K \qquad (2.147)\\
&\equiv \frac{9}{32\pi}\left|g_{K_1^*}\right|^2\cdot\gamma_{\mathrm{PT}}\sin^2\theta_K. \qquad (2.148)\end{aligned}$$

In Eqns 2.146 and 2.148, five parameters $\{\gamma_{\mathrm{S}},\ \gamma_{\mathrm{P0}},\ \gamma_{\mathrm{PT}},\ x_1,\ x_2\}$ are introduced. These are observables, but are not of direct interest[14]. Pragmatically, they control the relative size of the P- and S-wave components in the angular distribution, with the lineshape information provided by $g_{K_J^*}$. The fraction of longitudinally polarised amplitudes in the P-wave state, introduced in Eqn. 2.136, becomes

$$F_{\mathrm{L}} = \frac{\gamma_{\mathrm{P0}}\cdot\left|g_{K_1^*}(m_{K\pi})\right|^2}{\gamma_{\mathrm{P0}}\cdot\left|g_{K_1^*}(m_{K\pi})\right|^2+\gamma_{\mathrm{PT}}\cdot\left|g_{K_1^*}(m_{K\pi})\right|^2} = \frac{\gamma_{\mathrm{P0}}}{\gamma_{\mathrm{P0}}+\gamma_{\mathrm{PT}}} \qquad (2.149)$$

in this notation, and the $m_{K\pi}$ dependence divides out.

The analysis presented in this thesis is concerned with the size of the S-wave component, but not other angular observables. In the ideal case of constant angular efficiency in $\cos\theta_\ell$ and ϕ these angles, which provide no information about the S-wave component, would be able to be integrated out. The angular distribution of the data would be

$$\begin{aligned}\frac{\mathrm{d}^3\Gamma}{\mathrm{d}q^2\mathrm{d}m_{K\pi}\mathrm{d}\cos\theta_K} &= 2\pi\mathcal{J}_1^c+4\pi\mathcal{J}_1^s-\frac{2\pi}{3}(\mathcal{J}_2^c+2\mathcal{J}_2^s)\\
&= \frac{8\pi}{3}\mathcal{J}_1^c+\frac{32\pi}{9}\mathcal{J}_1^s \qquad (2.150)\\
&= \frac{2}{3}\left|g_{K_0^*}\right|^2\gamma_{\mathrm{S}}+2\left|g_{K_1^*}\right|^2\gamma_{\mathrm{P0}}\cos^2\theta_K+\left|g_{K_1^*}\right|^2\gamma_{\mathrm{PT}}\sin^2\theta_K\\
&\qquad+\frac{4\sqrt{3}}{3}\mathfrak{Re}\left((g_{K_0^*}g_{K_1^*}^*)\cdot(x_1+ix_2)\right)\cos\theta_K\end{aligned}$$

[14]The subscript notation P0 and PT is used since the parameters $\gamma_{\mathrm{P0,PT}}$ are the sum of the squares of the longitudinal and transverse P-wave amplitudes (after factorising the $m_{K\pi}$ dependence).

The angular efficiency is described in more detail in Section 5.3 and discussion of folding the efficiency effects into the model of the data reserved for Section 6.2.1 in particular, Eqn. 6.8. Suffice to say that in reality, the angular efficiency function in $\cos\theta_\ell$ and ϕ is not constant. Performing the integral over $\cos\theta_\ell$ and ϕ will therefore not describe the data, as the efficiency effects cannot not be ignored. The full angular distribution is therefore required. It retains contributions from the remaining $\mathcal{J}_i$, which are:

$$\begin{aligned}
\mathcal{J}_3 &= \frac{3}{16\pi}\left|g_{K_1^*}\right|^2\left(\left|A_{L\perp}\right|^2-\left|A_{L\parallel}\right|^2+\left|A_{R\perp}\right|^2-\left|A_{R\parallel}\right|^2\right)\sin^2\theta_K && (2.151)\\
&= \frac{3}{16\pi}\left|g_{K_1^*}\right|^2\tilde{J}_3\sin^2\theta_K && (2.152)\\
\mathcal{J}_6 &= \frac{3}{4\pi}\left|g_{K_1^*}\right|^2\mathfrak{Re}\left(A_{L\parallel}A_{L\perp}^*-A_{R\parallel}A_{R\perp}^*\right)\sin^2\theta_K && (2.153)\\
&= \frac{3}{4\pi}\left|g_{K_1^*}\right|^2\tilde{J}_6\sin^2\theta_K && (2.154)\\
\mathcal{J}_9 &= \frac{3}{8\pi}\left|g_{K_1^*}\right|^2\mathfrak{Im}\left(A_{L\parallel}A_{L\perp}^*+A_{R\parallel}A_{R\perp}^*\right)\sin^2\theta_K && (2.155)\\
&= \frac{3}{8\pi}\left|g_{K_1^*}\right|^2\tilde{J}_9\sin^2\theta_K && (2.156)
\end{aligned}$$

and contribute one further parameter $(\tilde{J}_i)$ each. Substituting Eqns. 2.146, 2.148, 2.152, 2.154, and 2.156 into Eqn. 2.101, yields the full distribution:

$$\begin{aligned}
\frac{\mathrm{d}^5\Gamma}{\mathrm{d}q^2\mathrm{d}m_{K\pi}\mathrm{d}\cos\theta_K\mathrm{d}\cos\theta_\ell\mathrm{d}\phi} = &\frac{1}{4\pi}\left|g_{K_0^*}\right|^2\cdot\gamma_{\mathrm{S}}(1-\cos2\theta_\ell)\\
&+\frac{3}{4\pi}\left|g_{K_1^*}\right|^2\cdot\gamma_{\mathrm{P0}}\cos^2\theta_K(1-\cos2\theta_\ell)\\
&+\frac{\sqrt{3}}{2\pi}\mathfrak{Re}\left(\left(g_{K_0^*}g_{K_1^*}^*\right)\cdot(x_1+ix_2)\right)\cos\theta_K(1-\cos2\theta_\ell)\\
&+\frac{9}{16\pi}\left|g_{K_1^*}\right|^2\cdot\gamma_{\mathrm{PT}}\sin^2\theta_K\left(1+\frac{1}{3}\cos2\theta_\ell\right)\\
&+\frac{3}{8\pi}\left|g_{K_1^*}\right|^2\tilde{J}_3\sin^2\theta_K\sin^2\theta_\ell\cos2\phi\\
&+\frac{3}{2\pi}\left|g_{K_1^*}\right|^2\tilde{J}_6\sin^2\theta_K\cos\theta_\ell\\
&+\frac{3}{4\pi}\left|g_{K_1^*}\right|^2\tilde{J}_9\sin^2\theta_K\sin^2\theta_\ell\sin2\phi.
\end{aligned} \quad (2.157)$$

Discussion of the parameters is presented in Chapter 6. It should be noted, however, that the parameters $\left\{\tilde{J}_3,\ \tilde{J}_6,\ \tilde{J}_9\right\}$ are related to the parameters $\{S_3,\ A_{\mathrm{FB}},\ S_9\}$ introduced in Eqns 2.133 and 2.138 (for the pure P-wave case). To translate the tilde parameters into conventional P-wave angular observables, measured in previous analyses, one is obliged to integrate over the $m_{K\pi}$ window and scale by the total P-wave rate. In this notation this is,

$$S_3 = \frac{\tilde{J}_3 \int_{796}^{996} \left|g_{K_1^*}\right|^2 \mathrm{d}m_{K\pi}}{(\gamma_{\mathrm{P0}} + \gamma_{\mathrm{PT}}) \cdot \int_{796}^{996} \left|g_{K_1^*}\right|^2 \mathrm{d}m_{K\pi}} = \frac{\tilde{J}_3}{\gamma_{\mathrm{P0}} + \gamma_{\mathrm{PT}}}, \tag{2.158}$$

$$A_{\mathrm{FB}} = \frac{3}{4}\frac{\tilde{J}_6}{\gamma_{\mathrm{P0}} + \gamma_{\mathrm{PT}}}, \tag{2.159}$$

$$S_9 = \frac{\tilde{J}_9}{\gamma_{\mathrm{P0}} + \gamma_{\mathrm{PT}}}, \tag{2.160}$$

which are independent of $m_{K\pi}$ in an analogous way to Eqn 2.149. These three parameters only appear in the model due to the efficiency shape in $\cos\theta_\ell$ and ϕ, since the last three terms of Eqn 2.157 disappear with a uniform efficiency. They are therefore nuisance parameters in this analysis. As will be shown, the model has a low sensitivity to them.

Chapter 3

The LHCb experiment

This chapter cites Ref. [52] throughout unless specified otherwise.

The LHCb detector is one of the four main detectors at the LHC at CERN, Geneva. The LHC is a proton–proton (pp) collider installed in the tunnel previously occupied by the Large Electron Positron collider. The tunnel has a circumference of approximately 27 km and spans the Franco-Swiss border between the city of Geneva and the Jura mountain range. The LHC is designed to collide protons at a centre-of-mass energy ($\sqrt{s}$) of 14 TeV. The datasets used in this thesis, were taken at reduced energies, $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$. They correspond to integrated luminosities of $1\,\text{fb}^{-1}$ and $2\,\text{fb}^{-1}$, in the 2011 and 2012 running of the LHC respectively. This is often referred to as 'LHC Run-I' in the literature. For the LHCb experiment, the proton beams are slightly offset at the collision point such that a lower but uniform luminosity is provided during the entire period when a single fill of the LHC circulates.

The LHCb detector is designed to study the properties of b hadrons. Located at interaction point 8 on the LHC just outside Ferney-Voltaire, LHCb is a spectrometer covering the azimuthal beam angle range $(10 < \theta < 250)\,\text{mrad}$ or equivalently, the pseudorapidity[1] range $2 < \eta < 5$. The detector is often referred to as a 'single-arm, forward spectrometer'. 'Single-arm' since there is no instrumentation about $\theta = \pi$ and 'forward' due to the angular acceptance. As shown in Fig. 3.1, the production of $b\bar{b}$ quark-pairs is enhanced at small angle, driving the design choice of the detector acceptance.

The detector itself comprises a dipole magnet and several subdetectors for tracking, particle identification (PID), and calorimetry. Figure 3.2 shows a cross-section view of the detector. The pp collision point is on the left hand side of this figure, at the origin of the cartesian coordinates shown. In this diagram, the detector instrumentation and the forward-going particles flow from left to right. Following this direction, the subdetectors are introduced as follows. High resolution tracking around the pp collision point is performed by the vertex locator (VELO) subdetector which is described in Section 3.1. The first of two

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

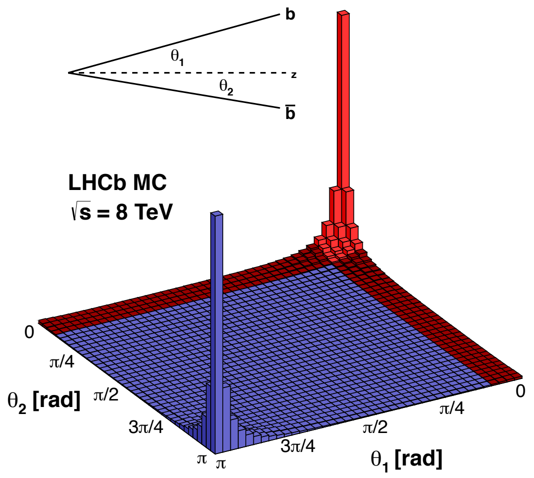
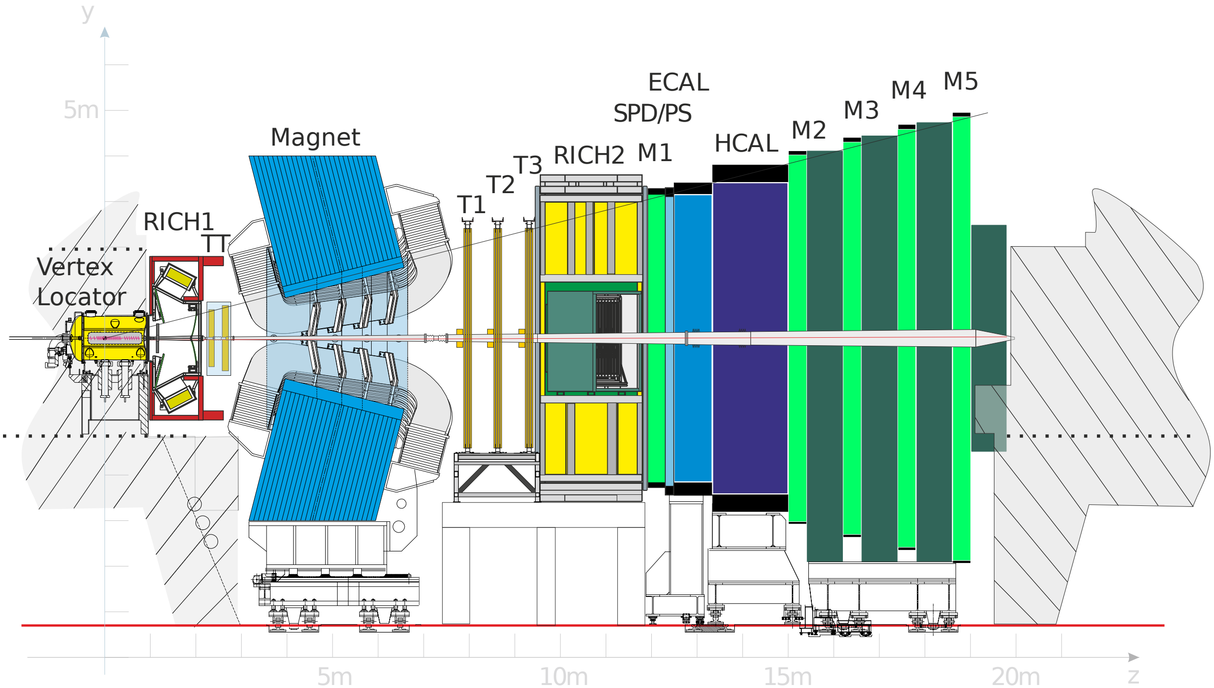


Figure 3.1: The simulated production of $b\bar{b}$ quark pairs in proton-proton collisions as a function of the opening angle, θ_1 (θ_2) of the b- ($\bar{b}$-) quark, shown in the inlay. The red coloured section shows the acceptance of the LHCb detector. Figure from Ref. [53].

ring-imaging Cherenkov (RICH) detectors, which contribute PID information, is followed by the upstream tracking station (TT) and then the magnet. Downstream of the magnet are further tracking stations (T1–3) and the second RICH detector. The tracking stations TT and T1–3 and the magnet are also described in Section 3.1. The RICH detectors and hadron PID are described in Section 3.2. Following the second RICH detector is the first of the muon chambers and the calorimetry system. The muon system is described in Section 3.3. The calorimetery system consists of: the scintillating pad detector (SPD), the preshower detector (PS), the electromagnetic calorimeter (ECAL), and finally the hadronic calorimeter (HCAL). Information from the calorimetry system is not used in the present analysis, but is used in other triggers and for the analysis of complementary radiative or electron decay modes such as $B^0 \to K_1^*(892)^0\gamma$, and $B^+ \to K^+e^+e^-$ for example. Triggering is described in Section 3.4. The calorimetry system is followed by the remaining four muon chambers, each of which are interleaved with iron shielding. The chain of algorithms used to generate simulated data samples is described in Section 3.5.

This thesis concerns $B^0 \to K^+\pi^-\mu^+\mu^-$ decays. The four charged particles produced in these decays, and from elsewhere in the collision event, will pass through the detector leaving deposits of energy in the tracking system. These deposits are referred to as 'hits'. The tracking system is used to perform a fit to all hits recorded in a collision event, and associate the hits to 'tracks'. Tracks are associated to vertices, and some PID or 'mass hypothesis' is associated to each track. An example of a reconstructed collision event is shown in Fig. 3.3. Successful reconstruction and selection of a true $B^0 \to K^+\pi^-\mu^+\mu^-$ decay corresponds to a candidate four-track vertex or 'four-body object' with the correct PID hypothesis. The nomenclature 'candidate' is used to denote that a given four-body object in an event may be a genuine $B^0 \to K^+\pi^-\mu^+\mu^-$ decay and could also be due to some background process. Furthermore a single event, such as the one shown in Fig. 3.3, may contain more than one candidate four-body object.

Figure 3.2: A cross-sectional view of the detector showing the main subdetectors. The collision point is at the origin of the cartesian coordinate system surrounded by the vertex locator (VELO). From left-to-right the remaining subdetectors are: the first ring-imaging Cherenkov (RICH) detector, the upstream tracking station (TT) the magnet, the downstream tracking stations (T1–3), the second RICH detector, the Electromagnetic and Hadronic CALorimeters (ECAL and HCAL), and the muon stations (M1–5). Figure taken from Ref. [54].



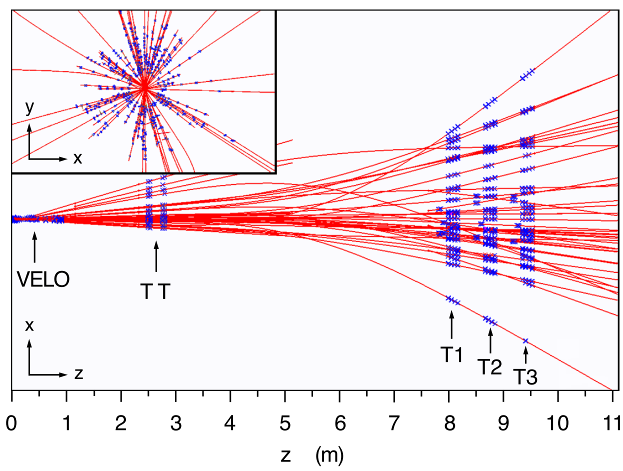
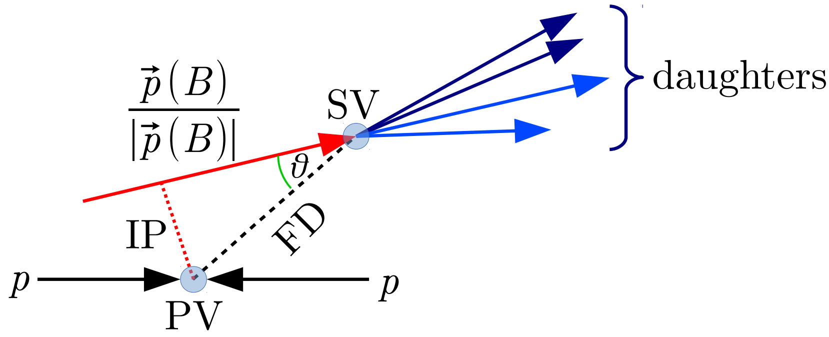


Figure 3.3: An example of an offline-reconstructed event. Hits (blue points) in the vertex locator (VELO), the upstream tracking station (TT) and tracking stations (T1–3) are fit globally associating hits to tracks (red lines) and tracks to vertices. In the case of muon tracks, hits in the muon system would also be included although they are not shown in this figure. Figure taken from Ref. [55].

Figure 3.4: A schematic decay topology for a B meson production and decay in the lab frame at LHCb. The B meson is produced in the proton–proton collision, the primary vertex (PV). The particle flies and decays at the secondary vertex (SV). The momenta of the daughters are used to reconstruct the momentum of the parent B meson, $\vec{p}(B)$. The shortest distance from the reconstructed direction of the parent, $\vec{p}(B)/|\vec{p}(B)|$, to the PV is the impact parameter (IP). An IP may also be calculated for any of the daughter tracks with respect to any vertex. The pointing angle (ϑ), is the angle between $\vec{p}(B)/|\vec{p}(B)|$ and the FD direction.



Candidate b hadron decay events (without PID information) are selected first by the trigger. The trigger applies initially simple tracking and vertex reconstruction algorithms which become more complex (and therefore time-consuming) as the candidates are selected by the algorithms at earlier stages. All candidates following the trigger are re-analysed in a so-called 'offline reconstruction' step. The nomenclature 'offline' is used to mean any analysis to stored data.

3.1 Tracking charged particles

The lifetime of b hadrons at LHC energies means that they fly on the order of millimetres before decaying. This flight distance (FD) is resolved with the VELO subdetector. The B meson is produced in the proton-proton interaction or primary vertex (PV), it then travels for some FD before decaying at its secondary vertex (SV). Figure 3.4 shows a schematic topology of a decay event highlighting these vertices.

The VELO is a fine-grained tracker consisting of silicon strips. The silicon strips instrument so-called 'modules' of which there are 42. The modules are mounted on two retractable bellows (21 modules per side). When data taking during stable pp collisions, the bellows are moved in to 7 mm from the collision point. Each module has two sets of strips arranged in either radial (R) or angular (ϕ) geometry[2] as is shown in Fig. 3.5. An overview of the VELO is shown in Fig. 3.6.

The choice of R, ϕ coordinate geometry enables fast reconstruction of the impact

[2]In the context of this chapter, ϕ denotes the angular beam coordinate. For all other chapters it is used for the helicity angle of $B^0 \to K^+\pi^-\mu^+\mu^-$ decays defined in Section 2.5 (Fig. 2.10, Eqns 2.97–2.100).

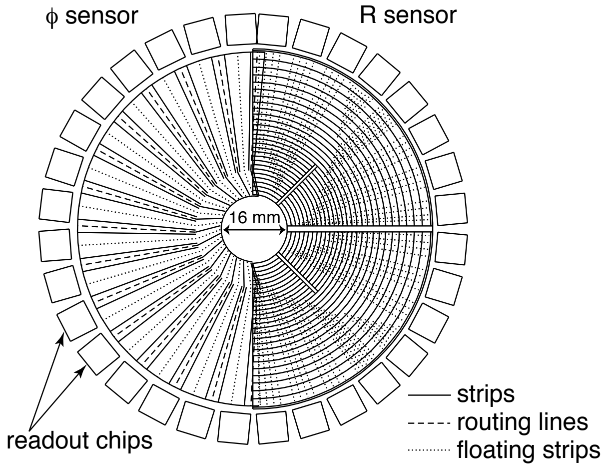


Figure 3.5: The geometry of the VELO silicon strips. Strips are arranged radially to measure angle, ϕ, (left) and in concentric curves to measure radial position, R, (right). Figure taken from Ref. [52].

parameter (IP), which is the minimum distance of a track to a vertex. The IP is also shown in Fig. 3.4. The resolution of this IP is dependent on the component of the momentum transverse to the beam line (p_{T}) and is measured be $(15 + 29\,\mathrm{GeV}^{-1}c/p_{\mathrm{T}})\,\mu\mathrm{m}$. The single hit resolution in the VELO is a function of the microstrip positioning and pitch. In general, the hit resolution is smaller than $30\,\mu\mathrm{m}$ [56]. The PV position resolution is dependent on the number of tracks, N, in the vertex. The resolution of the x position is measured to be $(-1 + 97.7/N^{0.61})\,\mu\mathrm{m}$ which is very slightly worse than the y position resolution. The resolution of the z position is measured to be $(-16 + 923/N^{0.69})\,\mu\mathrm{m}$ which is roughly an order of magnitude worse. For PVs with fewer than 10 tracks, the x or y position resolution is between 15 and $35\,\mu\mathrm{m}$, the z position resolution is between 100 and $300\,\mu\mathrm{m}$ [56].

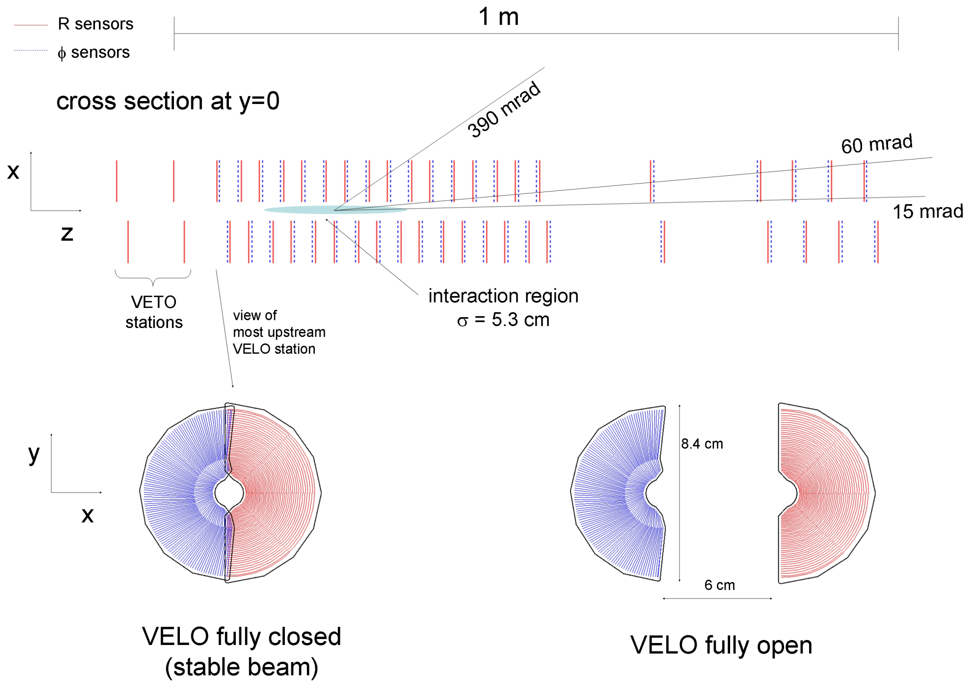
The TT is comprised entirely of silicon strip detectors whereas T1–3, are each divided into silicon strip inner tracker (IT), and straw tube outer tracker (OT) sections. The layout of T1–3 is shown in Fig. 3.7. For the TT and the IT parts of T1–3, the single hit resolution is between 50 and $55\,\mu\mathrm{m}$ [55]. For the straw tubes in the OT parts of T1–3 the single hit resolution is around $200\,\mu\mathrm{m}$ [55].

The magnet has a bending power of about 4 Tm. The curvature of tracks in the magnetic field allows for charge separation and momentum measurement. The combined tracking system and magnet measure momentum with a relative uncertainty ($\delta p/p$) in the range 0.4% (at low momentum) to 0.6% (for high momentum) [55]. The momenta of tracks used by this measurement are typically within this range. The magnet polarity was periodically reversed throughout data-taking in 2011 and 2012 such that measurements of *CP* violation may be performed and the magnet polarity bias assessed.

3.2 Hadron identification

The ability to correctly identify hadron tracks is important for selecting the $K\pi$ system in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays.

Figure 3.6: An overview of the VELO layout. The upper part shows the arrangement of the VELO modules. The lower left (right) show the active area of the modules in closed (open) position. Figure taken from Ref. [52].



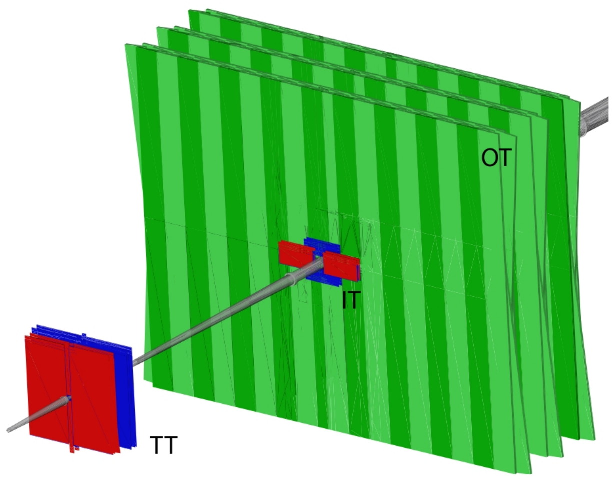


Figure 3.7: A schematic layout of the tracking stations with the beam-pipe. The inner tracker (IT) sections of the upstream tracking station (TT) and tracking stations (T1–3) are shown in red and blue. The outer tracker (OT) sections of T1–3 are shown in green. Figure taken from Ref. [57].

Two RICH detectors, with sensitivity over different track momentum regions, are employed for this purpose. As charged particles pass through matter faster than the speed of light in that matter, they emit radiation conically. This phenomenon is known as 'Cherenkov radiation'. The RICH detectors collect these Cherenkov photons and measure the opening angle (Cherenkov angle, θ_C) of the photons. The Cherenkov angle of the radiation and momentum of the passing particle are related through

$$\cos\theta_C = \frac{\sqrt{m^2c^2 + p^2}}{pn}, \tag{3.1}$$

where n is the refractive index of the radiating medium, which follows from a geometric argument.

The first of the RICH detectors is shown in Fig. 3.8 and makes use of two media, aerogel and C_4F_{10} gas. These media provide $K\pi$ hypothesis separation for tracks in the $(2 < p < 60)\,\mathrm{GeV}/c$ and $(10 < p < 60)\,\mathrm{GeV}/c$ momentum ranges. The second RICH detector makes use of CF_4 as a radiating medium and is able to separate tracks in the $(15 < p < 100)\,\mathrm{GeV}/c$ range. The Cherenkov angle resolution is found to be 1.6 mrad and 0.6 mrad in C_4F_{10} and CF_4 respectively [58].

The Cherenkov angle is combined with reconstructed momentum information and each track is fit to a particle mass hypothesis using Eqn. 3.1. The Difference in Log-Likelihoods (DLL) between two hypotheses is then calculated. The DLL between a kaon and pion hypothesis,

$$\mathrm{DLL}_{K\pi} = \log\mathcal{L}(\theta_C, p|K) - \log\mathcal{L}(\theta_C, p|\pi), \tag{3.2}$$

is the discriminating variable to differentiate kaon and pion tracks. An analogous variable is computed for protons, $\mathrm{DLL}_{p\pi}$.

The left-hand side of Fig. 3.9 illustrates the separation of particle types. Distinct bands corresponding to a particle of given mass are visible as a function of the Cherenkov angle and the reconstructed momentum. The $K\pi$ separation performance of the RICH system is shown to the right of Fig. 3.9. The $\mathrm{DLL}_{K\pi}$ variable provides close to 100% separation for low momentum tracks. Separation worsens with higher momentum tracks and with a less severe requirement on the $\mathrm{DLL}_{K\pi}$ variable.

3.3 Muon system and muon identification

Energetic muons from B meson decays are separated from other tracks by the LHCb muon system. The system, shown in Fig. 3.10, comprises five muon tracking stations M1–5. The stations are shielded in order to absorb hadrons and lower momentum particles. Each station is subdivided into regions (R1–4) of decreasing mean occupancy. Stations M1–5 comprise of 276 multi-wire proportional chambers (MWPCs) for hit detection, with the

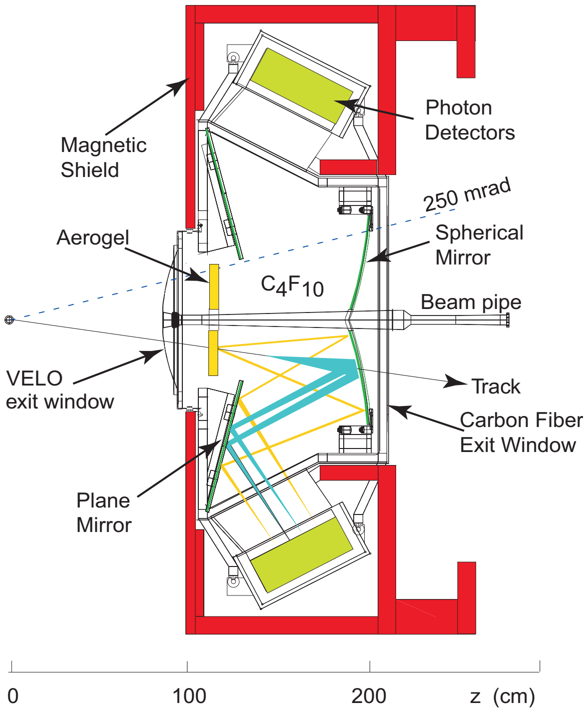


Figure 3.8: A side-view schematic layout of the first ring-imaging Cherenkov (RICH) detector in the same orientation as Fig. 3.2. Particles travel from left to right through the subdetector. Following the arrow labelled 'track', an incident particle travels faster than the speed of light in the C_4F_{10} radiator and emits Cherenkov radiation depicted as the blue (darker grey) solid cone. The cone of light is reflected and focused into a ring by the spherical mirror, and reflected up into an array of photo-multiplier tubes by a plane mirror. From the cone of reflected Cherenkov photons, a measurement of the Cherenkov angle is performed for each particle track in the subdetector. Figure taken from Ref. [52].

Figure 3.9: The reconstructed Cherenkov angle versus reconstructed momentum for selected calibration data (left), and the performance of $K\pi$ separation (right). In the left-hand plot, distinct bands are visible for particle types according to their mass. The curves follow the function in Eqn. 3.1. The right-hand plot shows kaon identification efficiency and pion misidentification rate for the RICH system as a function of track momentum. The efficiencies are measured using $D^0 \to K\pi$ calibration data. Figures taken from Ref. [58].

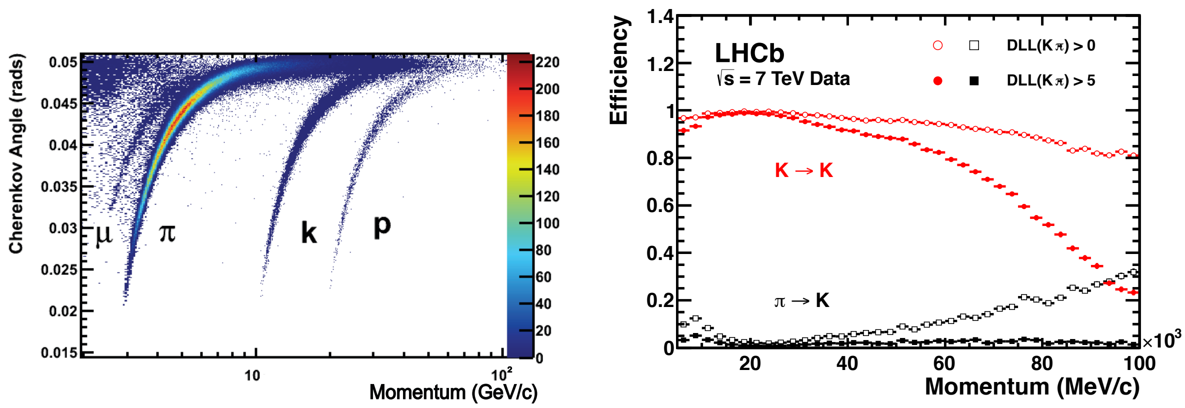
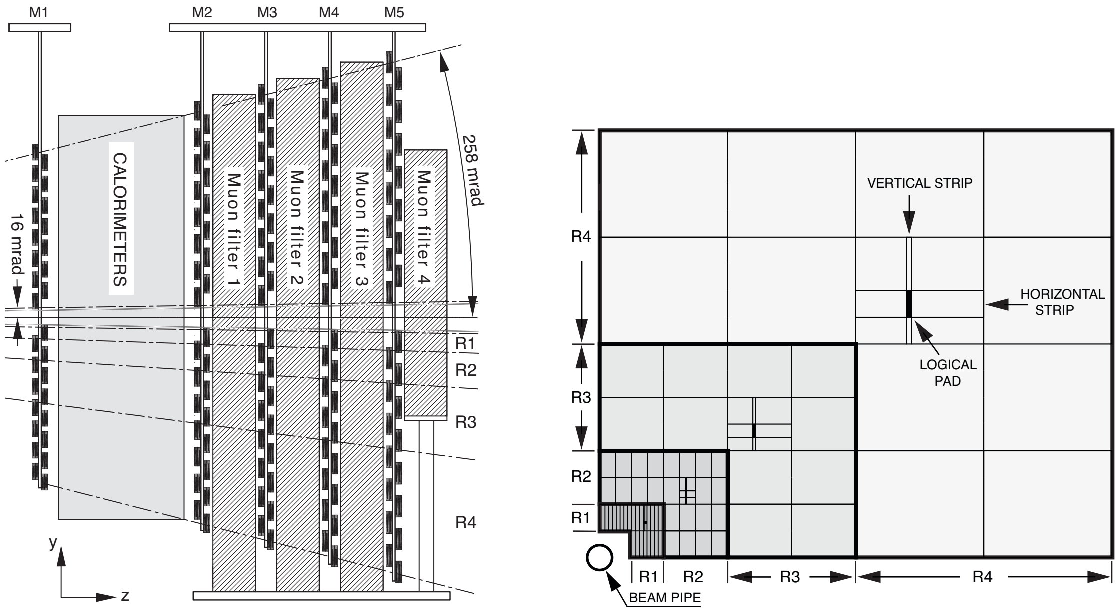


Figure 3.10: A side view of the muon system (left) with the muon stations (M1–5), and a planar view of one quadrant of the muon stations M2 and M3. The left-hand figure is orientated as in Fig 3.2 with particle travel from left to right. Station M1 is shielded by the RICH detector, M2 by the calorimeters, and M3–5 by iron muon filters. The right-hand figure depicts the regions (R1–4) with decreasing mean occupancy. A logical pad is shown for each region shaded in black. Figures taken from Ref. [52].



exception of M1R1 which comprises gas electron multipliers (GEMs), where the occupancy is highest. The MWPCs and GEMs have a time resolution of 5 ns and 3 ns respectively. Good time resolution enables the muon information to be used in the trigger, which is discussed in Section 3.4.

The linear dimensions and segmentation of the regions R1–4 scale in the ratio 1 : 2 : 4 : 8. This design choice is so that particle flux and occupancy are roughly equal throughout all regions. To reduce the number of input channels to the trigger, the stations M2+3, and M4+5 are paired into 'logical' layers. Each of the regions is then divided into 'logical pads' as shown in Fig. 3.10. The number of logical pads in a region determines the spatial resolution. The muon identification used in the trigger is discussed in Section 3.4 and is not discussed any further in the present Section.

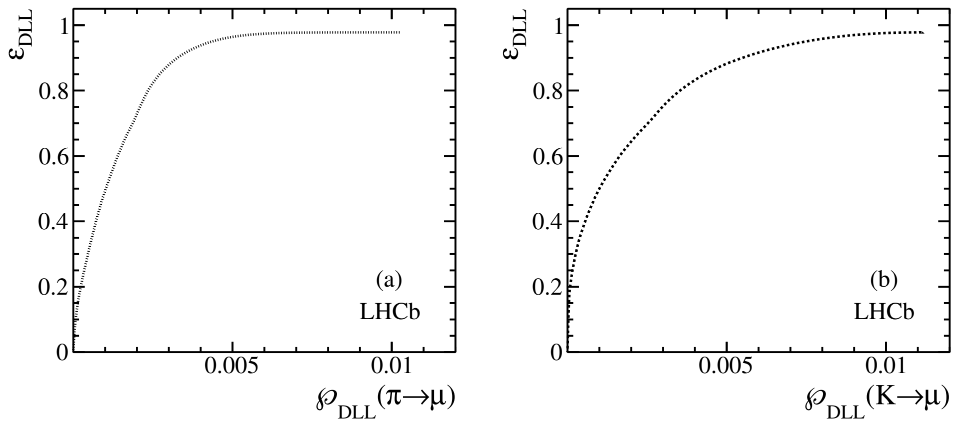
The muon system provides two main variables for offline muon identification [59]. They are a boolean muon identification variable (`IsMuon`) and a continuous DLL. The `IsMuon` variable is based on the presence of hits in logical pads in the stations following an extrapolated field of interest. The field of interest is determined by the track momentum and designed to account for the multiple scattering one expects for a muon traversing the iron. The exact requirements of `IsMuon` are presented in Table 3.1. The DLL variable is defined in an similar way to the $\mathrm{DLL}_{K\pi}$ variable defined in Eqn. 3.2,

$$\mathrm{DLL}_{\mu\pi} = \mathcal{L}(\theta_C, p, D^2|\mu) - \mathcal{L}(\theta_C, p, D^2|\pi). \tag{3.3}$$

Momentum range	Hits required in muon stations
$3\,\mathrm{GeV}/c < p < 6\,\mathrm{GeV}/c$	M2 and M3
$6\,\mathrm{GeV}/c < p < 10\,\mathrm{GeV}/c$	M2, M3 and either M4 or M5
$p > 10\,\mathrm{GeV}/c$	M2, M3, M4 and M5

Table 3.1: The `IsMuon` identification requirements. Taken from Ref. [59].

Figure 3.11: The reciever operating characteristic (ROC) curves for the $\mathrm{DLL}_{\mu\pi}$ variable defined in Eqn. 3.3. The efficiency of selecting true muons using this variable is shown as a function of pion misidentification (a) and kaon misidentification (b) probabilities. ROC curves are described in more detail in Section 4.3. Figures adapted from Ref. [59].



In this expression, θ_C is the Cherenkov angle provided by the RICH subdetectors, and D^2 is the average squared distance significance of hits in the muon chambers with respect to the linear extrapolation of the tracks from the tracking system. True muons have a much narrower D^2 distribution than other particles incorrectly accepted by `IsMuon`. The precise definitions of the field of interest and the D^2 variable are given in Ref. [59].

The `IsMuon` variable has an efficiency between 95 and 100% for true muons, and misidentification probability below 2% for protons, pions, or kaons. The $\mathrm{DLL}_{\mu\pi}$ efficiency is shown in Fig. 3.11 as a function of the pion and kaon misidentification probability.

3.4 Triggers

A particle detector's triggering system is designed to reduce the data readout to a manageable level. Generally the collision rate of an experiment is much higher than the rate at which a detector readout electronics, and the reconstruction algorithms can cope. Furthermore, disk storage considerations limit the output. The problem of this high rate is alleviated by selecting or rejecting events, at the time of data-taking. This drives the design of the triggering system.

The LHCb trigger [60, 61] is required to reduce the data rate from the LHC collision frequency of approximately 40 MHz to around 3 kHz. The trigger is split into three stages, one of which is based in hardware, and two of which are software stages. Each of these trigger stages are described in the following sections. The aim of this split is that a coarse decision precedes the more computationally intensive decision making to improve time efficiency. A coarse decision is made initially, as to whether a given event is worth reading out the whole detector. Given the event readout, a more complex decision is made as to whether it is worthwhile to perform the full reconstruction, and so on.

Of relevance to the analysis described in this thesis, the hardware stage uses a simple search for high p_{T} muons, the first software stage aims to reconstruct energetic muon and hadron tracks, and the final software stage performs a full event reconstruction and is able to select data events based on topology.

3.4.1 Hardware trigger

The hardware-level or 'level-zero' (L0) trigger takes information from the calorimeters and the muon system. A trigger decision is made based on this limited information. The L0 trigger comprises bespoke hardware and runs with the LHC clock, it reduces the rate to below 1 MHz. After the L0 stage, the bandwidth is low enough such that the whole detector may be read out.

The L0 trigger algorithm is formed of three trigger 'paths', which use calorimeter information, muon system information, and pile-up information. Only events that satisfy the trigger path using information from the muon system (`L0Muon`) are used in the analysis described in this thesis.

For the `L0Muon` algorithm, the muon stations are divided into quadrants. In each quadrant, the p_{T} of the muon candidate tracks in M1–3 is measured with a resolution of $\sim 20\%$ [62]. Each quadrant reports the track with the highest and second-highest p_{T}. The `L0Muon` algorithm passes the event if the threshold for the highest p_{T} track is exceeded by at least one of the eight candidates. This `L0Muon` threshold was set to $1.48\,\mathrm{GeV}/c$ in 2011 and increased to $1.76\,\mathrm{GeV}/c$ in 2012 [61].

A second algorithm that is not used in this analysis requires a threshold for the product of the highest and second-highest p_{T} tracks (`L0DiMuon`) [60]. In Fig 3.12 the efficiency of these trigger algorithms is presented. There is a significant overlap between `L0Muon` and `L0DiMuon` candidates, as well as a somewhat lower efficiency for `L0DiMuon`. Therefore the analysis described in this thesis requires only that the signal candidate was selected by the `L0Muon` trigger, a so-called trigger on signal (TOS) event[3].

[3]The alternative, being a trigger independently of signal (TIS) event, is used for control purposes, testing the trigger efficiency (described in Section 7.3.10 of this thesis) and in some cases, selecting background events.

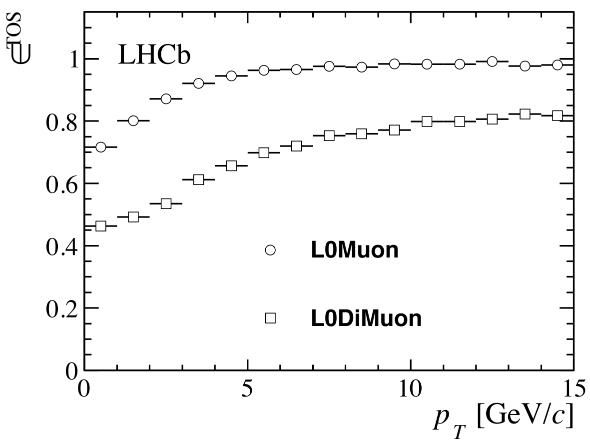


Figure 3.12: The efficiency to trigger on signal (TOS) for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ decays for the `L0Muon` and `L0DiMuon` algorithms (described in the text) as a function of the component of the momentum transverse to the beam line (p_{T}) of the J/ψ. Candidates are required to have been selected by the `L0Muon` algorithm for the analysis described in this document. Figure taken from Ref. [60].

3.4.2 The software triggers

Data events selected at L0 are forwarded to the software trigger algorithms. Both of the software trigger algorithms, described in the two sections that follow, are compiled C++ programs. Copies of the two programs run on a dedicated farm of computers located at ground-level above the experimental cavern.

The computer nodes on the dedicated software trigger farm are able to write events from the L0 stage to hard-disk, for a limited time [61]. Since the LHC only delivers stable beams for $\sim 30\%$ of the time, these events are processed between fills of the LHC. About 20% of the events accepted by L0 are 'deferred' in this way. The thresholds of the L0 trigger are therefore able to be somewhat looser than they would otherwise be.

The first-level software trigger

The VELO reconstruction algorithms are fast enough to be run over the events output from L0, however the full detector event reconstruction is not. The first software trigger (HLT1) algorithm for `L0Muon` candidates is able to perform a fast muon identification, described as follows. For every track stem in the VELO, a search window is opened in M3 defined by the deflection of a track of at least $6\,\mathrm{GeV}/c$ in the bending plane. Hits found inside the search window in M3 are combined with the VELO track stem to define search windows in M2, M4, and M5. If any hits are found in the secondary search windows, the track is tagged as a provisional muon candidate. The quality of the provisional muon candidates is assessed using the χ^2 divided by the number of degrees of freedom (χ^2/ndf) of a track fit. The first provisional muon candidate is fit, if the χ^2/ndf is less than 25, the algorithm tags the track as a muon candidate and exits, otherwise the next provisional candidate track is fit.

A VELO track may be selected without being tagged as a muon if it has an large IP with respect to any primary vertex. However the minimum IP threshold is set high in order to keep the output rate manageable.

Following these selections 'forward tracking' is performed on both muon candidate tracks and large-IP VELO tracks. In the forward tracking procedure, minimum momentum and p_{T} are imposed on these tracks, defining search windows in the T1–3 stations. Hits found in the search windows in the tracking stations are attached to the track and the track is refit using a Kalman filter [63, 64]. This forward tracking algorithm is slightly simpler than the offline tracking algorithm for timing reasons. The invariant mass resolution of the $J/\psi \to \mu^+\mu^-$ decays is found to be $\sim 3\%$ larger than that achieved offline [60].

The algorithms used to trigger $B^0 \to K^+\pi^-\mu^+\mu^-$ events are the HLT1 track trigger path (`Hlt1TrackAllL0`) or the HLT1 muon track trigger path (`Hlt1TrackMuon`). The former selects all track candidates with a $p_{\mathrm{T}} > 1.6\,\mathrm{GeV}/c$. The latter selects forward tracks with a lower threshold $p_{\mathrm{T}} > 1\,\mathrm{GeV}/c$ but also requires the `IsMuon` variable. Both algorithms require a minimum IP from the primary vertex. The efficiency of `Hlt1TrackMuon` is shown to the left of Fig. 3.13.

The second-level software trigger

The second software trigger (HLT2) runs forward-tracking over all VELO tracks with $p_{\mathrm{T}} > 500\,(300)\,\mathrm{MeV}/c$ in 2011 (2012) [60,61]. The `IsMuon` variable is evaluated for all tracks and is used to select muon candidates. With this level of reconstruction: track quality, vertex quality, and event topology are used to select events consistent with containing a b hadron decay.

The analysis described in this thesis requires candidate events pass either dedicated exclusive muon algorithms (`Hlt2SingleMuon`, `Hlt2DiMuonDetatched`) or any of six topological algorithms.

The single muon algorithm requires either a very high p_{T} ($10\,\mathrm{GeV}/c$) muon candidate, or a good quality candidate with a somewhat lower p_{T} which is also inconsistent with originating at the PV. Furthermore, this algorithm is pre-scaled by a factor 0.5 to reduce the rate to an acceptable level. The dimuon algorithm requires two reasonable quality muon candidate tracks which make a vertex and that both tracks are inconsistent with originating at the PV. For these tracks, the dimuon invariant mass ($m_{\mu\mu}$) is computed and required to be above $1\,\mathrm{GeV}/c^2$. The exact requirements are shown in Table 3.2.

The topological algorithms are somewhat more sophisticated and make use of a multivariate analysis (MVA) technique. The technique is based on a modification of the boosted decision tree (BDT) [65, 66] algorithm known as 'Bonsai' boosted decision tree (BBDT) [67]. The BBDT classifier forms bins of continuous distributions and trains the BDT algorithm in a discrete way. This has the advantage of being very fast to evaluate the BBDT response for a set of tracks forming an n-body topology candidate. There are six topological BBDT algorithms for 2-, 3-, and 4-body topologies with or without one or more muon candidates (`Hlt2TopoMuNBodyBBDT` or `Hlt2TopoNBodyBBDT`). The topological algorithms do not require that all of the b hadron daughter tracks are reconstructed into the n-body object. Allowing for this possibility means that the n-body invariant mass

Table 3.2: Muon selections required by the HLT2 algorithms relevant to this analysis. The impact parameter χ^2 (IPχ^2) is the change in χ^2 of a vertex fit when including a track (in this case, the PV). This quantity will be described in more detail in Section . Other abbreviations used here are described in the text and shown in Fig 3.4. Table adapted from Ref. [60].

	`Hlt2SingleMuon`		`Hlt2DiMuonDetatched`
`IsMuon`	True	True	True
Track IPχ^2 with the PV	> 200	–	> 9
Track p_{T}	$> 1.3\,\mathrm{GeV}/c$	$> 10\,\mathrm{GeV}/c$	–
Track χ^2/ndf	< 2	–	< 5
Dimuon FD	–	–	> 49
Dimuon vertex χ^2/ndf	–	–	< 25
Dimuon p_{T}	–	–	$> 1.5\,\mathrm{GeV}/c$
$m_{\mu\mu}$	–	–	$> 1\,\mathrm{GeV}/c^2$
Pre-scale	0.5	1.0	1.0

may be less than that of the b hadron, the BBDT is therefore trained with the addition of the 'corrected' mass,

$$m_{\mathrm{corr.}} \equiv \sqrt{m^2 + |p_{\mathrm{T}}{'}_{\mathrm{miss.}}|^2} + |p_{\mathrm{T}}{'}_{\mathrm{miss.}}|\,, \tag{3.4}$$

where m is the invariant mass of the n-body object, and $p_{\mathrm{T}}{'}_{\mathrm{miss.}}$ is the missing momentum transverse to the FD vector. More information regarding this variable may be found in Ref. [60], for example. As well as the invariant mass and corrected mass, the sum of the daughters' p_{T}, the IP, the distance of closest approach (DOCA) between the tracks, the minimum p_{T} of all of the tracks, and the FD are included in the classifier. These variables are binned and the BBDT algorithm is trained using these discrete variables. Candidates are accepted if a threshold of the BBDT response is met in the lowest of the n-body classifiers. That is: a 2-body candidate is formed, and `Hlt2Topo2BodyBBDT` is evaluated, if this is not passed then a third track is added to the candidate, `Hlt2Topo3BodyBBDT` is evaluated, and so on. More details may be found in Refs. [60,61,67]. At the time of writing, this novel technique of using a MVA classifier in the triggering is unique to LHCb. The performance of the topological algorithms is found to be better than equivalent cut-based selections. The efficiency is shown to the right of Fig. 3.13.

3.4.3 Summary of triggering requirements

Candidates are required to pass the trigger requirements listed in Table 3.3. At HLT2 the logical 'or' of the various algorithms is employed. At all stages the offline-candidates are required to have been selected by the trigger.

Figure 3.13: The efficiency to TOS for $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ for the `Hlt1TrackMuon`, amongst other algorithms not described (left), and the topological HLT2 algorithms (right). Figure taken from Ref. [60].

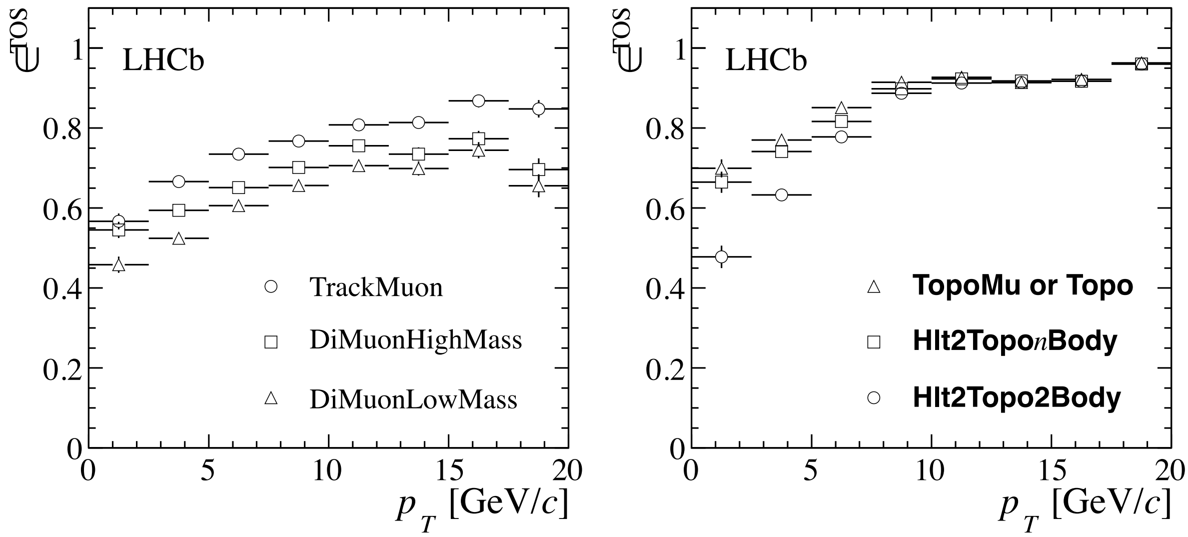


Table 3.3: Triggers required of candidate events.

Stage	Triggers	Note
L0	`L0Muon`	
HLT1	`Hlt1TrackAllL0` or `Hlt1TrackMuon`	
HLT2	`Hlt2SingleMuon` or `Hlt2DiMuonDetatched`	
	or `Hlt2TopoNBodyBBDT` or `Hlt2TopoMuNBodyBBDT`	$n = 2, 3, 4$

3.5 Generation of simulated data

The analysis described in this thesis makes use of simulated data (or *pseudodata*) events in several areas. Generation of simulated data makes use of existing generic open-source simulation software in a specific LHCb configuration [68].

In the first step, pp collision events are generated using Pythia 6.4 [69] and 8.1 [70] in approximately equal measures. The simulated $b\bar{b}$ quark pair is hadronised repeatedly until a target b hadron[4] is produced. The decay of the b hadron is described by EvtGen [71] and final-state radiation is generated using Photos [72]. The particles are then passed through a full simulation of the LHCb detector where the detector model, and the material interactions are implemented using Geant4 [73]. The data from the simulated detector is then stored and reconstructed as if it were real data.

[4] Of relevance to this measurement: B^0, B^0, B_s^0, and Λ_b^0.

Chapter 4

Selection of $B^0 \to K^+\pi^-\mu^+\mu^-$ candidate events

Candidate events which are selected by the trigger algorithms summarised in Table 3.3 are fully reconstructed offline. The present chapter describes the selection of $B^0 \to K^+\pi^-\mu^+\mu^-$ events following this reconstruction. The selection is developed with the goal of removing background events and preserving the yield of genuine $B^0 \to K^+\pi^-\mu^+\mu^-$ signal decays.

Background events may be separated into two categories: 'combinatorial' and 'specific'. Combinatorial background events are candidate four-body objects comprised of random combinations of tracks that happen to pass the preceeding selection stage. By contrast, specific background processes are candidates formed from similar B-hadron decays where tracks are misidentified, or the correct topology is misreconstructed.

The selection is applied in three main stages. The *preselection* stage, which reduces the size of the dataset and removes large contribution of combinatorial background events, is described in Section 4.1. Specific background processes are considered individually and criteria to reject such processes, so-called *vetoes*, are designed and implemented. The specific background rejection stage is described in Section 4.2. A multivariate classifier is trained in order to suppress combinatorial background events. The multivariate classifier stage is described in Section 4.3. Finally, the performance of the full selection chain is presented in Section 4.4.

4.1 Preselection stage

Candidate events are initially filtered by a set of preselection criteria. The preselection is designed to reduce the dataset to a manageable size. All preselected event and particle track information is stored in data files on the order of a few gigabytes in size, which should be compared to the full triggered and reconstructed dataset which is over three petabytes.

A simple requirement made at the preselection stage is that there be fewer than 600 hits in the SPD. This removes excessively noisy events with a lot of activity in the detector. The cartesian (x, y, z) position of the PV is required not to be excessively separated from the mean position. This removes interactions likely to originate from beam gas in the LHC pipe.

All tracks are required to be separated by 1 mrad from each-other (measured from the SV), and have a low probability of being *ghost tracks*. A ghost track is an erroneous combination of hits not produced by a single particle. A detailed description of the ghost probability algorithm can be found in Ref. [74]. Tracks are further required to lie inside the detector acceptance. These requirements ensure clean, well reconstructed tracks.

The remainder of the preselection requires kinematic and PID properties of the four-body object which are chosen to be consistent with a genuine $B^0 \to K^+\pi^-\mu^+\mu^-$ decay. A list of the full set of requirements may be found in Table 4.1. At the preselection stage, a window is placed on the full B^0 candidate invariant mass ($m_{K\pi\mu\mu}$) and an upper limit is placed on $m_{\mu\mu}$ and $m_{K\pi}$. Also, a requirement is made that ϑ (marked on Fig. 3.4) between the sum of the daughters' momenta and the FD vector be smaller than 14 mrad.

The `IsMuon` decision is required to be true for all of the candidate muon tracks, and the $\text{DLL}_{\mu\pi}$ is required to be larger than -3 (*i.e.* not excessively pion-like). The candidate pion track is required to have a $\text{DLL}_{K\pi}$ smaller than 25 (*i.e.* very loosely pion-like). The candidate kaon track is required to have a $\text{DLL}_{K\pi}$ greater than -5 (*i.e.* not excessively pion-like).

Finally, requirements are made on the impact parameter χ^2 (IPχ^2), and the flight distance χ^2 (FDχ^2) of vertices and tracks respectively. The former is defined as the change in χ^2 of a vertex fit when including a track. Much like a small IP, a smaller IPχ^2 is more consistent with that track originating from the vertex. Combinatorial combinations of $K\pi\mu\mu$ tracks will usually originate at the PV. The daughter candidate tracks are therefore required to have individually large IPs and IPχ^2s with respect to the PV and small IP and IPχ^2 with respect to the SV. The momentum vector of the four-body object, which is the vector sum of the daughters' momenta, should have a small IP and IPχ^2 with respect to the PV. This is since the genuine B^0 will originate at the PV. The FDχ^2 is defined, for two vertices, as the change in χ^2 when the two vertices are combined into a single vertex fit. Vertices formed by pairs of tracks, or by all four daughter tracks should have a large FD and FDχ^2 with respect to the PV. In some literature, the FDχ^2 is called the 'vertex separation χ^2'.

4.2 Specific background processes

Several specific backgrounds processes are considered. For each process, a sample of simulated events is generated. Using this sample, the total efficiency of misidentifying and selecting the process, ϵ_{bkg}, is measured. An estimate for the number of expected events

Table 4.1: Preselection requirements applied to triggered data. In this table: θ is the azimuthal beam angle, ϑ is the pointing angle, and $\langle x\rangle_{\mathrm{PV}}$, $\langle y\rangle_{\mathrm{PV}}$, $\langle z\rangle_{\mathrm{PV}}$ denote the mean PV position. All other variables are introduced in the text.

	Selection
Event SPD multiplicity	< 600
PV separation	$\lvert x - \langle x\rangle_{\mathrm{PV}}\rvert < 5\,\mathrm{mm}$
PV separation	$\lvert y - \langle y\rangle_{\mathrm{PV}}\rvert < 5\,\mathrm{mm}$
PV separation	$\lvert z - \langle z\rangle_{\mathrm{PV}}\rvert < 200\,\mathrm{mm}$
Tracks azimuthal angle	$(0 < \theta < 400)\,\mathrm{mrad}$
Tracks ghost prob.	< 0.4
Tracks minimum IPχ^2 with the PV	> 9
Track pair angular separation	> 1 mrad
Muon track	`IsMuon`
Muon track DLL$_{\mu\pi}$	> -3
Kaon track DLL$_{K\pi}$	> 5
Pion track DLL$_{K\pi}$	< 25
$m_{\mu\mu}$	$< 7100\,\mathrm{MeV}/c^2$
Dimuon vertex χ^2/ndf	< 9
$m_{K\pi}$	$< 6200\,\mathrm{MeV}/c^2$
Dihadron vertex χ^2/ndf	< 9
Dihadron vertex FDχ^2	> 9
Four-body invariant mass	$(4600 < m_{K\pi\mu\mu} < 7000)\,\mathrm{MeV}/c^2$
Four-body object IPχ^2 with the PV	< 16
Four-body object ϑ	$< 14\,\mathrm{mrad}$
Four-body vertex FDχ^2	> 121
Four-body vertex χ^2/ndf	< 8

Table 4.2: The values of the hadronsiation fractions used in estimates of specific backgrounds. A p_{T} dependence of the Λ_b^0 production fraction is observed [76] thus the last row of this table contains an approximate average.

	Value	Ref.
$\mathcal{B}\left[b \to B^0\right]$	$(40.1 \pm 0.8)\%$	[4]
$\mathcal{B}\left[b \to B^+\right]$	$(40.1 \pm 0.8)\%$	[4]
$\mathcal{B}\left[b \to B_s^0\right]$	$(10.3 \pm 0.9)\%$	[4]
$\mathcal{B}\left[b \to \Lambda_b^0\right]$	$(20 \pm 8)\%$	[76]

from each sample is then computed using,

$$N_{\mathrm{bkg,expected}} = \int \mathcal{L} \mathrm{d}t \cdot 2\sigma_{b\bar{b}} \cdot \epsilon_{\mathrm{bkg}} \cdot \mathcal{B}\left[b \to X\right] \cdot \mathcal{B}\left[X \to \mathrm{bkg}\right]. \tag{4.1}$$

In this expression, the total integrated luminosity $\int \mathcal{L} \mathrm{d}t = 3.05 \pm 0.11\,\mathrm{fb}^{-1}$. The production cross-section, $\sigma_{b\bar{b}}$, is measured in the LHCb detector acceptance and extrapolated to 4π using PYTHIA 6.4, this results in $\sigma_{b\bar{b}} = 284 \pm 69\,\mu\mathrm{b}$ [75]. The hadronisation fraction, denoted[1] by $\mathcal{B}\left[b \to X\right]$, is for the relevant B hadron/baryon and listed in Table 4.2. Finally, $\mathcal{B}\left[X \to \mathrm{bkg}\right]$ is the total branching fraction for the specific process. For example the hadronisation and total branching fraction for $B^0 \to J/\psi K^{*0}$, forming a background as described in Sections 4.2.1, would be

$$\mathcal{B}\left[b \to B^0\right] \cdot \mathcal{B}\left[B^0 \to J/\psi K^{*0}\right] \cdot \mathcal{B}\left[J/\psi \to \mu^+\mu^-\right] \cdot \mathcal{B}\left[K^* \to K\pi\right], \tag{4.2}$$

the hadronisation and total branching fraction for $\Lambda_b^0 \to \Lambda^*(1520)^0\mu^+\mu^-$, forming a background as described in Section 4.2.2, would be

$$\mathcal{B}\left[b \to \Lambda_b^0\right] \cdot \mathcal{B}\left[\Lambda_b^0 \to \Lambda^*(1520)^0\mu^+\mu^-\right] \cdot \mathcal{B}\left[\Lambda^*(1520)^0 \to pK\right]. \tag{4.3}$$

And the probability to misidentify tracks is folded into ϵ_{bkg} as applicable. In cases where the decay of a $K_1^*(892)^0$ is involved, isospin arguments show that $\mathcal{B}\left[K_1^*(892)^0 \to K^+\pi^-\right] = 2/3$, and $\mathcal{B}\left[K^*(892)^+ \to K^+\pi^0\right] = 1/3$.

Criteria are designed to reject these specific background processes. The estimated yields, computed using Eqn. 4.1, of each background before these criteria, and after the full selection are summarised in Table 4.3. The simulated events that are not excluded by the full selection are shown in Fig 4.1. This figure shows stacked histograms scaled to the estimated yields.

[1] In the literature, hadronisation fractions are often denoted f_X. This thesis follows the PDG convention, $\mathcal{B}\left[b \to X\right]$ [4].

Table 4.3: Estimated yields, and percentage relative to estimated signal yield, of specific background events. Yields and percentages of signal are quoted before and after the vetoes and selection detailed in this chapter. The processes are described in the corresponding sections given as a reference (charges are omitted from daughters for brevity).

		Before vetoes		After vetoes and selection	
Process	§	Estimated	% signal	Estimated	% signal
$\Lambda_b^0 \to \Lambda^* \mu\mu$	4.2.2	$(2.0 \pm 1.0) \times 10^3$	24 ± 10	$(1.0 \pm 0.5) \times 10^2$	1.2 ± 0.5
$B^0 \to J/\psi K^*$	4.2.1	75 ± 24	1.31 ± 0.26	64 ± 21	1.12 ± 0.24
$B_s^0 \to \phi\mu\mu$	4.2.3	$(1.0 \pm 0.4) \times 10^3$	18 ± 6	43 ± 18	0.75 ± 0.26
signal swaps	4.2.4	$(8.4 \pm 2.1) \times 10^2$	14.7 ± 1.2	28 ± 7	0.49 ± 0.05
$\Lambda_b^0 \to pK\mu\mu$	4.2.2	$(2.7 \pm 1.3) \times 10^2$	4.7 ± 2.0	13 ± 6	0.22 ± 0.09
$B^+ \to K\mu\mu$	4.2.6	70 ± 18	1.21 ± 0.13	4.4 ± 1.2	0.077 ± 0.011
$J/\psi K\pi$ swaps	4.2.5	$(3.7 \pm 1.0) \times 10^2$	6.4 ± 0.7	5 ± 4	0.09 ± 0.07

Figure 4.1: Distributions of simulated events shown as stacked histograms and scaled to the number of expected events that survive the vetoes described in this section. The distributions are shown *before* the application of the multivariate descriminator, described in Section 4.3, since this removes most of the simulated background samples' statistics. The channels shown are: $B^0 \to K_1^*(892)^0\mu^+\mu^-$ (signal) in **cyan**; signal mode with hadron swaps in **pink**; $B_s^0 \to \phi(1020)\mu^+\mu^-$ in **green**; $B^+ \to K^+\mu^+\mu^-$ in **red**; and samples of $\Lambda_b^0 \to pK^-\mu^+\mu^-$ and $\Lambda_b^0 \to \Lambda^*(1520)^0\mu^+\mu^-$ in **black**. Supplementary figures are included in Appendix A in Fig. A.1.

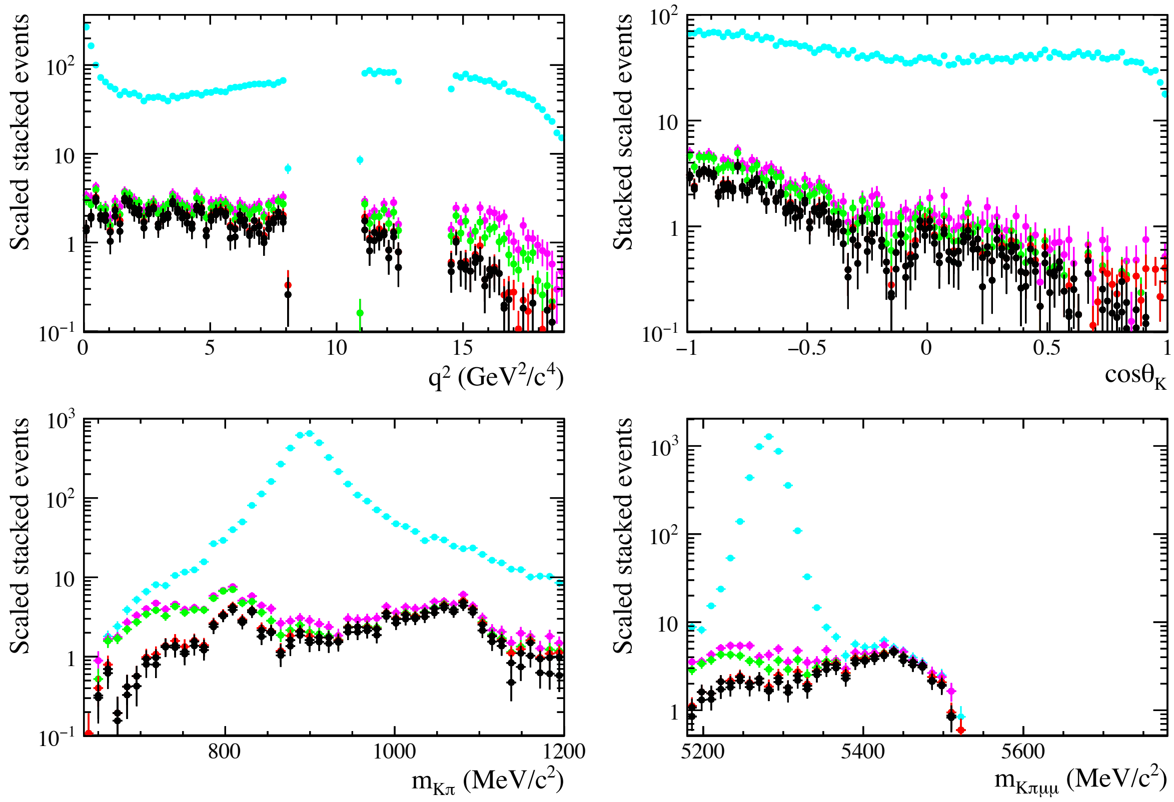
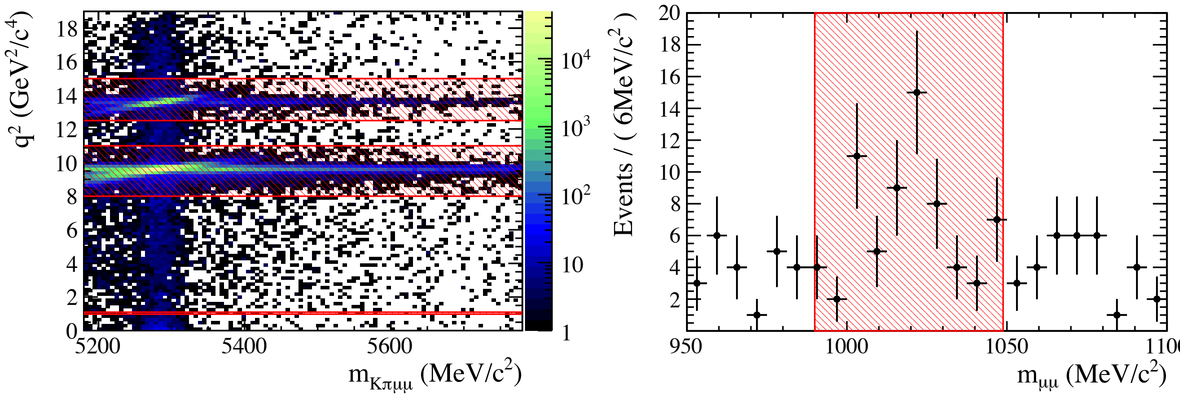


Figure 4.2: The $m_{K\pi\mu\mu}$-q^2 plane for selected data candidates (left) and the dimuon invariant mass spectrum in the vicinity of the $\phi(1020)$ (right). The red shaded regions of q^2 are excluded from the analysis due to the large contributions from $B^0_{(s)} \to J/\psi K^{*0}$ and $B^0_{(s)} \to \psi(2S)K^{*0}$. The region $q^2 \in [0.98, 1.1]\,\mathrm{GeV}^2/c^4$ is also excluded since weak evidence is seen for the decay $B^0_{(s)} \to \phi(1020)K^{*0}$ (right).



4.2.1 Contribution from correctly reconstructed quarkonia resonances

Decays of the type $B \to XK^*$ where X is some quarkonium resonance, which decays $X \to \mu^+\mu^-$, can form a background when correctly reconstructed. This is discussed in the present section. These quarkonia decays also form a background when particle tracks are misreconstructed. The misreconstructed contribution is discussed below in Section 4.2.5.

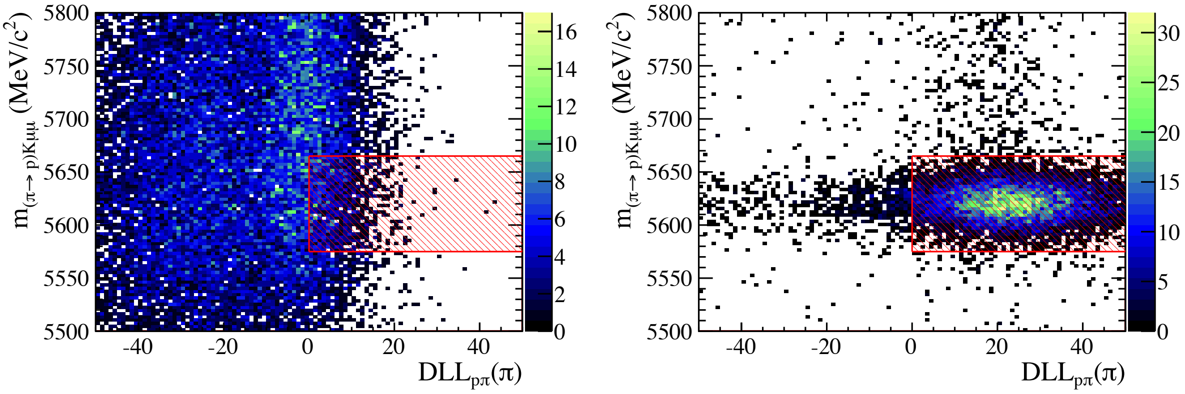
In the case of correctly reconstructed decays they are excluded from the analysis by removing regions of q^2. The decays: $B^0_{(s)} \to \phi(1020)K^{*0}$; $B^0_{(s)} \to J/\psi K^{*0}$; and $B^0_{(s)} \to \psi(2S)K^{*0}$ are excluded by removing: $q^2 \in [0.98, 1.1]\,\mathrm{GeV}^2/c^4$; $q^2 \in [8, 11]\,\mathrm{GeV}^2/c^4$; and $q^2 \in [12.5, 15]\,\mathrm{GeV}^2/c^4$, respectively. These exclusions are incorporated into the choice of binning scheme described in Section 6.5.1. The branching fraction for $B^0 \to \phi(1020)K^{*0}$ and $\phi(1020) \to \mu^+\mu^-$ is $(2.8 \pm 0.24) \times 10^{-9}$ [4] and interference between the $B^0 \to \phi(1020)K^{*0}$ and the penguin $B^0 \to K^{*0}\mu^+\mu^-$ diagrams may be significant. Therefore only weak evidence for the decay is observed. The excluded regions and the region of dimuon invariant mass in the vicinity of the $\phi(1020)$ are shown in Fig 4.2.

In order to select $B^0 \to J/\psi K^{*0}$ events, for use as control channel data, the region $q^2 \in [9.22, 9.96]\,\mathrm{GeV}^2/c^4$ is used. This bin is equivalent to a $60\,\mathrm{MeV}/c^2$ window around the nominal mass of the J/ψ, $m_{\mu\mu} \in [3036, 3156]\,\mathrm{MeV}/c^2$.

4.2.2 $\Lambda^0_b \to pK^-\mu^+\mu^-$

A background from $\Lambda^0_b \to pK^-\mu^+\mu^-$ decays, inclusive of all possible pK^- resonances, arises when the p is reconstructed as either of the hadron candidates. The dominant contribution to the $pK^-\mu^+\mu^-$ final state is expected to be from the $\Lambda^*(1520)^0 \to pK^-$

Figure 4.3: Planes of Λ_b^0 candidate mass ($m_{(\pi\to p)K\mu\mu}$) and pion candidate $\text{DLL}_{p\pi}$ for simulated $B^0 \to K_1^*(892)^0\mu^+\mu^-$ events (left) and simulated phase-space $\Lambda_b^0 \to \Lambda^*(1520)^0\mu^+\mu^-$ events (right). The red shaded box shows the area that is excluded by the veto (Eqns 4.4 and 4.5).



resonance. However there is evidence for other structure in the data, as can be seen in Fig. 4.4 (right).

Decays of this type are rejected by making alternative mass hypotheses for a candidate and removing events in regions of mass and $\text{DLL}_{p\pi}$, as detailed below. The background from $\Lambda_b^0 \to pK^-\mu^+\mu^-$ can occur in two ways.

The simplest is when the p is reconstructed as the "pion" candidate. In this case, changing the hypothesis of the pion track; from the pion mass to the proton mass; and remaking the invariant mass, $m_{(\pi\to p)K\mu\mu}$, will give an invariant mass consistent with that of the Λ_b^0 baryon. The contribution from these events is removed by vetoing events with $m_{(\pi\to p)K\mu\mu}$ around the nominal Λ_b^0 mass where the "pion" candidate track has a proton-like $\text{DLL}_{p\pi}$. Candidates are removed if:

$$(5575 < m_{(\pi\to p)K\mu\mu} < 5665)\,\text{MeV}/c^2 \tag{4.4}$$

and

$$\text{DLL}_{p\pi}(\pi) > 0. \tag{4.5}$$

Figure 4.3 shows the $m_{(\pi\to p)K\mu\mu}$ and $\text{DLL}_{p\pi}(\pi)$ plane for simulated signal and phase-space $\Lambda_b^0 \to \Lambda^*(1520)^0\mu^+\mu^-$.

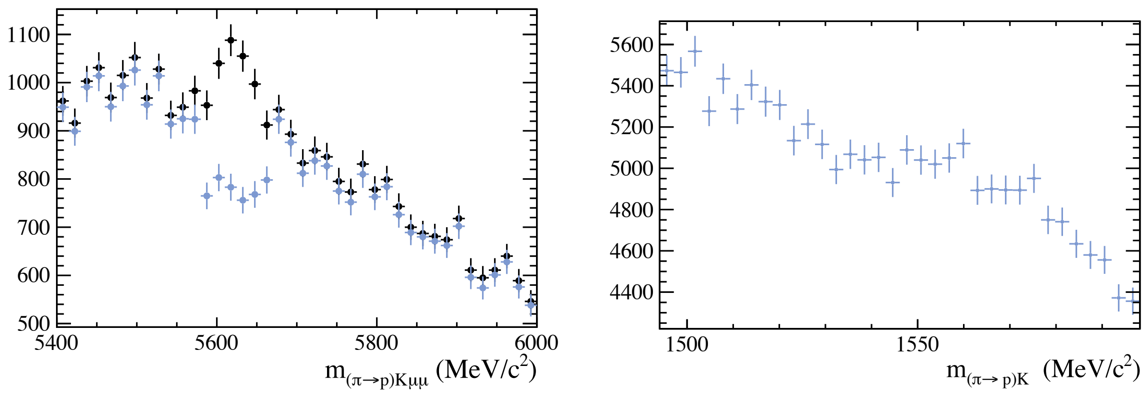
The second way $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays can contribute is when the p is reconstructed as the "kaon" candidate and the K^- is reconstructed as the "pion". In this case, both hadron track mass hypotheses need to be changed. The resulting invariant mass, $m_{(K\to p)(\pi\to K)\mu\mu}$, will similarly be consistent with the Λ_b^0 baryon. This contribution is removed by a similar requirement on $m_{(K\to p)(\pi\to K)\mu\mu}$ and by requiring that the pion candidate has a kaon-like $\text{DLL}_{K\pi}$. That is:

$$(5575 < m_{(K\to p)(\pi\to K)\mu\mu} < 5665)\,\text{MeV}/c^2 \tag{4.6}$$

and

$$\text{DLL}_{K\pi}(\pi) > 0. \tag{4.7}$$

Figure 4.4: The $m_{(\pi\to p)K\mu\mu}$ (left) and $m_{(\pi\to p)K}$ (right) spectra in selected data candidates. The black points show the spectrum with the preselection and vetoes used in the analysis presented in Ref. [1]. There is evidence for a peak at the nominal Λ_b^0 mass. The blue points are the spectra after the vetoes described in Section 4.2.2.



The planes of $m_{(K\to p)(\pi\to K)\mu\mu}$ and $\mathrm{DLL}_{K\pi}(\pi)$ (*i.e.* the equivalent to Fig. 4.3), as well as figures justifying the DLL cut values used, are included in Appendix A, Figs A.2 and A.3.

4.2.3 $B_s^0 \to \phi(1020)\mu^+\mu^-$

The decay process $B_s^0 \to \phi(1020)\mu^+\mu^-$ where the $\phi(1020) \to K^+K^-$ forms a background if either of the kaons are misidentified as pions. If a true $\phi(1020)$ decay is misidentified as $K^{*0} \to K\pi$ then changing the mass hypothesis of the "pion" track to that of a kaon and remaking the dihadron ($m_{(\pi\to K)K}$) and four-body object ($m_{(\pi\to K)K\mu\mu}$) invariant masses should be consistent with a $\phi(1020)$ and the B_s^0 respectively.

Candidates are therefore rejected if all of

$$(5321 < m_{(\pi\to K)K\mu\mu} < 5411)\,\mathrm{MeV}/c^2; \tag{4.8}$$

$$(1010 < m_{(\pi\to K)K} < 1030)\,\mathrm{MeV}/c^2; \tag{4.9}$$

$$\mathrm{DLL}_{K\pi}(\pi) > -10; \tag{4.10}$$

are fulfilled, or if

$$(5321 < m_{(\pi\to K)K\mu\mu} < 5411)\,\mathrm{MeV}/c^2; \tag{4.11}$$

$$(1030 < m_{(\pi\to K)K} < 1075)\,\mathrm{MeV}/c^2; \tag{4.12}$$

$$\mathrm{DLL}_{K\pi}(\pi) > 10. \tag{4.13}$$

The former veto (Eqns 4.8–4.10) removes candidates with a fairly kaon-like $\mathrm{DLL}_{K\pi}$ for the "pion" track and that peak near the nominal $\phi(1020)$ mass. The latter veto (Eqns 4.11–4.13) is designed to remove candidates in the tail of the $\phi(1020)$ distribution that will sit under the $K_1^*(892)^0$ when misreconstructed, in this case a stronger requirement is made on $\mathrm{DLL}_{K\pi}$ of the "pion".

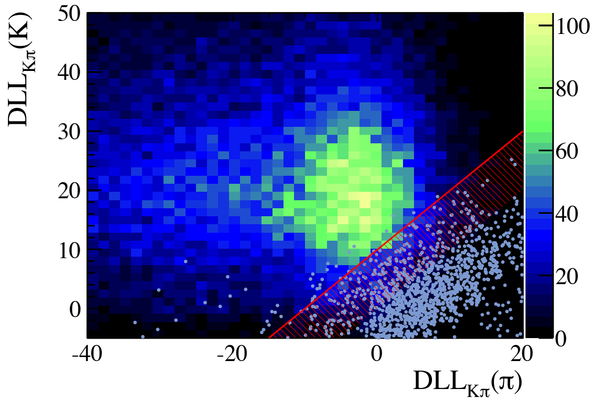


Figure 4.5: Simulated signal candidates (color scale) overlaid with simulated hadron PID swap events (grey points). The red line is at $\mathrm{DLL}_{K\pi}(K) - \mathrm{DLL}_{K\pi}(\pi) = 10$. Events to the lower right of this line, indicated by the direction of the shading, are excluded.

4.2.4 Hadron hypothesis swaps

Signal decay events and control channel events can form a background if the kaon and pion are misidentified as a pion and kaon respectively. These '*hadron PID swap*' events are removed with a requirement on the difference in the $\mathrm{DLL}_{K\pi}$ of the hadrons. The selection imposed on candidates is that

$$(\mathrm{DLL}_{K\pi}(K) - \mathrm{DLL}_{K\pi}(\pi)) > 10. \tag{4.14}$$

This has the effect of excluding events where the "kaon" candidate looks more pion-like than the "pion" candidate. The excluded region forms a diagonal region in the plane of the $\mathrm{DLL}_{K\pi}$ variable, as shown in Fig. 4.5.

4.2.5 Misreconstructed $B^0 \to J/\psi\, K^+\pi^-$ and $B^0_s \to J/\psi\, K^+\pi^-$

In addition to hadron hypothesis swaps, $B^0 \to J/\psi\, K^+\pi^-$ decays can form a background if either of the hadron candidates are misidentified as a muon and the muon of corresponding charge is misidentified as the hadron. These hadron-muon swaps are only relevant for decays where the quarkonium resonance is the J/ψ, since the branching fraction of $B^0 \to J/\psi\, K^+\pi^-$ is so large that the double misidentification becomes significant.

For these processes, changing the hadron mass hypothesis to a muon and remaking the invariant mass with the other muon candidate (either $m_{(\pi\to\mu)\mu}$ or $m_{(K\to\mu)\mu}$) will give an invariant mass consistent with a $J/\psi \to \mu^+\mu^-$ decay. For the case of the pion, candidates are rejected if

$$3036 < m_{(\pi\to\mu)\mu} < 3156\,\mathrm{MeV}/c^2 \tag{4.15}$$

and the either $\mathrm{DLL}_{\mu\pi}(\pi) > 5$ or the pion track fulfills `IsMuon`. For the case of the kaon, candidates are rejected if

$$3036 < m_{(K\to\mu)\mu} < 3156\,\mathrm{MeV}/c^2 \tag{4.16}$$

and the either $\mathrm{DLL}_{\mu\pi}(K) > 5$ or the kaon track fulfills `IsMuon`.

4.2.6 $B^+ \to K^+\mu^+\mu^-$

A background contribution can be formed if a pion from elswhere in the event is combined with the decay $B^+ \to K^+\mu^+\mu^-$ forming a four-track final state. In this case, the $K\mu\mu$ invariant mass should be consistent with originating from a B^+ decay. Furthermore adding a pion track will effect the upper side of the distribution of $m_{K\pi\mu\mu}$. The upper side of $m_{K\pi\mu\mu}$ is used in training the multivariate classifier (Section 4.3) and included in fits in order to control the background component. Candidates are removed if

$$mkpimm > 5380\,\mathrm{MeV}/c^2 \tag{4.17}$$

$$(5220 < m_{K\mu\mu} < 5340)\,\mathrm{MeV}/c^2. \tag{4.18}$$

4.3 Multivariate classifier

In the final step of the selection chain, a multivariate analysis (MVA) technique is employed to suppress combinatorial background events and any remaining specific background processes. In general, many individual variables will separate signal events from background. Defining a list of selection criteria based on these variables, such as the preselection listed in Table 4.1, does not usually exploit all of the available information. Extra separating power can be achieved using MVA methods which take correlations between the variables into account.

For this analysis, the boosted decision tree (BDT) algorithm [65,66] is the specific MVA algorithm used. The implementation of the BDT algorithm is taken from the toolkit for multivariate analysis (TMVA) package [77]. A brief description of the algorithm is given in Section 4.3.1. The algorithm is trained to select control channel $B^0 \to J/\psi\, K^{*0}$ data events (as a proxy to the $B^0 \to K^{*0}\mu^+\mu^-$ signal) and reject background proxy events from the region $m_{K\pi\mu\mu} \in [5350, 7000]\,\mathrm{MeV}/c^2$ in data. The statistical power of the technique is maximised using the k-folding procedure which is described in Section 4.3.2. The data samples are prepared for training as described in Section 4.3.3. The algorithm requires a set of input variables. A description of the 13 variables chosen for this analysis is described in Section 4.3.4. The performance of the training is presented in Section 4.3.5.

4.3.1 Description of the boosted decision tree algorithm

The BDT algorithm is an extension of the decision tree algorithm. A single decision tree is a sequence of splits chosen to maximise separation of signal from background for a given set of variables. The algorithm requires a *training sample* of events flagged as 'signal' (s) or 'background' (b)[2]. This training sample is then split into subsamples, or *nodes*. The splitting is based on one variable at a time and is chosen to maximise some measure

[2] To be explicit: roman 's' and 'b' is used to avoid any possible confusion with the b- and s-quarks.

of separation at each node. The splitting continues until some termination criterion is fulfilled, typically having achieved a predefined level of signal purity, P, or node depth. Once termination criterion is fulfilled at each node, the terminated *leaf nodes* are classified as signal (S) or background (B), based on the dominant portion of events in each. A schematic example of a single decision tree is shown in Fig. 4.6. Misclassified signal events that end up in a background leaf node and vise versa are given a larger weighting and a subsequent tree is trained. This procedure is known as *boosting*. Boosting and training a new tree is repeated to obtain a *forest* of decision trees. The final BDT classifier is the average of the individual decision trees in the forest.

The algorithm used in this analysis trains decision trees with a subset of four out of the 13 variables chosen at random. Each node splitting is decided according to the `Gini` scheme, The scheme requires the definition,

$$\texttt{Gini} \equiv \left(\sum_i w_i \right) P(1-P), \tag{4.19}$$

where the sum runs over all events in the node (subsample), each with a weight, w_i. The purity of each sample is defined as

$$P \equiv \frac{\sum_{\mathrm{s}} w_{\mathrm{s}}}{\sum_i w_i} = \frac{\sum_{\mathrm{s}} w_{\mathrm{s}}}{\sum_{\mathrm{s}} w_{\mathrm{s}} + \sum_{\mathrm{b}} w_{\mathrm{b}}} \tag{4.20}$$

where the sum in the denominator runs over all events in the node, and the sum in the numerator over only the events flagged as signal. The split into two daughter nodes is then chosen to maximise the quantity

$$\texttt{Gini}_{\text{parent}} - \texttt{Gini}_{\text{daughter 1}} - \texttt{Gini}_{\text{daughter 2}}. \tag{4.21}$$

The configuration used for this analysis, is that trees are terminated at a maximum depth of three nodes, and/or leaf nodes are classified as signal like for purity > 0.5. Misclassified events are given a new event weight, defined by the adaptive boosting procedure [66]. In this boosting scheme, a common multiplicative weight for all of the misclassified events in the tree,

$$\beta \ln \left[\frac{1 - R_{\text{mis.}}}{R_{\text{mis.}}} \right] \tag{4.22}$$

is calculated. The factor β is a tunable parameter. In the configuration used in this analysis $\beta = 0.05$ is chosen. The rate of misclassification is

$$R_{\text{mis.}} \equiv \frac{\sum_{\mathrm{s|B}} w_{\mathrm{s|B}} + \sum_{\mathrm{b|S}} w_{\mathrm{b|S}}}{N}, \tag{4.23}$$

where N is the sum of the weights of all events in the tree and the numerator contains the sum of signal events classified as background (s|B) and background as signal (b|S). The

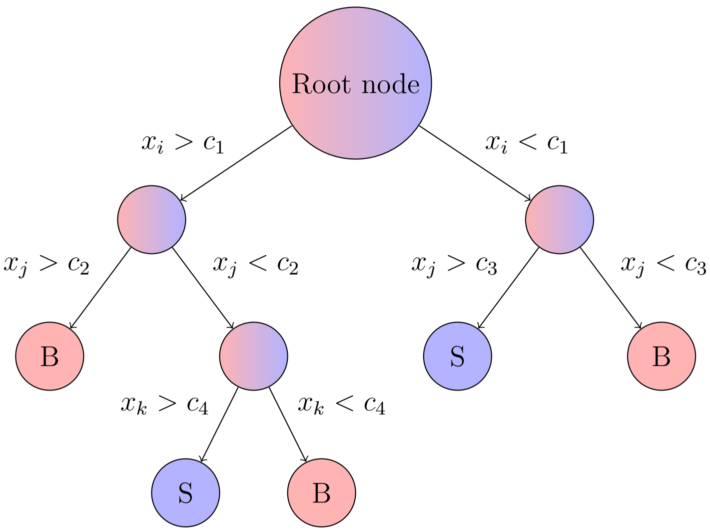


Figure 4.6: A schematic example of a single decision tree. The training sample of signal and background events are split in order to maximise some measure of separation at each node. The splitting continues until some termination criterion is fulfilled and the sample is designated as a signal node (S) or a background node (B). In a boosted algorithm, events that are misclassified (signal events that end up in a background node and vice versa) are assigned a higher weighting and the decision tree is retrained.

weights are updated:

$$w_{\mathrm{s|B}} \longmapsto \beta \ln \left[\frac{1-R_{\mathrm{mis.}}}{R_{\mathrm{mis.}}}\right] \cdot w_{\mathrm{s|B}} \qquad w_{\mathrm{b|S}} \longmapsto \beta \ln \left[\frac{1-R_{\mathrm{mis.}}}{R_{\mathrm{mis.}}}\right] \cdot w_{\mathrm{b|S}} \tag{4.24}$$

$$w_{\mathrm{s|S}} \longmapsto w_{\mathrm{s|S}} \qquad w_{\mathrm{b|B}} \longmapsto w_{\mathrm{b|B}} \tag{4.25}$$

and then rescaled in such a way as to preserve the total weight of the sample. Following the weighting of misclassified events, a new tree is trained with a different random subsample of four the variables. This is repeated 2000 times. The average of the 2000 trees in the forest is taken as the BDT classifier.

4.3.2 *k*-Folding the data samples to maximise statistics

The limiting factor in such an MVA technique is typically the low statistics available for the training sample. The available statistics are further reduced by the need to evaluate the performance of the algorithm on a second data sample that has not been used by the algorithm for training. The use of this so-called *testing sample* ensures that the algorithm is not biased by the statistical fluctuations of the training sample provided. This effect is known as *overtraining*. Furthermore, the BDT classifier obtained by the procedure described above is a continuous variable. A single 'cut' threshold is required which requires an optimisation on an unbiased dataset. For this analysis in particular, the algorithm suffers from low statistics in the sample of background proxy events.

In order to combat this problem, the k-folding technique is employed. The available data are split into $k = 10$ subsamples. The BDT algorithm described above is trained and tested on 90% of the data. The resulting BDT classifier is evaluated on the remaining 10%

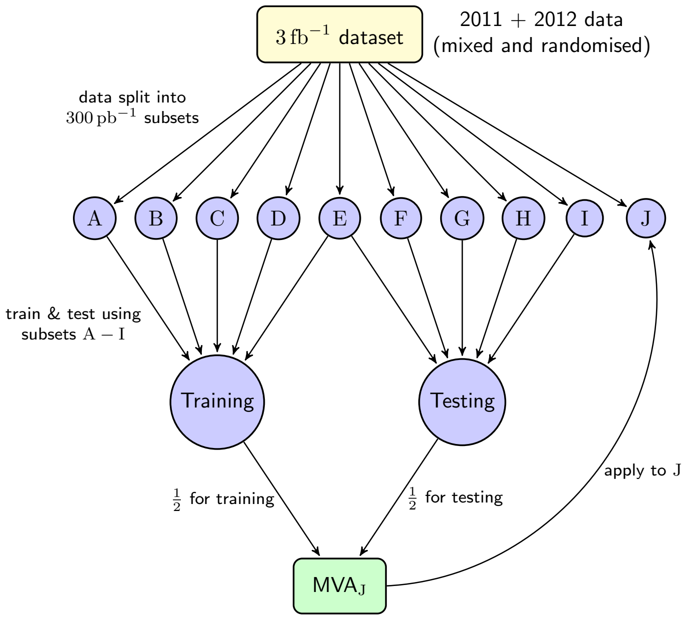


Figure 4.7: A schematic diagram showing the data flow with the k-folding technique. The full data sample is split randomly into ten subsamples. Ten individual classifiers are trained and tested/optimised each using 90% of the data. Each classifier is assigned to the remaining subsample that was not used in training or testing.

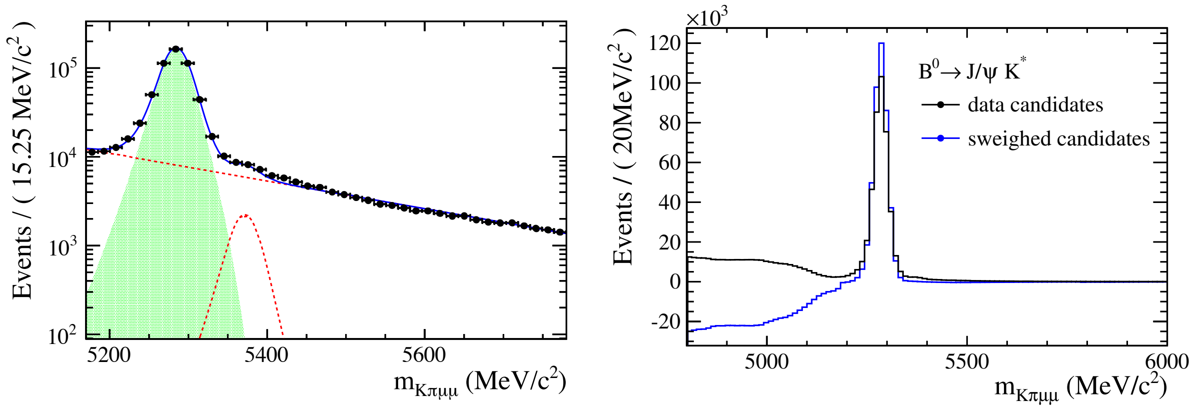
of the data. This procedure is repeated and ten statistically independent BDT classifiers are obtained, each one is written only to the 10% subsample on which it was not trained. In this way, all of the available background events are used for training but may also be used in subsequent stages of the analysis without bias. The 10-folding procedure is shown schematically in Fig 4.7.

4.3.3 Sample preparation using the sPlot technique

The sPlot technique [78] is used to prepare the signal proxy sample prior to the BDT algorithm. This technique is employed in order to remove other components from the signal proxy sample. Namely the combinatorial background component to and the $B_s^0 \to J/\psi K^{*0}$ component which contaminate the sample. The candidate events are fit with the sum of four Crystal Ball (CB) functions [79] and an exponential function to describe the background component in the signal proxy sample. The CB function, F_{CB}, consists of a Gaussian core with a power-law tail. It is described precisely in Appendix B.1 in terms of the Gaussian width, σ, the mean m_B, the tail power n, and threshold of transition α. Two of the four CB functions describe the $B^0 \to J/\psi K^{*0}$ events, a so-called *double-CB*, and the other two describe the $B_s^0 \to J/\psi K^{*0}$ component. In a double-CB function, F_{dCB}, two different width parameters combined with a relative fraction, f_{σ_1}. The definition used in this thesis is,

$$\begin{aligned} F_{\text{dCB}}(m_{K\pi\mu\mu}; m_B, \vec{\alpha}) \equiv & f_{\sigma_1} \cdot F_{\text{CB}}(m_{K\pi\mu\mu}; m_B, \alpha, n, \sigma_1) \\ & + (1 - f_{\sigma_1}) \cdot F_{\text{CB}}(m_{K\pi\mu\mu}; m_B, \alpha, n, \sigma_2). \end{aligned} \tag{4.26}$$

Figure 4.8: The result of the fit of Eqn. 4.28 to the control channel data (left) and the sPlot weights (right). The sPlot procedure defines an event weight based on the first component of Eqn. 4.28 (describing $B^0 \to J/\psi K^{*0}$ decays) and to suppress the other components. The data on the right are shown before (black histogram) and after applying the event weight (blue histogram).



The tail parameters, α and n are shared between the two. For brevity, the notation $\vec{\alpha}$ is introduced as a shorthand for a set of double-CB component parameters:

$$\vec{\alpha} \equiv \{\alpha,\, n,\, \sigma_1,\, \sigma_2,\, f_{\sigma_1}\}\,. \tag{4.27}$$

The $B_s^0 \to J/\psi K^*$ component is described with an identical function F_{dCB} except that the mean m_B is replaced with the nominal mass of the B_s^0 meson. The full model is then,

$$\begin{aligned}(1 - f_{B_s^0}) \cdot n_{J/\psi K^*} \cdot F_{\mathrm{dCB}}(m_{K\pi}, m_{B^0}, \vec{\alpha})\\ + f_{B_s^0} \cdot n_{J/\psi K^*} \cdot F_{\mathrm{dCB}}(m_{K\pi}, m_{B_s^0}, \vec{\alpha})\\ + n_{\mathrm{bkg}} \cdot \frac{e^{p_0 m_{K\pi\mu\mu}}}{\int e^{p_0 m_{K\pi\mu\mu}} \mathrm{d}m_{K\pi\mu\mu}}\end{aligned} \tag{4.28}$$

where the components are, in order, the B^0 component; the B_s^0 component; and the combinatorial background component.

Event weights are then generated based on the probability that the event contained a $B^0 \to J/\psi K^{*0}$ decay. The $B_s^0 \to J/\psi K^*$ and combinatorial background events are suppressed by small or negative weights. The results of the fitting and sPlot weighing procedure is shown in Fig 4.8. In addition to the BDT training, the sPlot event weights are also used to derive and assess corrections to account for mismodelling in the simulated samples. Corrections to the simulated data are described in Section 5.1.

4.3.4 Classifier inputs

The choice of input variables for the BDT algorithm drive the performance. The more information provided to the algorithm, the better the resulting classifier will be. However, inclusion of certain variables is known to have a biasing effect on the variables of interest.

For the purposes of this analysis, the selection should not bias the $m_{K\pi\mu\mu}$, $m_{\mu\mu}$, or $m_{K\pi}$ spectra, nor the helicity angles ($\cos\theta_K$, $\cos\theta_\ell$, ϕ). The choice of variables to use as inputs is therefore a balance between maximal selection power and whilst incurring minimal bias.

Variables such as the momenta of the individual particles, or the IP of individual tracks to the vertex are avoided since these are correlated with the angles, despite the fact that they are powerful when separating signal from background. Variables related to the four-body object are preferred since they are generally more weakly correlated to the variables of interest [1]. The pointing angle of the four-body object; ϑ (depicted in Fig 3.4); the χ^2/ndf and lifetime, τ, of the four-body vertex; and the scalar momentum and p_T of the four-body object are all uncorrelated to the variables of interest [1]. These variables each provide good separation power individually. The signal proxy, and background dominated sample distributions are shown in Fig. 4.9.

A substantial improvement in performance is also observed by the inclusion of PID information for each of the track types. For the hadrons, the $\mathrm{DLL}_{K\pi}$ variables are included and for the muon candidates, the $\mathrm{DLL}_{\mu\pi}$. The distributions of these PID are shown in Fig. 4.10.

Track isolation variables

The performance of the BDT algorithm is improved with the inclusion of track isolation variables [80]. Isolation variables are derived for the hadron and muon tracks which make up the $B^0 \to K^+\pi^-\mu^+\mu^-$ candidate. All tracks not assigned to the signal $B^0 \to K^+\pi^-\mu^+\mu^-$ candidate; *i.e.* tracks that are not tagged as the kaon, pion or either muon; are considered as a potential tertiary decay from each of these candidate tracks. This is shown schematically in Fig. 4.11. The isolation variables used here are a count of the number of these 'extra' tracks that could reasonably have originated from a tertiary decay. Signal candidate tracks with a low 'isolation count' are 'isolated'. Tracks with a high count are more likely be the result of a poorly reconstructed cascade-type decay.

For hadrons: the number of extra tracks which can form a tertiary vertex, which passes the criteria in Table 4.4 is recorded. This provides the isolation count variable which describes how isolated the hadron tracks are [80].

For muon tracks: the quantities listed in Table 4.4 as well as the IP, p_T, and χ^2/ndf of the extra track are combined into a *track-wise BDT*. The track-wise BDT is trained to select or reject single tracks (in contrast to the main BDT for *event* selection). The muon isolation BDT is trained on simulated signal tracks to reject simulated background tracks. The resulting muon isolation count is the number of extra tracks, for each signal muon track in an event, that are selected by the muon isolation BDT > 0.15 (the optimum selection). This procedure uses techniques developed for Ref. [81] and described in more detail in Ref. [82][3]. The samples of simulated background tracks were generated for the

[3] Sections 6.5 and 6.6 in Ref. [82].

Figure 4.9: The kinematic variables related to the four-body object that are given as input to the BDT algorithm. The distributions are shown for weighted signal proxy candidates (blue) and background dominated candidates (red).

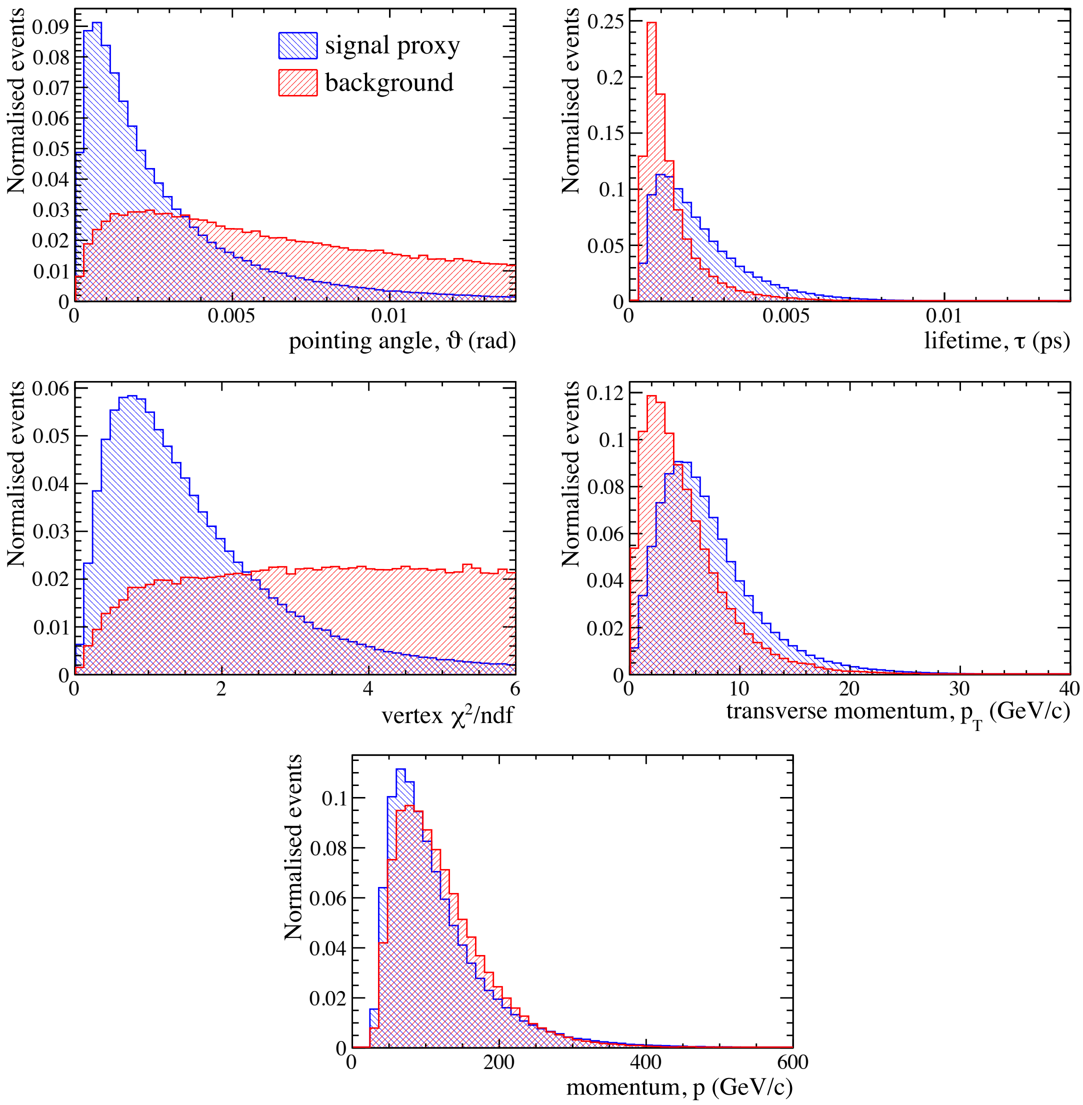
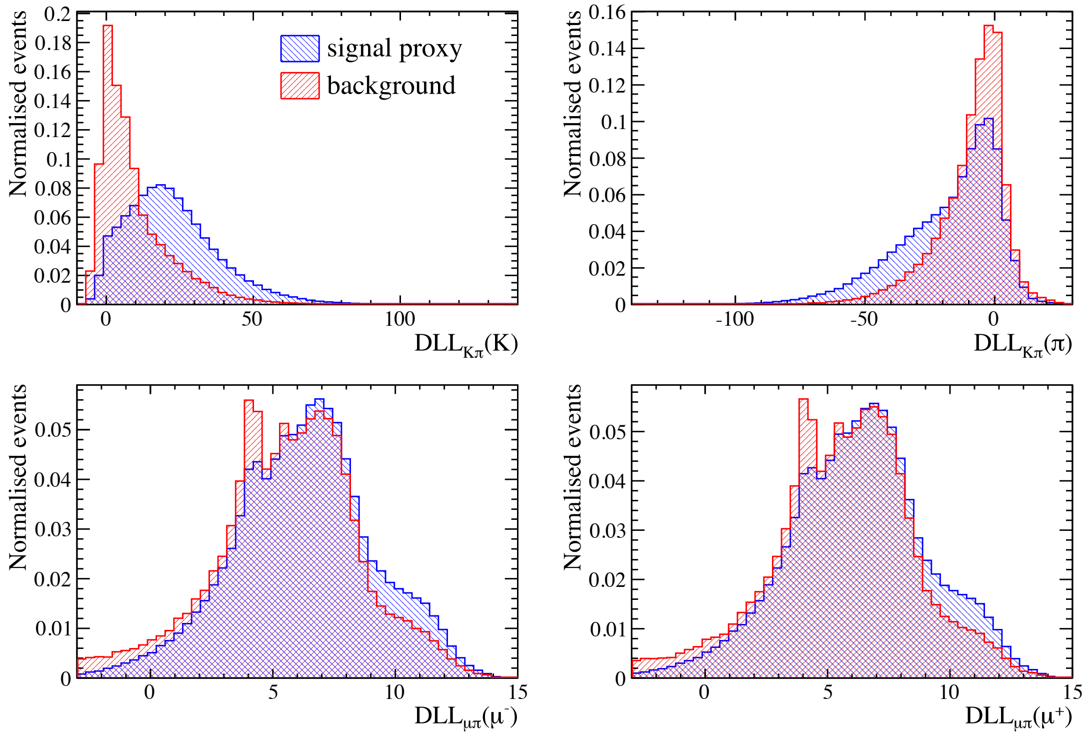


Figure 4.10: The hadron and muon candidate PID variables that are given as input to the BDT algorithm. The distributions are shown for weighted signal proxy candidates (blue) and background dominated candidates (red).



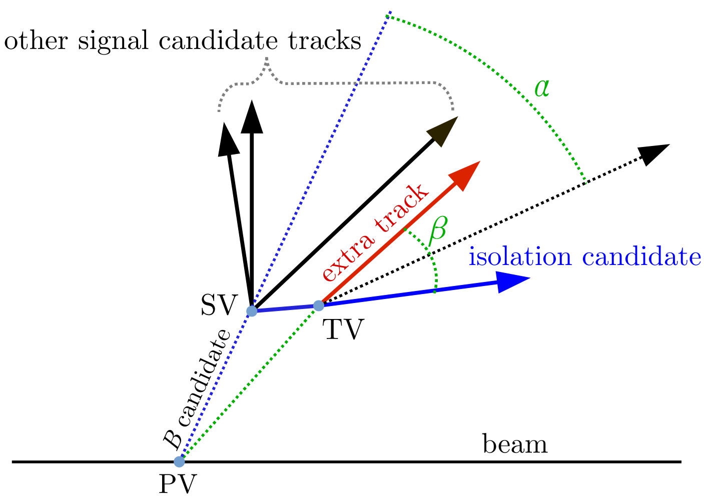


Figure 4.11: A schematic diagram showing the addition of an extra track (marked in red) to form a tertiary vertex (TV) The number of extra tracks that may be added to each signal candidate track (marked in blue) fulfilling some selection criteria is recorded. The angle β is the angle between the hadron candidate and the extra track. The angle α is the angle between the original line of flight and the vector momentum of the hadron candidate combined with the extra track.

Table 4.4: The selection applied to extra tracks forming a tertiary vertex (TV) with a signal candidate hadron track as shown in Fig. 4.11. The number of tertiary vertices selected according to these criteria is recorded for each signal candidate hadron track. This provides an integer variable for each event which describes how isolated the hadron tracks are. The angle β is the angle between the hadron candidate and the extra track. In the last row, the quantity is built from the vector momenta and transverse momenta of the hadron candidate and the track. The angle α is the angle between the original line of flight and the vector momentum of the hadron candidate combined with the extra track. The angles α and β are marked in green on Fig. 4.11.

	Selection
Distance between the PV and the TV	$(0.5 < (\mathrm{TV} - \mathrm{PV}) < 40)\,\mathrm{mm}$
Distance between the SV and the TV	$(-0.15 < (\mathrm{TV} - \mathrm{SV}) < 30)\,\mathrm{mm}$
The DOCA between the hadron and the extra track	$< 0.13\,\mathrm{mm}$
The IP in units of its resolution	> 3
The angle between the hadron and the extra track	$\beta < 0.27\,\mathrm{mrad}$
$\dfrac{\lvert\vec{p}_h + \vec{p}_{\mathrm{track}}\rvert\,\alpha}{\lvert\vec{p}_h + \vec{p}_{\mathrm{track}}\rvert\,\alpha + p_{\mathrm{T}h} + p_{\mathrm{T\ track}}}$	< 0.6

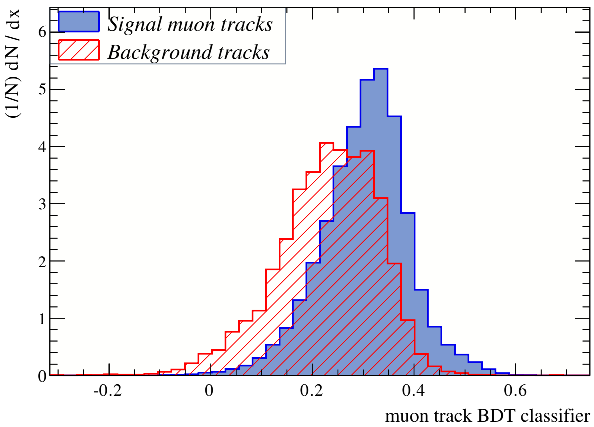


Figure 4.12: The muon isolation BDT classifier distribution. The y-axis of the histogram shows the (normalised) number of tracks. The muon isolation variable is the number of tracks that pass a cut of 0.15. Figure from Ref. [80] following the procedure described in [82].

analysis presented in Ref. [83]. The track-wise BDT algorithm is configured in a similar way to the main BDT for event selection (described in Section 4.3.1) except that the trees are allowed to go five nodes deep before termination, and the forest contains 850 trees from which the average is made. The track-wise muon isolation BDT classifier is shown in Fig. 4.12.

The distributions of isolation count variables for both track types are shown in Fig. 4.13. These distributions are inputs to the event selection BDT algorithm.

4.3.5 Results

The BDT algorithm is run with the above configuration and input variables (Figs 4.9, 4.10, and 4.13). The training results are shown in Fig. 4.14 in the form of a reciever operating

Figure 4.13: The hadron and muon candidate integer isolation variables. The hadron variable is the number of tracks in the event that pass the selection given in Table 4.4. The muon variable is the number of tracks in the event which have a muon isolation BDT classifier response > 15. The distributions are shown for weighted signal proxy candidates (blue) and background dominated candidates (red).

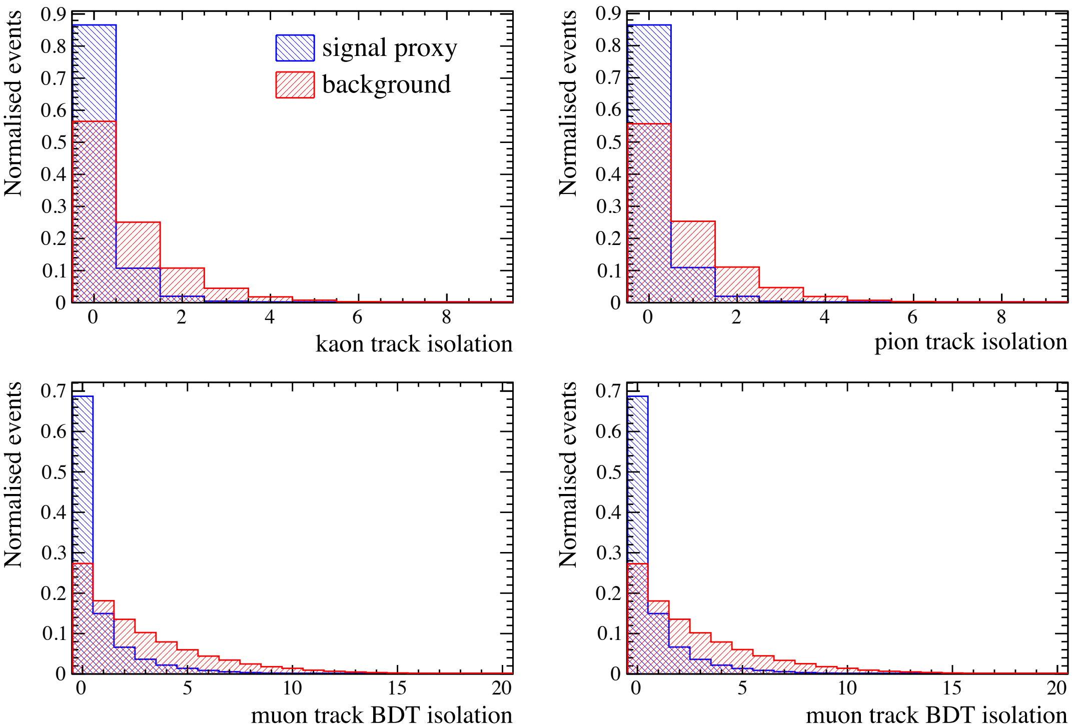
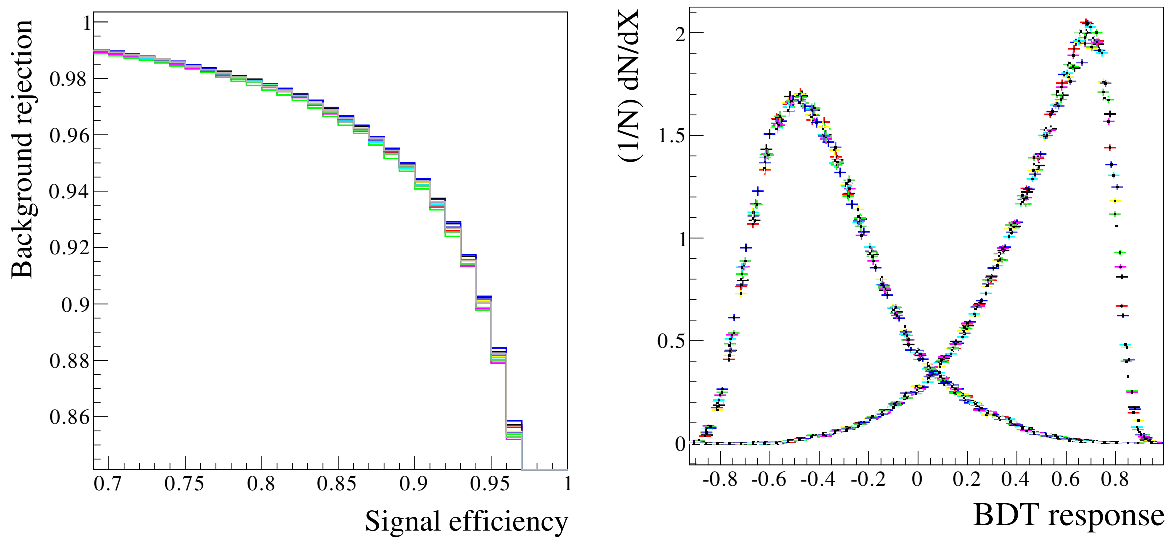


Figure 4.14: The ROC curves (left), and classifier distributions (right) for all ten folds overlaid. In the right-hand plot, the peaks correspond to the classifier response for background (signal) where the peak is centered at around -0.5 (0.7). Figure from Refs [80].



characteristic (ROC) curve. A ROC curve is any plot of a true-positive rate, against a false positive rate and maps the performance of a boolean selection (such as signal versus background) for a given discriminating variable. A more severe curve indicates better performance since it maps a lower false-positive probability for a higher true-positive probability, whereas more gentle curves indicate a worse performance.

The value of the BDT classifier at which to select events is found by an optimisation procedure. For each value of a cut on the BDT, the signal yield, $n_{K^{*}\mu\mu}$, for the range $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$ is estimated by performing a fit to control channel $B^0 \to J/\psi\, K^{*0}$ data events in a window around the nominal B^0 mass: $m_{K\pi\mu\mu} \in [5230, 5350]\,\mathrm{MeV}/c^2$. The yield is scaled by the product of the ratio of the selection efficiency (excluding the BDT cut) and the ratio of branching fractions,

$$\frac{\mathcal{B}\left[B^0 \to K^{*0}\mu^+\mu^-\right]}{\mathcal{B}\left[B^0 \to J/\psi\, K^{*0}\right]}. \tag{4.29}$$

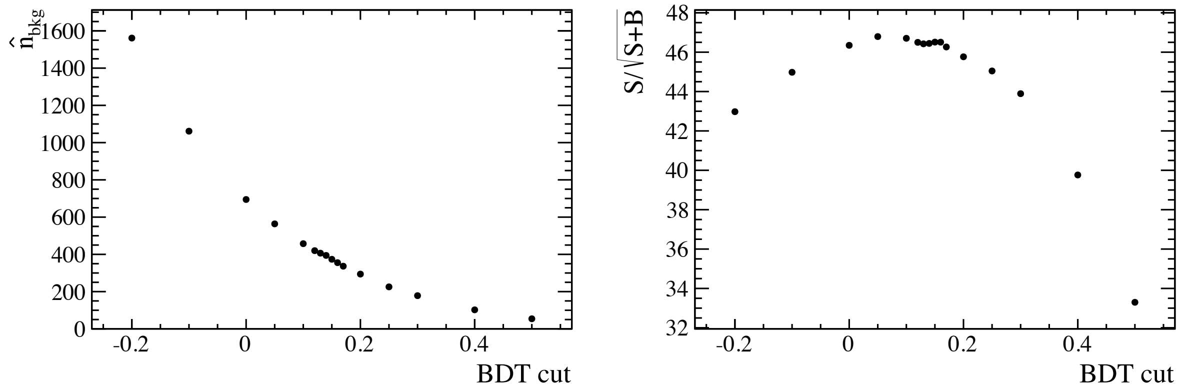
The background yield, n_{bkg}, is estimated by fitting the sum of two exponential functions below ($m_{K\pi\mu\mu} \in [5000, 5180]\,\mathrm{MeV}/c^2$) and above ($m_{K\pi\mu\mu} \in [5500, 7000]\,\mathrm{MeV}/c^2$) the signal region for $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$.

For each value of the BDT cut, the figure of merit,

$$\frac{\hat{n}_{K^{*}\mu\mu}}{\sqrt{\hat{n}_{K^{*}\mu\mu} + \hat{n}_{\mathrm{bkg}}}} = \frac{\mathrm{S}}{\sqrt{\mathrm{S}+\mathrm{B}}}, \tag{4.30}$$

is evaluated. The hat notation is used to denote that these are estimated quantities. Following this procedure, the selection cut, BDT > 0.2 is chosen. A plot of the estimated

Figure 4.15: Estimated background yield (left) and the figure of merit (right) as a function of the selection cut of the BDT classifier. An aggressive cut is chosen at BDT > 0.2 to reduce the effect of the background and to guard against the possibility of miscalibration in the extrapolation procedure used to obtaining the expected background yield. This threshold is the same as the angular analysis presented in Ref. [29].



background yield, and the significance in steps of incremental BDT cut is shown in Fig. 4.15.

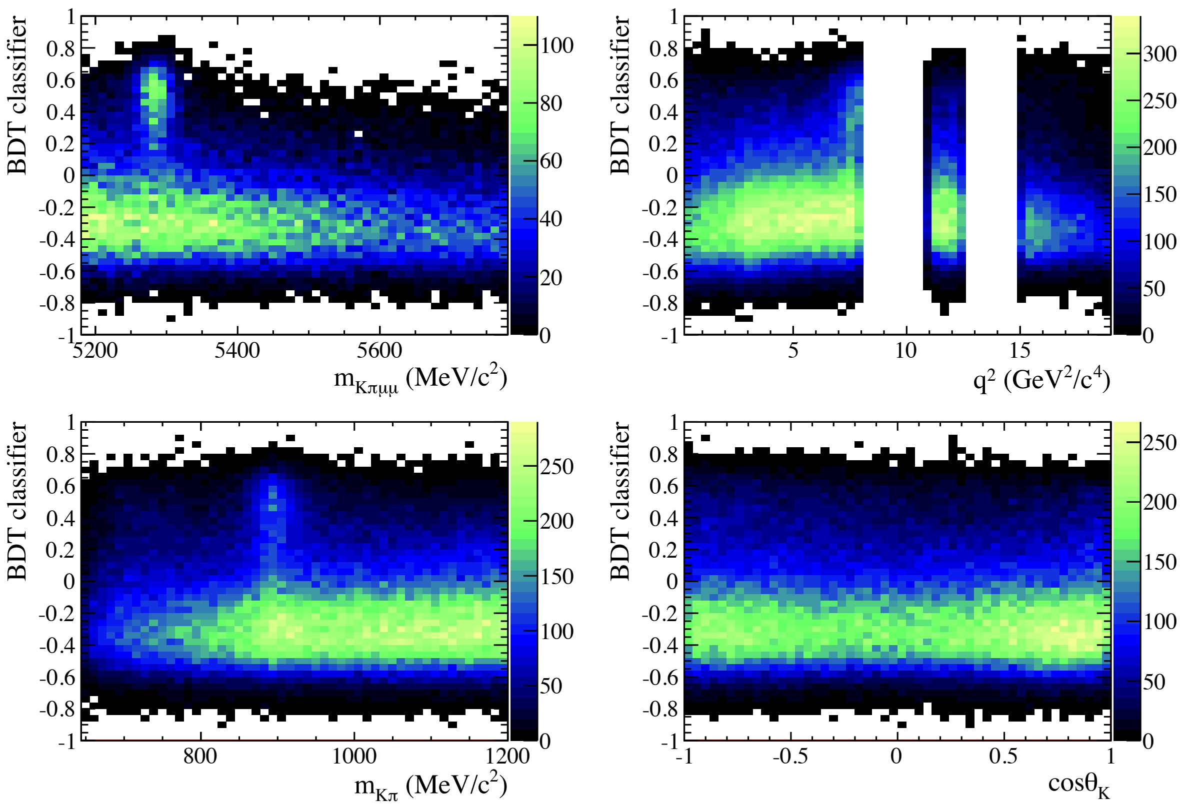
4.4 Selection performance

The efficiency to select signal and control channel events is calculated using simulated signal and control channel events. This is shown in Table 4.5. A discussion of the angular efficiency shape, and the relative efficiency binned in q^2 is given in Chapter 5. Figure 4.16 shows the correlation between several parameters and the BDT classifier. No significant correlations are observed.

Table 4.5: Efficiencies of various stages of the selection as measured on simulation for the signal and control channel. The efficiencies are split into absolute (abs.) and relative (rel.), namely the number of selected simulation events with respect to the original number simulated and with respect to the output of the previous stage respectively. Uncertainties are due to statistics of the pseudodata sample at each stage.

	$B^0 \to K_1^*(892)^0\mu^+\mu^-$		$B^0 \to J/\psi\, K_1^*(892)^0$	
stage	abs. (%)	rel. (%)	abs. (%)	rel. (%)
LHCb acceptance	15.3 ± 0.01	15.3 ± 0.01	14.9 ± 0.01	14.9 ± 0.01
reconstruction and preselection	1.30 ± 0.004	8.5 ± 0.03	1.32 ± 0.003	8.9 ± 0.02
triggers	1.02 ± 0.004	78.2 ± 0.14	1.06 ± 0.003	80.2 ± 0.09
hadron swap (Eqn. 4.14)	0.92 ± 0.004	90.3 ± 0.11	0.96 ± 0.003	90.9 ± 0.08
all other specific vetoes (§4.2)	0.89 ± 0.004	96.7 ± 0.07	0.92 ± 0.003	96.0 ± 0.05
BDT	0.80 ± 0.003	89.7 ± 0.12	0.83 ± 0.002	90.2 ± 0.08

Figure 4.16: The data distributions for the BDT classifier against several variables of interest for this analysis. No significant correlation effects are observed.



Chapter 5

Describing the efficiency using simulated events

Knowledge of the efficiency to detect $B^0 \to K^+\pi^-\mu^+\mu^-$ decays is required in two main steps of the analysis presented in this thesis. The efficiency in terms of the variables of interest is required when performing a fit to extract the S-wave component. And the ratio of efficiencies for $B^0 \to K^+\pi^-\mu^+\mu^-$ relative to $B^0 \to J/\psi K^{*0}$ events, binned in q^2, is required when normalising the differential branching fraction. The efficiencies are taken from simulation in the form of large pseudodata samples.

These pseudodata samples are taken from the output of the algorithms described in Section 3.5. Before using the samples to extract the efficiency, they are first corrected for differences between data and simulation. The correction procedure is described in Section 5.1. The efficiency is extracted in two ways: the first method produces the binned ratio of efficiencies in q^2, and is described in Section 5.2; the second method results in a functional description of the angular efficiency and is described in Section 5.3.

5.1 Correcting for differences between data and simulation

The procedure described in Section 3.5 is, in general, able to reproduce realistic pseudodata. However significant differences are observed in several distributions. The symptoms of this mismodelling are corrected using control channel data. The corrections are derived in two steps and validated using control channel data and simulation. The first step involves correcting the DLL variables which are not, in general, well reproduced by the simulation. The correction procedure is to *re-sample* the DLL distributions based on high purity calibration samples of muon, kaon, pion and proton tracks. Re-sampling is described in Section 5.1.1. The second step is to derive an event weight to account for the discrepancy observed in the control channel samples of $B^0 \to J/\psi K^{*0}$ data and simulated events. The derivation of these event weights is described in Section 5.1.2. Finally, the improvement observed in several variables of interest is presented in Section 5.1.3.

5.1.1 Re-sampling the particle identification variables

The selection includes requirements on the $\mathrm{DLL}_{K\pi}$, $\mathrm{DLL}_{\mu\pi}$, and $\mathrm{DLL}_{p\pi}$ variables. The former two are included in the training of the BDT algorithm (Section 4.3.4). In simulated data, these variables rely on the simulation of the RICH subdetector. The RICH detector response is affected by the number of tracks in an event (N_{tracks}), the detector alignment, and the gas properties. Accurate simulation of the RICH response is difficult. Agreement between data and simulation in the PID variables is therefore rather poor. In order to achieve good agreement, the simulated DLL distributions are re-sampled from calibration sample events from the data.

Calibration data samples comprised of $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$, $\Lambda^0 \to p\pi^-$ and $J/\psi \to \mu^+\mu^-$ decays are used as a pure source of kaons and pions, protons, and muons respectively. These samples are obtained from data without requirements made on the PID variables, except for a requirement on $\mathrm{DLL}_{p\pi}$ of the proton tracks from $\Lambda^0 \to p\pi^-$ decays. The calibration samples are collected into 3D bins of η, N_{tracks}, and p_{T}. For each 3D bin, histograms of all of the DLL variables are stored as template distributions.

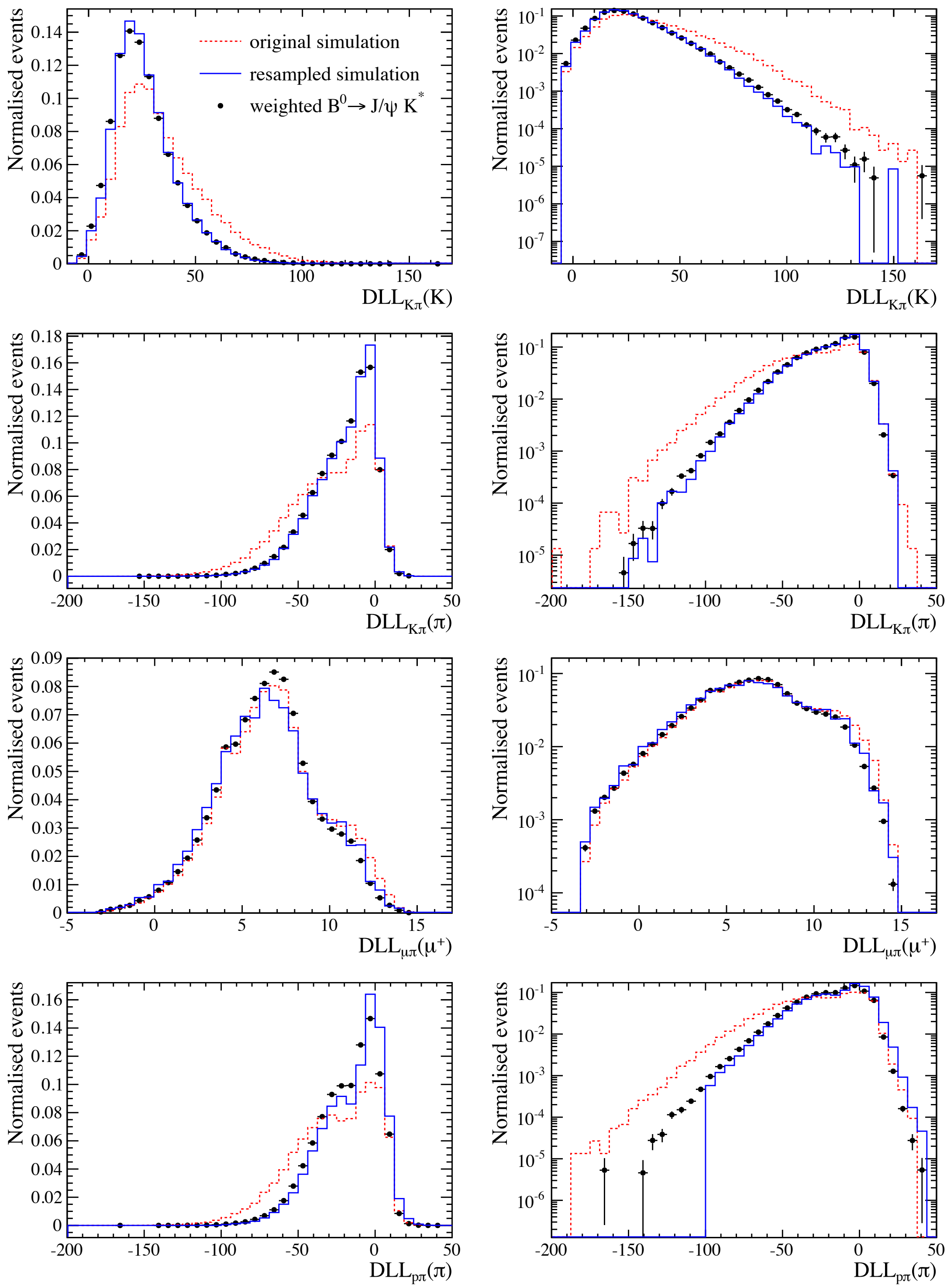
The re-sampling procedure is as follows. All signal candidate tracks in all simulated events are iterated over. For a given simulated track, the 3D bin corresponding to the η, N_{tracks} and p_{T} is recorded. Since the true identity of the simulated particle which generated the track is known, the correct calibration sample in the bin is retrieved. The histograms of each of the DLL variables of the corresponding track type are used to generate a new *re-sampled DLL* value. This re-sampled variable is used as a replacement for the original poorly simulated variable throughout.

The procedure improves the agreement between simulation and data in the PID variable distributions. However, it only preserves the correlations of the sampled variables (*i.e.* η, N_{tracks}, and p_{T}) and breaks any other correlations. Figure 5.1 shows the improvement achieved by the re-sampling by comparing re-sampled simulation, uncorrected simulation and sFit weighted $B^0 \to J/\psi\, K^{*0}$ data. The agreement found by re-sampling is found to be better than by weighting kinematic distributions of the daughters. An estimate of the systematic uncertainty incurred due to breaking these correlations is discussed in Section 7.3.8.

5.1.2 Event weights to correct for residual differences

An event weighting for all simulated samples is derived from the distributions of three kinematic variables that show differences between data and simulation. The event weights are calculated directly from the difference in distributions of sFit weighted $B^0 \to J/\psi\, K^{*0}$

Figure 5.1: Validation plots for the re-sampling procedure. Original simulated PID (red) is re-sampled (blue) and compared to sFit weighted $B^0 \to J/\psi K^{*0}$ data (black points). The boundary at $\text{DLL}_{K\pi}(\pi) < -100$ (bottom row) is due to a cut in the calibration samples. This variable is only used in the veto against Λ_b^0 peaking background, described in Section 4.2.2. The fraction of missing events amounts to $O(10^{-4})$ of the $B^0 \to J/\psi K^{*0}$ data.



data and the simulation. It should be noted that the $B^0 \to J/\psi K^{*0}$ data itself has a genuine S-wave component, whereas the simulation does not. This S-wave component in data is minimised by performing the comparison in a narrow window of $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$. The presence of this component in the data which is used to correct the pseudodata, results in small discrepancies in the kinematic distributions of the hadron momentum and p_T. This is treated as a source of systematic uncertainty, described in Section 7.3.3. The distributions from which the weights are derived are: N_tracks; the p_T of the four-body candidate B meson; and the χ^2/ndf of the $K\pi\mu\mu$ vertex fit. These distributions are chosen such that the effect of the selection, which might bias the distributions, is minimal. The weights are derived cumulatively: an event weight is derived for N_tracks which is then applied before deriving the weight for p_T and the product of the event weights for N_tracks and p_T is applied before deriving the weight for the vertex χ^2/ndf. The order of the variables in this procedure is chosen such that the largest discrepancy (in N_tracks) is corrected first, and so on. Although no difference is observed when changing the order.

These event weights are stored in histograms which are then used to weight all simulation samples. This procedure is illustrated in Fig. 5.2 which shows all three simulation distributions agreeing with the data by construction.

5.1.3 The effect on the multivariate classifier variable

The most important variable, in terms of correctly modelling the efficiency is the BDT classifier. The BDT classifier evaluated using re-sampled DLL variables and corrected by the event weights is shown in Fig. 5.3. A significant improvement in agreement with data is observed in the simulated BDT classifier.

Additional comparison plots for the isolation variables may be found in Appendix C in Fig. C.1.

5.2 Binned ratio of efficiencies

The measurement of the differential branching fraction result, discussed in Chapter 7, requires the ratio of efficiencies for $B^0 \to K^+\pi^-\mu^+\mu^-$ signal decays and $B^0 \to J/\psi K^{*0}$ decays. The ratio is required because the control channel $B^0 \to J/\psi K^{*0}$ is used to normalise the rate.

The ratio of efficiencies is taken from two simulated data samples. One sample is generated according to the full simulation chain described in Section 3.5 with a physics model based on the form factors given in Ref. [84]. The corrections described above are applied, and events are required to pass the selection requirements described in Chapter 4. This sample will be referred to as the selected and reconstructed (reco.) sample.

The second sample is generated using the Pythia and EvtGen steps, but without any simulation of the detector or its response. This sample is generated with the same physics

Figure 5.2: Illustration of the event weighting procedure. The original simulation (red) and sFit weighted $B^0 \to J/\psi\, K^{*0}$ data (black points) are used to derive weights for the simulated events. Applying these weights results in near-perfect agreement (blue) by construction. Minute differences are expected since the final weights are derived cumulatively. These weights are applied to all other selection-level simulation samples.

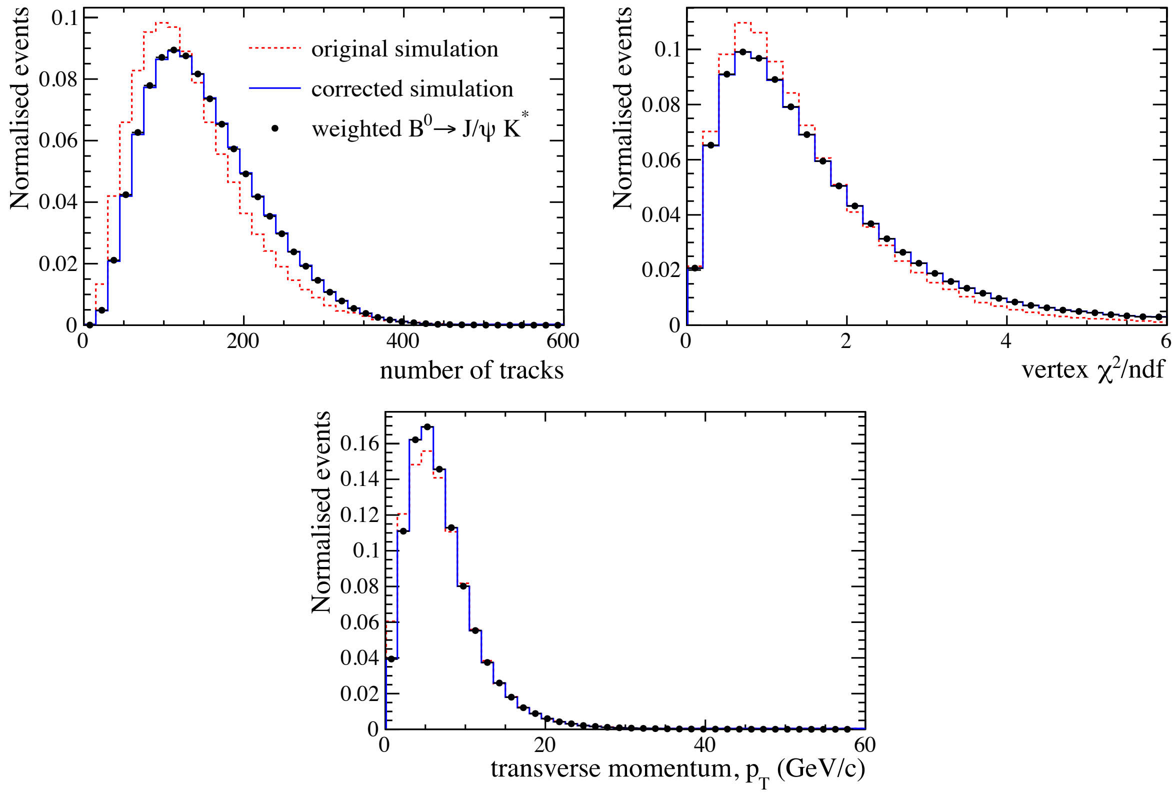
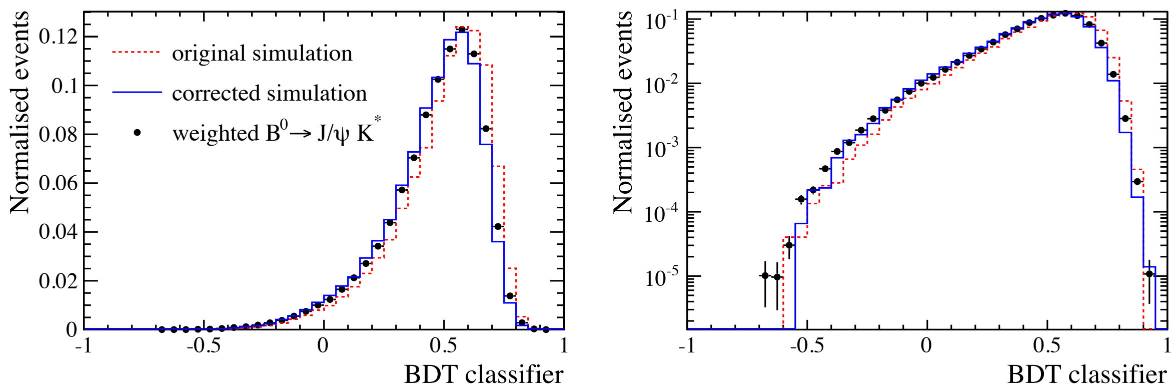


Figure 5.3: Comparison plot of the average BDT response on linear (left) and logarithmic (right) scales. The ten separate BDT variables (from k-folding) are averaged to form one response for simulated events. Uncorrected simulation (red) is weighted and re-sampled (blue) and compared to sFit weighted data (black points).



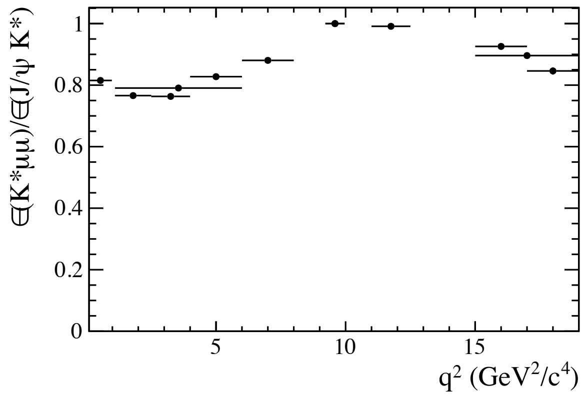


Figure 5.4: The ratio of efficiencies to reconstruct and select $B^0 \to K^+\pi^-\mu^+\mu^-$ and $B^0 \to J/\psi K^+\pi^-$ decays. This is obtained using generator-level and selected and reconstructed simulation samples. The horizontal error bar shows the q^2 bin size (Section 6.5.1) hence overlapping bins. The vertical uncertainty, not visible on this scale, is due to the statistics of the pseudodata samples. Several sources of systematic uncertainty (that contribute in addition to the pseudata sample statistics and are not shown) are discussed in Section 7.3.

model as the reco. sample. The sample will be referred to as the generator-level (gen.) sample.

The binned efficiency in q^2 is then the ratio of the reco. sample to the gen. sample for a given binning scheme. The ratio of efficiencies is simply the ratio of each q^2 bin to the bin $q^2 \in [9.22, 9.96]\,\mathrm{GeV}^2/c^4$. The resulting distribution is shown in Fig. 5.4.

Forming the efficiency ratio in this way implicitly integrates over all other variables. This introduces some model-dependence, since it relies on the assumption that the angular distribution in simulation is the same as that of the data. This model-dependence is assessed as a systematic effect and is described in Section 7.3.5.

5.3 Extraction of the angular efficiency

When performing the partial angular fit to extract the S-wave fraction, the physics is described by the angular distribution. However it is necessary to correct for biases induced by the LHCb detector geometry, the trigger and the selection. These effects, often referred to as *acceptance effects* elsewhere in the literature, have the effect of warping the angular distribution. For the analysis described in this thesis, the effects are corrected by multiplying the physical angular distribution, with the efficiency function[1],

$$\iiint \mathrm{d}\cos\theta_\ell \mathrm{d}\phi \mathrm{d}q^2{}_{\mathrm{bin}} \left[\frac{\mathrm{d}^5\Gamma}{\mathrm{d}q^2\mathrm{d}m_{K\pi}\mathrm{d}\cos\theta_K\mathrm{d}\cos\theta_\ell\mathrm{d}\phi} \times \epsilon(q^2, \cos\theta_K, \cos\theta_\ell, \phi, m_{K\pi})\right]. \quad (5.1)$$

The angular efficiency function $\epsilon(q^2, \cos\theta_K, \cos\theta_\ell, \phi, m_{K\pi})$ accounting for correlations between variables, is determined from simulated events.

Similarly to the extraction of the ratio of efficiencies, there are two simulated data samples required for this procedure (reco. and gen.). However no explicit physics model

[1]The triple integral makes the partial angular analysis (in variables $m_{K\pi}$ and $\cos\theta_K$) explicit.

is used, instead the samples are generated with a phase-space (phsp) model. In contrast to the ratio of efficiencies, this approach is model-independent since the efficiency is parameterised in terms of all relevant dimensions and the integral is explicitly performed.

The reco. sample follows the full chain of algorithms described in Section 3.5. Following this chain, a sample of 14×10^6 events, which are selected by the preselection stage, is generated. Since there are no model requirements in this case, the PYTHIA and EVTGEN steps are not required for the gen. sample. A gen. sample of 1.5×10^9 events is therefore generated with the phsp model.

Normally, to obtain an efficiency one would divide the distributions of gen. and reco. samples. However, the dimensions $\cos\theta_K$, $\cos\theta_\ell$ and ϕ are generated flat according to the phsp model description. It is therefore not necessary to divide the distributions by a gen. sample distribution, the distribution of the reco. sample in these dimensions is also the efficiency. By contrast, $m_{K\pi}$ and q^2 have a correlated 2D shape.

An event weighting procedure is employed which is equivalent to dividing the reco. $(m_{K\pi}, q^2)$ distribution by the gen. distribution. This simple division is not done performed as the event weighting results in a 5D un-binned sampling of the efficiency surface. The un-binned approach allows determination of the functional form of the efficiency without the need to implement a 5D histogram-fitting framework.

The event weighting procedure is as follows. The gen. sample is binned into a 2D histogram (with binned entries, $n(m_{K\pi}, q^2)$) from which a weight, $w_i^{(\mathrm{kin.}\to\epsilon)} \propto 1/n(m_{K\pi}, q^2)$, is derived. The 2D histograms for the gen. sample is shown in Fig. 5.5. Close to the kinematic edge of $(m_{K\pi}, q^2)$ the weights become arbitrarily large, as may be seen in the central hand plot of Fig. 5.5. A fiducial cut based on the analytically allowed phsp invariant mass function (or *triangle function* given in Appendix B.2) is therefore applied before applying the weights. This results in the more realistic distribution of weights shown in the right-hand plot of Fig. 5.5.

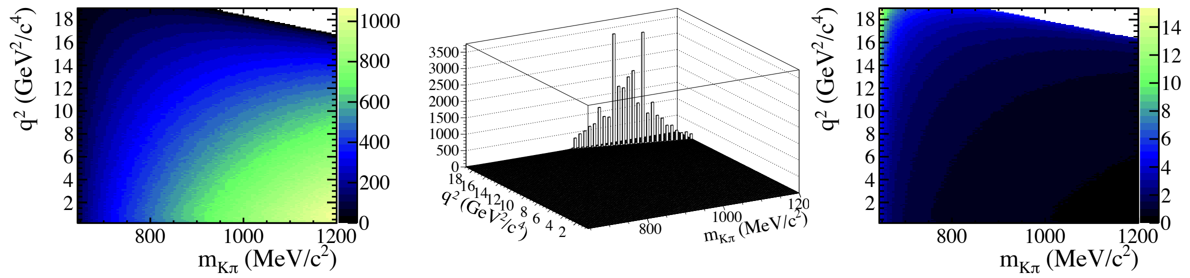
The reco. sample is then weighted by this efficiency weight, which converts the kinematic distributions of the reco. sample into the efficiency. The weights are scaled to preserve the number of data of the reco. sample. The method to extract the functional efficiency is described in the following Section 5.3.1. The results are presented in Section 5.3.2.

5.3.1 Method to extract the efficiency function

A functional description of the sampled efficiency is extracted using a parameter estimation technique called the method of moments (MoM)[2]. The MoM requires an orthogonal basis for which the set of Legendre polynomials of the first kind, $P_n(x)$ are chosen. They are

[2]The MoM technique is described in, for example, Ref. [85]. The technique may also be applied to extract physics parameters from angular distributions in data. Reference [86] describes this for the $B^0 \to K^{*0}\mu^+\mu^-$ angular distribution.

Figure 5.5: Two-dimensional ($m_{K\pi}, q^2$) histograms of the generator-level (gen.) kinematic distribution (left) and the inverted bin content distribution (centre). Simply inverting the bin contents results in unphysically large weights which overwhelm any structure in the rest of the ($m_{K\pi}, q^2$) plane. In order to remove the unphysically large weights in the inverted distribution, a fiducial cut is applied, resulting in the weight distribution (right). The weights in the central and right-hand plots have been scaled to preserve the number of events in the selection-level sample. The z-axis scale in both cases is therefore arbitrary.



defined by

$$P_0(x) = 1; \tag{5.2}$$

$$P_1(x) = x; \tag{5.3}$$

$$P_{n+1}(x) = \frac{1}{n+1}\left((2n+1)xP_n(x) - nP_{n-1}(x)\right); \tag{5.4}$$

and have the orthogonality condition,

$$\int_{-1}^{+1} P_m(x)P_n(x)\mathrm{d}x = \frac{2}{2n+1}\delta_{mn}, \tag{5.5}$$

which is valid for the domain $x \in [-1, 1]$.

Given that any function may be expressed as a decompostion into any orthogonal basis, the efficiency function is written

$$\epsilon = \sum_{ghijk} c_{ghijk} P_g(m_{K\pi}{}') P_h(q^{2\prime}) P_i(\cos\theta_\ell) P_j(\cos\theta_K) P_k(\phi'), \tag{5.6}$$

where the primes indicate a variable is transformed over some domain $y \in [a, b]$ to the domain $[-1, 1]$ by

$$y \longmapsto y' = 2\frac{y-a}{b-a} - 1. \tag{5.7}$$

The cosines are already in the required domain. The relation of Eqn. 5.6 is exact if the quintuple sum is infinite and provided that ϵ is a smooth continuous integrable function.

The coefficients of the expansion are determined using the relation:

$$\begin{aligned} c_{ghijk} = \iiiint\!\!\int & \mathrm{d}m'_{K\pi}\mathrm{d}q^{2\prime}\mathrm{d}\cos\theta_\ell\mathrm{d}\cos\theta_K\mathrm{d}\phi'\, \epsilon(m'_{K\pi}, q^{2\prime}, \cos\theta_\ell, \cos\theta_K, \phi') \\ & \times \Big[P_g(m'_{K\pi}) P_h(q^{2\prime}) P_i(\cos\theta_\ell) P_j(\cos\theta_K) P_k(\phi') \\ & \times \frac{(2g+1)}{2}\frac{(2h+1)}{2}\frac{(2i+1)}{2}\frac{(2j+1)}{2}\frac{(2k+1)}{2}\Big]. \end{aligned} \tag{5.8}$$

Which is obtained from Eqns 5.5 and 5.6[3].

However, the efficiency is not a known continuous integrable function; the efficiency function is sampled by the simulated events, as described above. The exact quintuple integral therefore becomes a discrete sum over weighted simulated events, e, which provides statistical estimator of the coefficients. Namely,

$$c_{ghijk} = \sum_{\text{events}, e} \Bigg[w_e^{(\text{kin.}\to\epsilon)} w_e^{(\text{data-sim.})} P_g(m'_{K\pi e}) P_h(q^{2\prime}_{\ e}) P_i(\cos\theta_{\ell e}) P_j(\cos\theta_{K e}) P_k(\phi'_{\ e}) \\ \times \frac{(2g+1)}{2}\frac{(2h+1)}{2}\frac{(2i+1)}{2}\frac{(2j+1)}{2}\frac{(2k+1)}{2} \Bigg], \tag{5.9}$$

where $w_e^{(\text{kin.}\to\epsilon)}$ and $w_e^{(\text{data-sim.})}$ are the weights which turn the kinematic sample into the efficiency (Section 5.3) and which correct for differences between data and simulation (Section 5.1) respectively. This procedure of summing over events to obtain estimators for the coefficients as shown in Eqn. 5.9, is the Legendre basis implementation of the MoM. It is also possible to calculate the variances between estimators [85] and form a covariance matrix. The variances are defined by

$$v^2_{ghijk,\,mnopq} = \sum_{\text{events}, e} \Bigg[w_e^{(\text{kin.}\to\epsilon)} w_e^{(\text{data-sim.})} P_g(m'_{K\pi e}) P_h(q^{2\prime}_{\ e}) P_i(\cos\theta_{\ell e}) P_j(\cos\theta_{K e}) P_k(\phi'_{\ e}) \\ \times \frac{(2g+1)}{2}\frac{(2h+1)}{2}\frac{(2i+1)}{2}\frac{(2j+1)}{2}\frac{(2k+1)}{2} \\ - c_{mnopq} \Bigg]^2 . \tag{5.10}$$

As a brief aside, the formulae above may be made more palatable by making some notational changes. The coefficients, c_{ghijk} may be ordered into a single list. Therefore a simple map of the five-dimensional indices to the one dimensional position in the list, I or J, may be defined. Making the definitions,

$$\underline{\Theta}_e \equiv \left\{ w_e^{(\text{kin.}\to\epsilon)},\ w_e^{(\text{data-sim.})},\ m'_{K\pi e},\ q^{2\prime}_{\ e},\ \cos\theta_{\ell e},\ \cos\theta_{K e},\ \phi'_{\ e} \right\} \tag{5.11}$$

and

$$\boldsymbol{P}_{ghijk}(\underline{\Theta}_e) \equiv w_e^{(\text{kin.}\to\epsilon)} w_e^{(\text{data-sim.})} P_g(m'_{K\pi e}) P_h(q^{2\prime}_{\ e}) P_i(\cos\theta_{\ell e}) P_j(\cos\theta_{K e}) P_k(\phi'_{\ e}) \\ \times \frac{(2g+1)}{2}\frac{(2h+1)}{2}\frac{(2i+1)}{2}\frac{(2j+1)}{2}\frac{(2k+1)}{2}, \tag{5.12}$$

Eqns 5.9 and 5.10 simplify to

$$c_I = \sum_{\text{events}, e} \boldsymbol{P}_I(\underline{\Theta}_e) \tag{5.13}$$

[3] A brief derivation is provided in Appendix C.2.

and

$$v_{IJ}^2 = \sum_{\text{events}, e} \left[\boldsymbol{P}_I(\underline{\Theta}_e) - c_J \right]^2. \tag{5.14}$$

By inspection of Eqn. 5.9 it is clear that the method requires some configuration. The choice of a set of maximum values $\{g, h, i, j, k\}$ for the expansion (Eqn. 5.6) is required. Choosing large values is computationally expensive in calculating the coefficients, more so in calculating the variances, and in evaluating and integrating the functional efficiency polynomial for use in a fit. It is also possible to generate a very high order polynomial efficiency function which erroneously accounts for the statistical fluctuations. Conversely, a low order polynomial is too coarse to correctly account for the shape of the efficiency distribution.

5.3.2 Resulting efficiency function

The Legendre method of moments is run including terms of all orders from zeroth to $(5, 6, 4, 7, 4)^{\text{th}}$ in $(m_{K\pi}, q^2, \cos\theta_\ell, \cos\theta_K, \phi)$ respectively.

One-dimensional projections of the efficiency sampling (the weighted simulated events) are shown with the integral of the polynomial function returned by the MoM in for the variables of interest in Fig 5.6. Similar plots for $\cos\theta_\ell$ and ϕ are included in Appendix C in Fig. C.2. Two-dimensional projections of the planes between variables of interest are shown alongside the integrated polynomial function in Fig 5.7. Similar plots of planes including $\cos\theta_\ell$ and ϕ are included in Appendix C in Figs C.3–C.5.

A good functional description is observed in the projections. The polynomial efficiency is used, as described in Chapter 6 and Eqn 6.8, to correct for the detector acceptance effects. The variances are used in assessing a contribution to the systematic uncertainty due to the statistics of the simulation sample. This is described in Section 7.3.

Figure 5.6: One-dimensional projections of the simulated efficiency distribution with the polynomial function returned by the Legendre method of moments (MoM) integrated and overlaid. Supplementary figures are included in Appendix C in Fig. C.2.

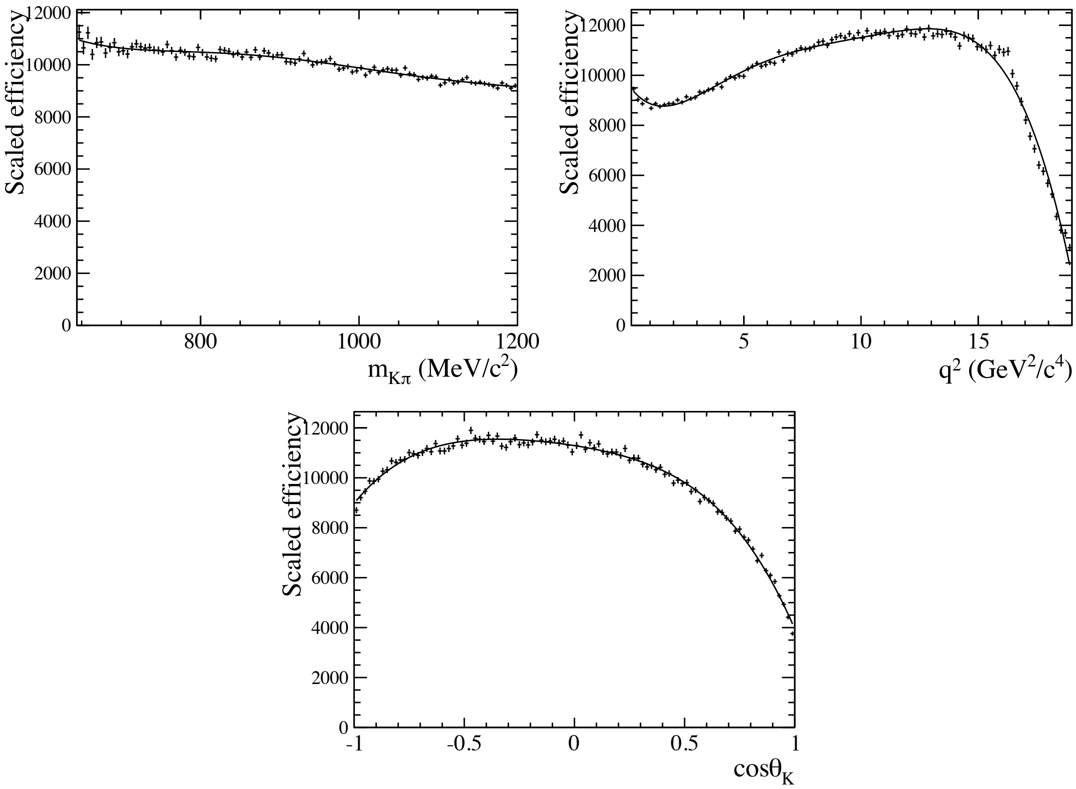
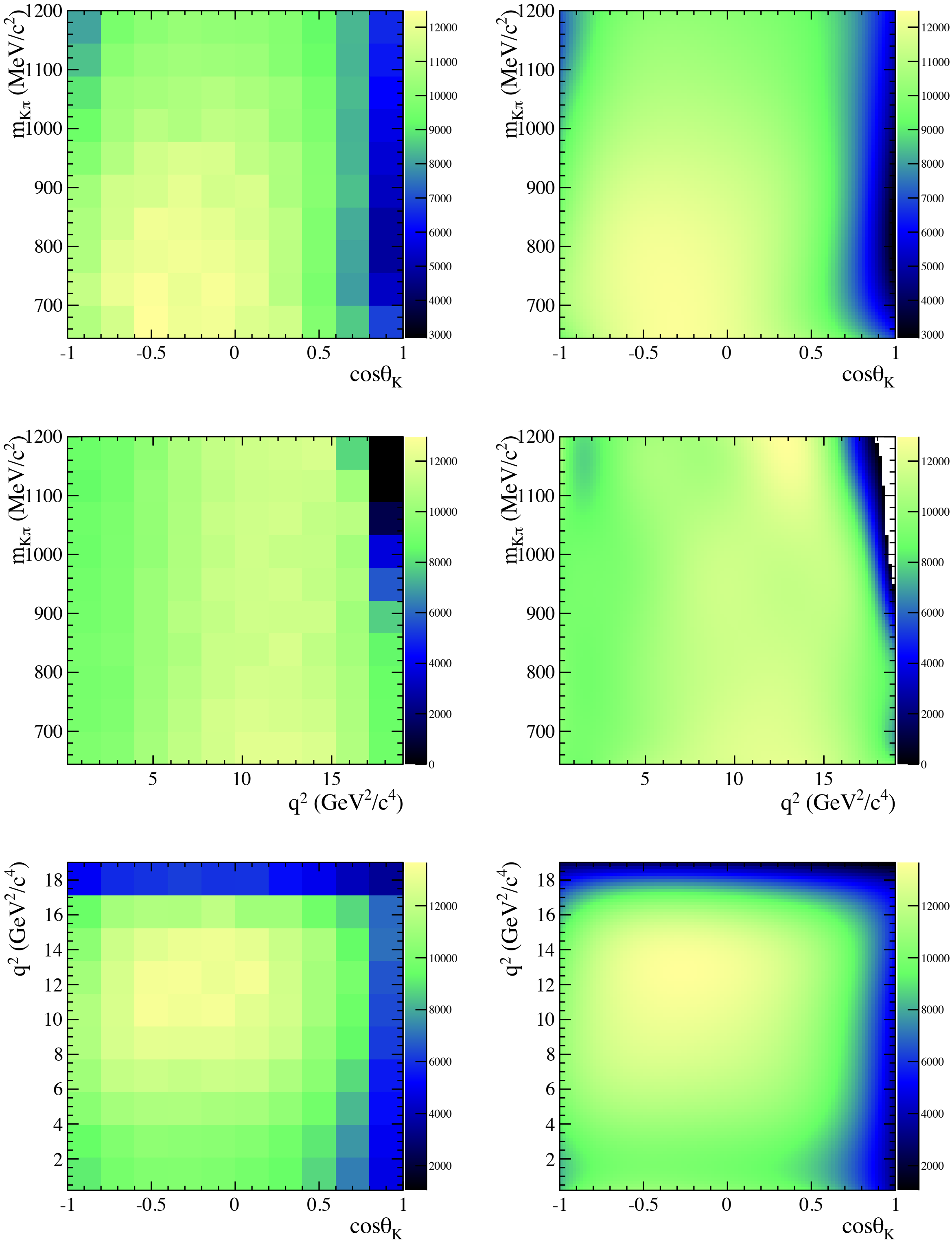


Figure 5.7: Two-dimensional projections of the sampled efficiency (left) and polynomial function returned by the Legendre method of moments (MoM) (right). The z-axis in both plots shows the projected efficiency scaled to the number of entries in the offline selected simulated sample, and is therefore arbitrary. Supplementary figures are included in Appendix C in Figs C.3–C.5.



Chapter 6

The model and fit procedure to extract the branching fraction and S-wave component

The total fitted model used to extract the S-wave component and the differential branching fraction is a combination of empirical models for the invariant mass spectra together with the theoretical angular distribution. The fit procedure consists of a multi-dimensional fit to $m_{K\pi\mu\mu}$, $m_{K\pi}$, and $\cos\theta_K$ in bins of q^2. For convenience, a fit to these dimensions will be referred to as the '3D fit'. For cross-checks it is also possible to perform a '2D fit' to $m_{K\pi\mu\mu}$ and $m_{K\pi}$ in order to determine $F_{\rm S}$, but with less precision than in the 3D case. In both cases, the $m_{K\pi\mu\mu}$ spectrum is included in order to constrain the signal and background yields.

The region of $m_{K\pi\mu\mu} \in [5230, 5330]\,\mathrm{MeV}/c^2$, equivalent to $\pm 50\,\mathrm{MeV}/c^2$ around the nominal mass of the B^0 meson will be referred to as the 'signal region' as it is dominated by genuine $B^0 \to K^+\pi^-\mu^+\mu^-$ events[1]. The region of $m_{K\pi\mu\mu}$ above the B^0 signal region is also included in the fit in order to constrain the shape of background events in the $m_{K\pi}$ and $\cos\theta_K$ distributions. This region, which consists of $m_{K\pi\mu\mu} \in [5400, 5780]\,\mathrm{MeV}/c^2$, will be referred to as the 'background sideband' as it is dominated by combinatorial background events. Remaining specific background events are treated as a source of systematic uncertainty as described in Section 7.3.6. Partially reconstructed decays are not included in the model. The full $m_{K\pi\mu\mu}$ fitting range,

$$(5180 < m_{K\pi\mu\mu} < 5780)\,\mathrm{MeV}/c^2, \tag{6.1}$$

excludes the significant contributions from partially reconstructed backgrounds. The selection and peaking background vetoes make the remaining such contributions negligible.

[1] The signal region has a signal-to-background ratio of roughly 0.6 across q^2, excluding the charmonia regions (Section 4.2.1).

In order to extract the differential branching fraction in q^2, the fit is performed in bins of q^2 which exclude the $B \to J/\psi K^+\pi^-$ and $B \to \psi(2S)K^+\pi^-$ resonance regions. The exact binning scheme to be used will be motivated in Section 6.5.1 and is presented in Table 6.3.

The full 3D model is

$$n_{\text{sig}} \cdot S(m_{K\pi\mu\mu}) \cdot S(m_{K\pi}, \cos\theta_K) + n_{\text{bkg}} \cdot B(m_{K\pi\mu\mu}) \cdot B(m_{K\pi}) \cdot B(\cos\theta_K). \tag{6.2}$$

The signal, S, and background, B, component probability density functions (pdfs)[2] are individually normalised and the fit is performed with an extended term in the likelihood such that the signal and background yields, n_{sig} and n_{bkg} float individually.

It is also possible to perform a fit to only the $m_{K\pi\mu\mu}$ spectrum with

$$n_{\text{sig}} \cdot S(m_{K\pi\mu\mu}) + n_{\text{bkg}} \cdot B(m_{K\pi\mu\mu}), \tag{6.3}$$

in order to extract the signal event yield. There is no way to distinguish between P- and S-wave components in this case. A fit to this spectrum in $B \to J/\psi K^+\pi^-$ is used to extract the yield of $B^0 \to J/\psi K^{*0}$ and the set of double-CB parameter values, used when a fit is made to the $B^0 \to K^+\pi^-\mu^+\mu^-$ data.

The simplified 2D model is

$$n_{\text{sig}} \cdot S(m_{K\pi\mu\mu}) \cdot \left[\int S(m_{K\pi}, \cos\theta_K) \mathrm{d}\cos\theta_K\right] + n_{\text{bkg}} \cdot B(m_{K\pi\mu\mu}) \cdot B(m_{K\pi}). \tag{6.4}$$

This model is used as a cross-check.

The exact form of the individual model components and associated assumptions will be described in the remainder of this chapter. In Section 6.1, the components $S(m_{K\pi\mu\mu})$ and $B(m_{K\pi\mu\mu})$ are described. The angular fit model including $m_{K\pi}$ is described in Section 6.2. Section 6.2.1 describes the angular distribution and Section 6.2.3 describes the modelling of the P- and S-wave lineshapes. The remaining background components, $B(m_{K\pi})$ and $B(\cos\theta_K)$ are described in Sections 6.2.4 and 6.2.4. Given the definition of these models, the procedures to extract the F_{S} parameter and the differential branching fraction are presented in Section 6.3. Finally, tests and validation of the model are presented in Sections 6.5 and 6.6.

6.1 Model for the $K\pi\mu\mu$ invariant mass spectrum

A model of the $m_{K\pi\mu\mu}$ spectrum was described in Section 4.3.3 in the context of preparing the $B \to J/\psi K^+\pi^-$ sample before training the BDT classifier. An identical model is used

[2] To be totally explicit: in Sections 4.3.1 and 4.3.5, roman S and B are used to represent the signal and background event yields with regards to training and optimising the multivariate classifier (as is conventional). In the present discussion, the italic S and B denote the pdfs. From here on, $n_{\text{sig}}, n_{\text{bkg}}$ are used for yields.

here. The signal component for $B^0 \to K^+\pi^-\mu^+\mu^-$ is described by a double-CB function,

$$S(m_{K\pi\mu\mu}) = F_{\mathrm{dCB}}(m_{K\pi\mu\mu}; m_B, \vec{\alpha}) = f_{\sigma_1} \cdot F_{\mathrm{CB}}(m_{K\pi\mu\mu}; m_B, \alpha, n, \sigma_1) + (1 - f_{\sigma_1}) \cdot F_{\mathrm{CB}}(m_{K\pi\mu\mu}; m_B, \alpha, n, \sigma_2), \quad (4.26)$$

and the combinatorial background component is described by an exponential shape,

$$B(m_{K\pi\mu\mu}) = \frac{e^{p_0 \cdot m_{K\pi\mu\mu}}}{\int e^{p_0 \cdot m_{K\pi\mu\mu}} \mathrm{d}m_{K\pi\mu\mu}}. \quad (6.5)$$

For $B^0 \to K^+\pi^-\mu^+\mu^-$ events where the $B \to J/\psi\, K^+\pi^-$ and $B \to \psi(2S)K^+\pi^-$ components are removed (discussed in Section 4.2.1) the functions above are the $m_{K\pi\mu\mu}$ components used in Eqns 6.2–6.3. The background model contributes one parameter, p_0, the coefficient of the exponential shape. The signal component contributes six parameters, $\vec{\alpha}$ and m_{B^0}, which are described in Section 4.3.3.

In order to extract the yield of $B^0 \to J/\psi\, K^{*0}$ events and a set of double-CB parameters, the control channel of $B^0 \to J/\psi\, K^+\pi^-$ decays in the range $q^2 \in [9.22, 9.96]\,\mathrm{GeV}^2/c^4$ is fitted. The yield of $B^0 \to J/\psi\, K^{*0}$ events, taken from a subsample of events in the range $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$, is used to normalise the differential branching fraction measurement. The double-CB parameters are taken from a fit to the full sample of $B^0 \to J/\psi\, K^+\pi^-$ events in the range $m_{K\pi} \in [644, 1200]\,\mathrm{MeV}/c^2$.

When fitting the control channel events in either $m_{K\pi}$ range, the signal model is extended to include a component for $B^0_s \to J/\psi\, K^+\pi^-$ decays.

$$S_{J/\psi\, K\pi}(m_{K\pi\mu\mu}) = (1 - f_{B^0_s}) \cdot F_{\mathrm{dCB}}(m_{K\pi}, m_{B^0}, \vec{\alpha}) + f_{B^0_s} \cdot F_{\mathrm{dCB}}(m_{K\pi}, m_{B^0_s}, \vec{\alpha}), \quad (6.6)$$

with the fraction of B^0_s events, $f_{B^0_s}$, which floats in the fit and account for the relative sizes of the two components. The $m_{K\pi\mu\mu}$ spectrum for control channel $B \to J/\psi\, K^+\pi^-$ data are fitted with this two-component signal model (with Eqn 6.5 and 6.6 inserted into Eqn. 6.3). The fitted $m_{K\pi\mu\mu}$ distributions are shown in Fig. 6.1.

The fit to the range $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$, which is $\pm 100\,\mathrm{MeV}/c^2$ around the nominal mass of the $K^*_1(892)^0$ gives the yield of $B^0 \to J/\psi\, K^{*0}$ events,

$$n_{B^0 \to J/\psi\, K^{*0}} = (1 - f_{B^0_s}) \cdot n_{\mathrm{sig}} = 324480 \pm 610. \quad (6.7)$$

This yield is needed to normalise the branching fraction calculation. The reason that the first fit is performed in the window $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$, is for compatibility with the measurement the branching fraction of the normalisation channel, $\mathcal{B}\,[B^0 \to J/\psi\, K^{*0}]$, which is taken from Ref. [87]. Extraction of the differential branching fraction, and the normalisation is discussed in Section 6.3. The values of the parameters returned by this fit are not used elsewhere in the analysis. They are given in Appendix E.1 in Table E.1 for completeness.

Table 6.1: The parameters from the fits to $B^0 \to J/\psi K^+\pi^-$ and $B^0_s \to J/\psi K^+\pi^-$ candidates in the ranges $q^2 \in [9.26, 9.96]$ GeV$^2/c^4$ and $m_{K\pi} \in [644, 1200]$ MeV/c^2. These parameters correspond to the lower plots in Fig. 6.1. The $\vec{\alpha}$, and m_{B^0} parameters are fixed in subsequent fits to the $B^0 \to K^+\pi^-\mu^+\mu^-$ sample.

parameter	value	units
n_{bkg}	41670 ± 380	–
n_{sig}	395590 ± 710	–
m_{B^0}	5284.40 ± 0.039	MeV/c^2
$m_{B^0_s} - m_{B^0}$	87 (fixed)	MeV/c^2
$f_{B^0_s}$	1.383 ± 0.032	10^{-2}
p_0	-4.872 ± 0.044	$10^{-3}[\text{MeV}/c^2]^{-1}$
α	1.360 ± 0.011	–
n	36 (fixed)	–
f_{σ_1}	0.643 ± 0.027	–
σ_1	14.97 ± 0.17	MeV/c^2
σ_2	24.25 ± 0.46	MeV/c^2

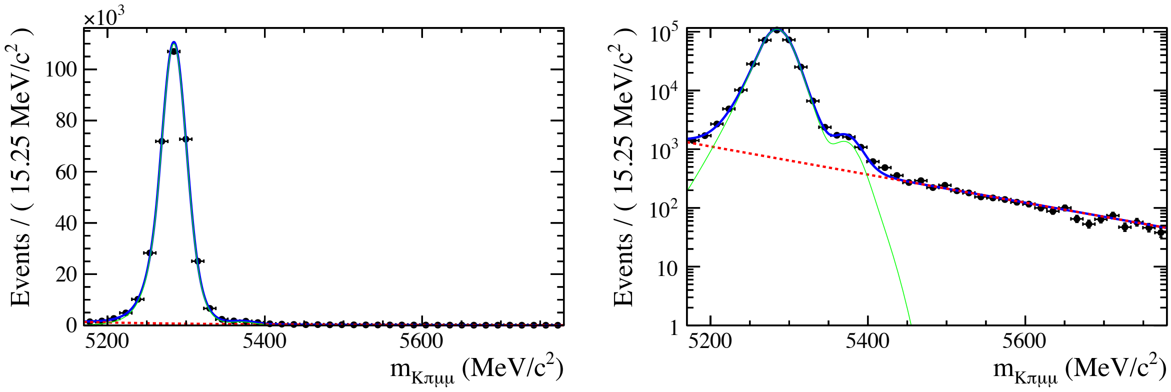
The fit to $m_{K\pi} \in [644, 1200]$ MeV/c^2, the $m_{K\pi}$ range in which the full 3D fit is performed, gives the parameters listed in Table 6.1. When performing the full 3D fit to $B^0 \to K^+\pi^-\mu^+\mu^-$ data away from the $B \to J/\psi K^+\pi^-$ and $B \to \psi(2S)K^+\pi^-$ regions, the six double-CB parameters are fixed to these values. This reduces the number of parameters in the fits to the $B^0 \to K^+\pi^-\mu^+\mu^-$ data samples which are statistically limited, having determined them rather precisely in a large sample of events. The resonant shape in $m_{K\pi\mu\mu}$ is universal between $B^0 \to J/\psi K^+\pi^-$ and $B^0 \to K^+\pi^-\mu^+\mu^-$ decays as it is dominated by the detector resolution. This procedure is the same as used in previous experimental analyses [26–28][3].

As a check, the signal model, $S(m_{K\pi\mu\mu})$, is fit to simulated $B^0 \to K^*_1(892)^0\mu^+\mu^-$ events. This fit is shown in Fig. 6.2.

[3]More details about which, may be found in Refs [1, 88].

Figure 6.1: The fitted $m_{K\pi\mu\mu}$ spectrum for $B^0 \to J/\psi K^+\pi^-$ and $B_s^0 \to J/\psi K^+\pi^-$ candidate decays shown on linear (left) and logarithmic (right) scales for two regions of $m_{K\pi}$. This fit is for $q^2 \in [9.92, 9.96]\,\mathrm{GeV}^2/c^4$ and the upper (lower) two plots are for $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$ ($m_{K\pi} \in [644, 1200]\,\mathrm{MeV}/c^2$). The yield of $B^0 \to J/\psi K^{*0}$ events, used to normalise the differential branching fraction of $B^0 \to K^+\pi^-\mu^+\mu^-$ is taken from the fit to the narrow window $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$. The set of shape parameters for the double-CB, $\vec{\alpha}$ and m_{B^0}, found for the wider window of $m_{K\pi} \in [644, 1200]\,\mathrm{MeV}/c^2$, are fixed in subsequent fits to $B^0 \to K^+\pi^-\mu^+\mu^-$ (away from the J/ψ and $\psi(2S)$ resonances). These parameters are given in Table 6.1.

$m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$



$m_{K\pi} \in [644, 1200]\,\mathrm{MeV}/c^2$

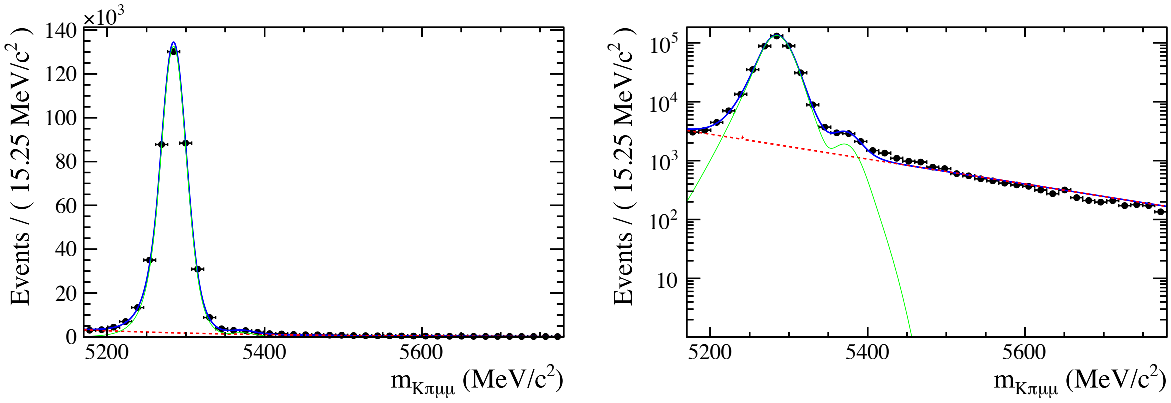
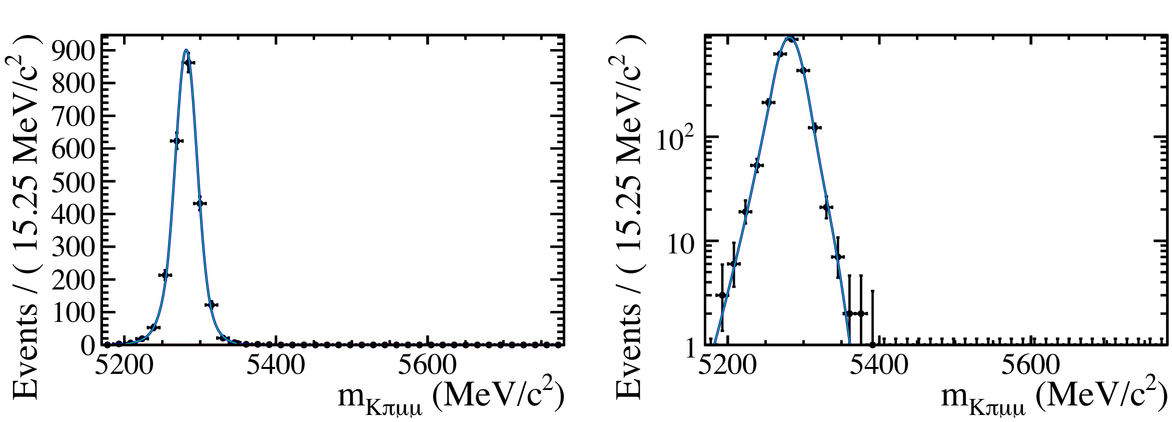


Figure 6.2: A fit to the $m_{K\pi\mu\mu}$ spectrum for reco. level simulated $B^0 \to K_1^*(892)^0\mu^+\mu^-$ events shown on linear (left) and logarithmic (right) scales.



6.2 Model for the angular fit including $K\pi$ invariant mass

6.2.1 The theoretical angular distribution

The full angular distribution,

$$\begin{aligned}\frac{\mathrm{d}^5\Gamma}{\mathrm{d}q^2\mathrm{d}m_{K\pi}\mathrm{d}\cos\theta_K\mathrm{d}\cos\theta_\ell\mathrm{d}\phi} = &\frac{1}{4\pi}\left|g_{K_0^*}\right|^2\cdot\gamma_{\mathrm{S}}(1-\cos 2\theta_\ell)\\ &+\frac{3}{4\pi}\left|g_{K_1^*}\right|^2\cdot\gamma_{\mathrm{P0}}\cos^2\theta_K(1-\cos 2\theta_\ell)\\ &+\frac{\sqrt{3}}{2\pi}\mathfrak{Re}\left(\left(g_{K_0^*}g_{K_1^*}\right)\cdot(x_1+ix_2)\right)\cos\theta_K(1-\cos 2\theta_\ell)\\ &+\frac{9}{16\pi}\left|g_{K_1^*}\right|^2\cdot\gamma_{\mathrm{PT}}\sin^2\theta_K\left(1+\frac{1}{3}\cos 2\theta_\ell\right)\\ &+\frac{3}{8\pi}\left|g_{K_1^*}\right|^2\tilde{J}_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi\\ &+\frac{3}{2\pi}\left|g_{K_1^*}\right|^2\tilde{J}_6\sin^2\theta_K\cos\theta_\ell\\ &+\frac{3}{4\pi}\left|g_{K_1^*}\right|^2\tilde{J}_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi.\end{aligned} \tag{2.157}$$

as well as a simplified angular distribution obtained by integrating over $\cos\theta_\ell$ and ϕ,

$$\begin{aligned}\frac{\mathrm{d}^3\Gamma}{\mathrm{d}q^2\mathrm{d}m_{K\pi}\mathrm{d}\cos\theta_K} = &\frac{2}{3}\left|g_{K_0^*}\right|^2\gamma_{\mathrm{S}}+2\left|g_{K_1^*}\right|^2\gamma_{\mathrm{P0}}\cos^2\theta_K+\left|g_{K_1^*}\right|^2\gamma_{\mathrm{PT}}\sin^2\theta_K\\ &+\frac{4\sqrt{3}}{3}\mathfrak{Re}\left(\left(g_{K_0^*}g_{K_1^*}^*\right)\cdot(x_1+ix_2)\right)\cos\theta_K\end{aligned} \tag{2.150}$$

were derived in Section 2.6.2. They are repeated above for the reader's convenience. The modelling of the $m_{K\pi}$ dependence, encoded in the functions $g_{K_J^*}$, is discussed in Sections 6.2.2 and 6.2.3.

In the 3D fit case, the signal is described by the theoretical angular distribution with the angular efficiency function folded in,

$$\begin{aligned}&S(m_{K\pi},\cos\theta_K)\propto\\ &\iint\mathrm{d}\cos\theta_\ell\mathrm{d}\phi\left[\frac{\mathrm{d}^5\Gamma}{\mathrm{d}q^2\mathrm{d}m_{K\pi}\mathrm{d}\cos\theta_K\mathrm{d}\cos\theta_\ell\mathrm{d}\phi}\times\int\mathrm{d}q^2\epsilon(q^2,\cos\theta_K,\cos\theta_\ell,\phi,m_{K\pi})\right].\end{aligned} \tag{6.8}$$

This expression represents the complete signal model which used for the final fit to $B^0\to K^+\pi^-\mu^+\mu^-$ data from which the results are extracted. Several approximations of Eqn 6.8 are used for cross-checks.

Assumptions used in cross-checks

In the ideal case of constant efficiency in $\cos\theta_\ell$ and ϕ, which is approximately true,

$$\epsilon\approx\epsilon(q^2,\cos\theta_K,m_{K\pi}), \tag{6.9}$$

or if the parameters $\{S_3, A_{\mathrm{FB}}, S_9\}$ are all zero, then Eqn. 6.8 reduces to

$$\frac{\mathrm{d}^3\Gamma}{\mathrm{d}q^2\mathrm{d}m_{K\pi}\mathrm{d}\cos\theta_K} \times \int \mathrm{d}q^2\epsilon(q^2, \cos\theta_K, m_{K\pi}). \tag{6.10}$$

This simplification is useful as a cross-check.

Integrating Eqn. 2.150 over $\cos\theta_K$, gives

$$S(m_{K\pi}) \propto \frac{\mathrm{d}^2\Gamma}{\mathrm{d}q^2\mathrm{d}m_{K\pi}} = \frac{4}{3}\left|g_{K_0^*}\right|^2\gamma_{\mathrm{S}} + \frac{4}{3}\left|g_{K_1^*}\right|^2(\gamma_{\mathrm{P0}} + \gamma_{\mathrm{PT}}). \tag{6.11}$$

In doing this, one assumes that the efficiency function is constant in $\cos\theta_K$ which is a much less valid approximation than constant efficiency in $\cos\theta_\ell$ and ϕ. However it may be fit to data ignoring the effect of the efficiency as rather coarse cross check.

6.2.2 Modelling the $K\pi$ invariant mass lineshape

In Eqns 2.142 and 2.143 from Section 2.6.2, the complex functions $g_{K_J^*} = g_{K_J^*}(m_{K\pi})$, $J = 0, 1$ were introduced. These functions model the $m_{K\pi}$ spectrum which contains several resonances.

The $m_{K\pi}$ dependence of a resonance, R, may be modelled as a relativistic Breit-Wigner function. Each resonance also has a spin-dependent Blatt-Weisskopf barrier factor, B'_L, and a spin-dependent phase-space factor for both the $B \to K_J^*[\mu\mu]$ and for the $K_J^* \to K\pi$ parts of the decay. The two parts of the decay are often referred to as the 'birth' and 'death' of the K_J^*.

The first three Blatt-Weisskopf barrier factors [89] are:

$$B'_0(p, p_R, d) = 1; \tag{6.12}$$

$$B'_1(p, p_R, d) = \sqrt{\frac{1 + (p_R d)^2}{1 + (pd)^2}}; \tag{6.13}$$

$$B'_2(p, p_R, d) = \sqrt{\frac{9 + 3(p_R d)^2 + (p_R d)^4}{9 + 3(pd)^2 + (pd)^4}}; \tag{6.14}$$

where p is the linear momentum available to the daughters, p_R is the linear momentum calculated at the resonance pole mass (*i.e.* $m_{ab} = m_R$) and d is the impact parameter, or 'meson radius' which is a parameter taken from scattering experiments. The Blatt-Weisskopf factors account for the fact that low momentum particles have difficulty generating sufficient angular momentum to conserve the spin of the resonance [4]. The meson radius parameter is fixed to $d = 1.6\,[\,\mathrm{GeV}/c]^{-1}$ which was the value used in the analysis presented in Ref. [90]. This choice is considered as a source of systematic uncertainty in Section 7.3.2.

The lineshape of a single resonance R, with orbital angular momentum J, may be written as:

$$f_R(m_{K\pi}, q^2, J) = \sqrt{kp} \cdot B'_{L_B}(p, p_R, d) \cdot \left(\frac{p}{p_R}\right)^{L_B} \cdot B'_{L_{K\pi}}(k, k_R, d) \cdot \left(\frac{k}{k_R}\right)^{L_{K\pi}} \\ \times \frac{1}{m_{K\pi}^2 - m_R^2 - im_R\Gamma_J(m_{K\pi})}, \tag{6.15}$$

where m_R is the pole-mass of the resonance; p (k) is the phase-space momentum for the K^* (kaon) in the rest frame of the B (K^*) evaluated at a given $m_{K\pi}$; p_R and k_R are the same, except that they are evaluated at the pole-mass. The function is formally dependent on q^2 since it is used to compute p and p_R. The factor $B'_{L_B}(p, p_R, d) \cdot (p/p_R)^{L_B}$ is therefore the 'birth factor' of the K^*, and $B'_{L_{K\pi}}(k, k_R, d) \cdot (k/k_R)^{L_{K\pi}}$ is the 'death factor'. The L_B and $L_{K\pi}$ are the relative angular momentum between the daughters. In literature this is sometimes referred to as the 'break-up angular momentum'. For all $K\pi$ states $K_J^* \to K\pi$ the daughters are spinless, so $L_{K\pi} = J$ of the resonance/state. For $B^0 \to [K^*{}_0]\,[\mu^+\mu^-]$ the dimuon object may be in a vector state and therefore $L_B = 1$ in this case. For $B^0 \to [K_1^*(892)^0]\,[\mu^+\mu^-]$ there are three possible values, $L_B \in \{0, 1, 2\}$. However, the difference is completely negligible in the kinematic ranges under consideration in this analysis. The $L_B = 0$ state is therefore used for the birth factor of the vector $K_1^*(892)^0$. The so-called running width, $\Gamma_J(m_{K\pi})$, is a function of the resonance width, Γ_R and is given by

$$\Gamma_J(m_{K\pi}) = \Gamma_R \cdot B'^2_{L_{K\pi}}(k, k_R, d) \cdot \left(\frac{k}{k_R}\right)^{2L_{K\pi}+1} \left(\frac{m_R}{m_{K\pi}}\right). \tag{6.16}$$

The description of a resonant lineshape given in Eqn. 6.15 follows the form used in EvtGen and by the PDG [4]. It differs slightly from Ref. [90], where the factor $(k/k_R)^{L_{K\pi}}$ was replaced by $(k/m_{K\pi})^{L_{K\pi}}$ for consistency with Belle [91]. This change is investigated as a source of systematic uncertainty, described in Section 7.3.1.

The function, f_R, is sufficient to describe the $m_{K\pi}$ shape of the P-wave component,

$$g_{K_1^*}(m_{K\pi}) = f_{892}(m_{K\pi}), \tag{6.17}$$

where f_{892} is used as a shorthand for $f_{K_1^*(892)^0}$.

6.2.3 Models for the S-wave component

Several possible models for an S-wave component exist. For the purposes of this analysis, which suffers from low statistics, two models that are found to perform well empirically are employed. The model used by default is referred to as the 'LASS model' [92]. It is combination of a relativistic Breit-Wigner shape to describe the $K_0^*(1430)^0$ with an empirical non-resonant (NR) component. Another model, known as the 'Isobar model', is used to assess the systematic variation due to this choice. The Isobar model is a coherent sum of components for the dominant K^* resonances which includes a component for the $K_0^*(800)^0$. As will be shown in Section 7.3.1, the choice of S-wave model is the dominant source of systematic uncertainty in the measurement of F_S.

The LASS model

The so-called LASS model was developed to empirically describe the $K\pi$ spectrum from the scattering experiment of the same name [92]. It is often used to describe the scalar component in the vicinity of the $K_1^*(892)^0$, although usually in a small window around the nominal mass. The functional form is,

$$g_{K_0^*}^{\mathrm{LASS}}(m_{K\pi}) \equiv B'_{L_B}(p, p_R, d) \cdot \left(\frac{p}{p_R}\right)^{L_B} \sqrt{p} \cdot \left(\frac{1}{\cot\delta_B - i} + e^{2i\delta_B}\frac{1}{\cot\delta_R - i}\right). \tag{6.18}$$

With the terms

$$\cot\delta_B = \frac{1}{ak} + \frac{rk}{2}; \tag{6.19}$$

and

$$\cot\delta_R = \frac{m_{1430}^2 - m_{K\pi}^2}{m_{1430}\Gamma_S(m_{K\pi})}. \tag{6.20}$$

Reference [93] quotes the running width corresponding to Eqn. 6.16 with $L = 0$ for the $K_0^*(1430)^0$:

$$\Gamma_S(m_{K\pi}) = \Gamma_{1430,L_{K\pi}=0}(m_{K\pi}) = \Gamma_{1430}\frac{k}{k_R}\frac{m_{1430}}{m_{K\pi}}. \tag{6.21}$$

The second term of Eqn. 6.18 is equivalent to a Breit-Wigner for the $K_0^*(1430)^0$. The first term of Eqn. 6.18 contains two empirical parameters $\{a, r\}$. These parameters are fixed to the values reported by the analysis of Ref. [90],

$$a = 2.86\,[\mathrm{GeV}/c]^{-1} \tag{6.22}$$

$$r = 3.83\,[\mathrm{GeV}/c]^{-1}. \tag{6.23}$$

In order to assess the systematic effect of this choice they are also fixed to values from the LASS experiment,

$$a = 1.94\,[\mathrm{GeV}/c]^{-1} \tag{6.24}$$

$$r = 1.76\,[\mathrm{GeV}/c]^{-1}, \tag{6.25}$$

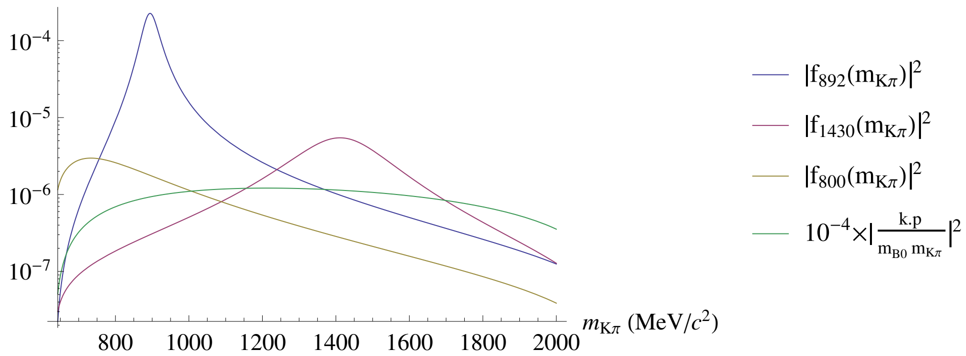
which is detailed in Section 7.3.1. The values in Eqns 6.24 and 6.25 are reported by BaBar [93] as coming from private communication with the LASS authors [92].

The Isobar model

The Isobar model consists of a coherent sum of components for the dominant S-wave resonances. Within the region $m_{K\pi} \in [644, 1200]\,\mathrm{MeV}/c^2$, the dominant resonances are the $K_0^*(800)^0$ and the $K_0^*(1430)^0$. A component, using Eqn 6.15, is included for each of these resonances. In addition, a NR component is included with a shape determined by phase-space alone. The $m_{K\pi}$ dependent S-wave shape is then given by

$$g_{K_0^*}^{\mathrm{Isobar}}(m_{K\pi}, q^2) = f_{800}(m_{K\pi}, q^2)\alpha_{800}e^{-i\delta\beta_{800}} + f_{1430}(m_{K\pi}, q^2) + \frac{pk}{m_{B^0}m_{K\pi}}\alpha_{\mathrm{NR}}e^{-i\delta\beta_{\mathrm{NR}}}, \tag{6.26}$$

Figure 6.3: The raw resonance functions $|f_R|^2$ with $J = 0, 1$ spin assignments and the non-resonant shape scaled by a factor 10^{-4}.



where, as in Eqn 6.17, the notation f_R with $R = 1430$ ($R = 800$) is used as shorthand for $K_0^*(1430)^0$ ($K_0^*(800)^0$). In Eqn. 6.26 above, α_{800} ($\delta\beta_{800}$) is the relative magnitude (phase) of the $K_0^*(800)^0$ relative to the $K_0^*(1430)^0$. The variables $\alpha_{\rm NR}$ and $\delta\beta_{\rm NR}$ are defined analogously for the NR component.

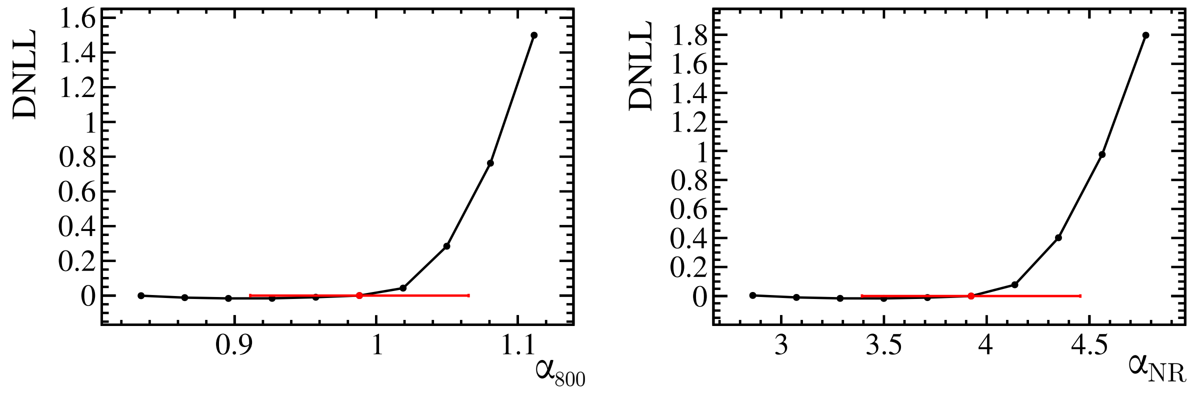
As might be expected, given the very similar $m_{K\pi}$ dependence (see Fig. 6.3) the model is found to be degenerate when the NR component is used in addition the $K_0^*(800)^0$. This is the case even in fits to subsamples of the control channel $B^0 \to J/\psi\, K^+\pi^-$ dataset with statistics far in excess of those expected for $B^0 \to K^+\pi^-\mu^+\mu^-$. Figure 6.4 shows likelihood profiles for the relative magnitude of the $K_0^*(800)^0$ and NR fit to a subsample of the control channel data. The degeneracy is removed by fixing either component. As the LASS model contains a NR component, the NR component in the Isobar model is removed leaving only the $K_0^*(1430)^0$ and $K_0^*(800)^0$ components. Whilst for a full analysis of $B^0 \to J/\psi\, K^+\pi^-$, this approach is not likely to be sufficient, the model is able to describe the $B^0 \to K^+\pi^-\mu^+\mu^-$ dataset very well.

Parameters for the $K_0^*(800)^0$

The function f_{800} describes the component for the $K_0^*(800)^0$ resonance. This function takes the mass, m_{800}, and the width, Γ_{800}, of the $K_0^*(800)^0$. Unfortunately, measurements of the properties of the $K_0^*(800)^0$ from scattering experiments and from production experiments are in poor agreement [4]. In addition, the value taken from the scattering experiments is itself a phenomenological average. In order to obtain realistic values, the mass and width to are fixed to a recomputed average of the production experiment values given in the PDG. This average is shown in Fig 6.5 with the counterpart for m_{800} from the PDG. The average values are found to be

$$m_{800} = (819.2 \pm 42.4)\,\mathrm{MeV}/c^2 \tag{6.27}$$

Figure 6.4: Likelihood profiles of the α_{800} (left) and α_{NR} (right) parameters performed on 25×10^3 $B^0\to J/\psi\,K^+\pi^-$ data events. The values of the parameters shown on the x-axis are arbitrary. The red point and error band corresponds to the minimum and uncertainty found by the fit minimisation algorithm. The parameters are observed to be degenerate with one another. Fixing α_{NR} to zero lifts this degeneracy.



and

$$\Gamma_{800} = (472.8 \pm 90.4)\,\text{MeV}/c^2. \tag{6.28}$$

In the Isobar model, these parameters are fixed to the central values above. They do not appear in the (default) LASS model.

6.2.4 Background models for $K\pi$ invariant mass and $\cos\theta_K$

$K\pi$ invariant mass parameterisation

The background shape in the $K\pi$ invariant mass is modelled as a smooth shape with a kinematic turn on. The simple function:

$$B(m_{K\pi}; m_{th}, n) = (m_{K\pi} - m_{th})^{\frac{1}{n}} \cdot \lambda \tag{6.29}$$

is found to perform well empirically. It contributes two parameters, the n^{th} root-parameter and the threshold mass, m_{th}, which is fixed to be

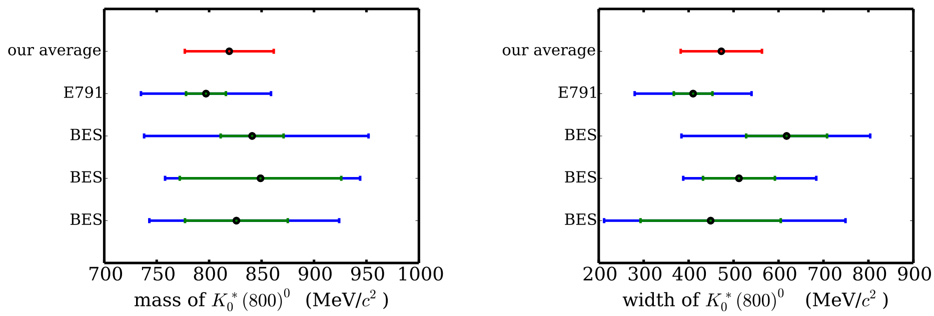
$$m_{th} = m_\pi + m_K = 634\,\text{MeV}/c^2. \tag{6.30}$$

At large values of q^2, the available phase-space becomes too small to produce events across the full range of $m_{K\pi}$. The factor λ, is the phase-space momentum for an object of $m_{K\pi\mu\mu} = 5780\,\text{MeV}/c^2$ (the end of the fitting range) to produce a $K\pi$ object at the q^2 of the centre of the bin. The phase-space factor is described in Appendix B.2.

As a cross check, the empirical 'D^*-D^0' shape function is used. The function was originally used by BaBar [94] to model background from slow pions in D^0 mixing measurements. This function is

$$B_{\text{alt.}}(m_{K\pi}; c, a, b) = \left(1 - e^{-\frac{(m_{K\pi}-m_{th})}{c}}\right) \cdot \left(\frac{m_{K\pi}}{m_{th}}\right)^a + b \cdot \left(\frac{m_{K\pi}}{m_{th}} - 1\right), \tag{6.31}$$

Figure 6.5: The measured values of mass (left) and width (right) of the $K_0^*(800)^0$ from production experiments taken from the PDG [4]. The statistical error is represented in blue and the systematic in green. The average values (used in this analysis) are represented by the red points. The PDG figure taken from [4] is shown below.



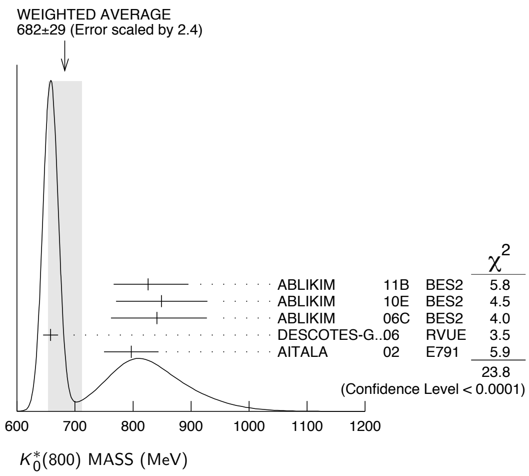
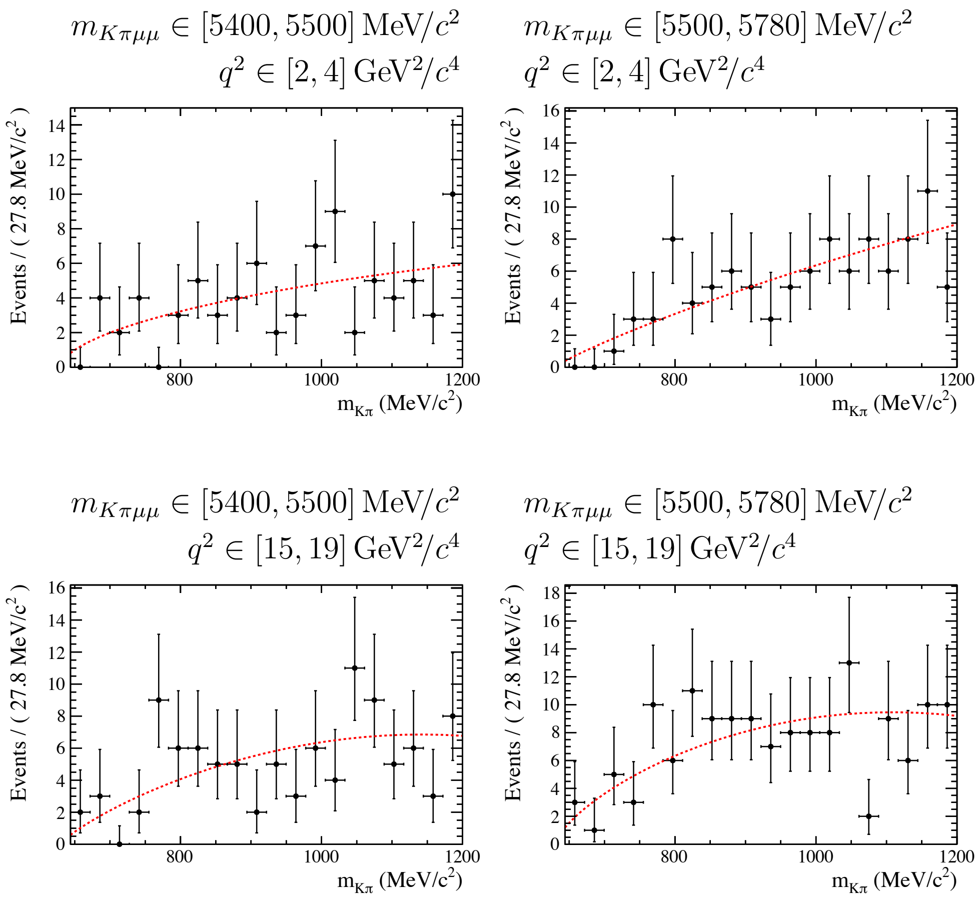


Figure 6.6: Background-only fits to $m_{K\pi}$ in the regions of $m_{K\pi\mu\mu}$ and q^2 specified above each plot.



where c is a parameter controlling the curvature at low $m_{K\pi}$, a is the power with which the function grows at high $m_{K\pi}$, and b is a further parameter determining the fall off at high $m_{K\pi}$. For $B_{\text{alt.}}$, the parameter m_{th} is also fixed. In fits to simulated data, the $\{a, b, c\}$ parameters were found to be excessively correlated, hence the default use of of $B(m_{K\pi}; m_{th}, n)$.

To further justify the choice of a single parameter, n, to describe the $m_{K\pi}$ background shape, fits to $m_{K\pi}$ are shown in Fig. 6.6 for the $m_{K\pi\mu\mu}$ background sideband. In this figure, the background-dominated data are well described by the simple function.

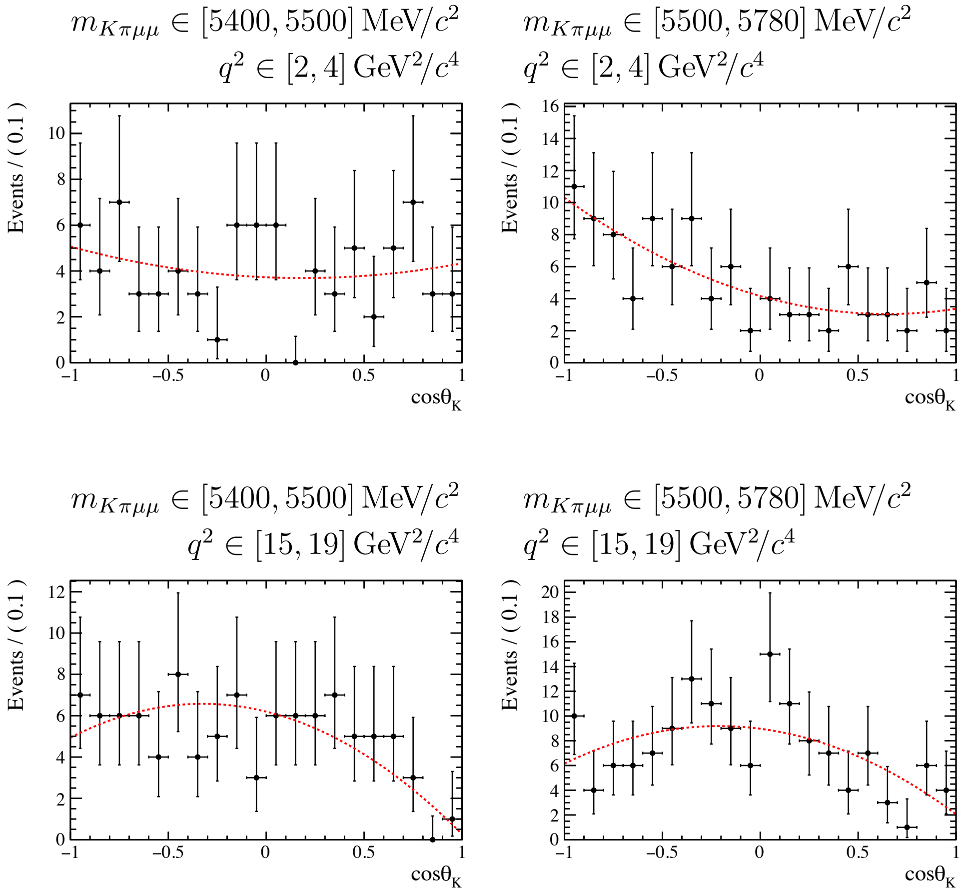
$\cos\theta_K$ parameterisation

The background component in $\cos\theta_K$ is described by a sum of Bernstein polynomials, $h_i(x)$, up to quadratic order,

$$B(\cos\theta_K; k_1, k_2) = 1 + k_1 h_1(\cos\theta_K) + k_2 h_2(\cos\theta_K). \tag{6.32}$$

The k_i are free parameters. This choice is studied, similarly to the $m_{K\pi}$ choice, with the 'background-only' fits which are shown in Fig. 6.7. Models containing cubic and higher

Figure 6.7: Background-only fits to $\cos\theta_K$ in the regions of $m_{K\pi\mu\mu}$ and q^2 specified above each plot.



order Bernstein polynomials were found to be degenerate when exploring the minimum found in the fits to $B^0 \to J/\psi\, K^+\pi^-$ data. A sum of Chebyshev polynomials was tried, however such a model is not forced to be positive, and fits then have convergence problems owing to negative pdf values. In order to estimate the systematic uncertainty due to this choice of model, a model up to cubic order is used. This is found to be a negligible contribution in the total systematic uncertainty, and is discussed in Section 7.3.7.

6.3 Extracting the size of the S-wave component and the differential branching fraction

Having included the $m_{K\pi}$-dependence explicitly in the angular distribution (Eqn 2.157), the fractional size of the S-wave component itself acquires $m_{K\pi}$-dependence. The quantity $\mathcal{F}_{\rm S}$, is defined as the fraction of the S-wave amplitudes to the total. Using the notation introduced in Section 2.6.2, this is

$$
\begin{aligned}
\mathcal{F}_{\mathrm{S}} &= \mathcal{F}_{\mathrm{S}}(m_{K\pi}) && (6.33)\\
&\equiv \frac{\left(|A_{0L0}|^2 + |A_{0R0}|^2\right)\cdot\left|g_{K_0^*}(m_{K\pi})\right|^2}{\left(|A_{0L0}|^2 + |A_{0R0}|^2\right)\cdot\left|g_{K_0^*}(m_{K\pi})\right|^2 + \left(\sum_H\sum_p |A_{1Hp}|^2\right)\cdot|f_{896}(m_{K\pi})|^2} && (6.34)\\
&= \frac{\gamma_{\mathrm{S}}\cdot\left|g_{K_0^*}(m_{K\pi})\right|^2}{\gamma_{\mathrm{S}}\cdot\left|g_{K_0^*}(m_{K\pi})\right|^2 + \left(\gamma_{\mathrm{P0}}+\gamma_{\mathrm{PT}}\right)\cdot|f_{896}(m_{K\pi})|^2} && (6.35)
\end{aligned}
$$

In order to obtain the fractional size of the S-wave contribution in a format that is useful to inform the measurement of $\mathcal{B}\,[B^0 \to K_1^*(892)^0\mu^+\mu^-]$, in which the $m_{K\pi}$ dependence is not included, one must integrate over some region of $m_{K\pi}$ in both the numerator and denominator individually. It is therefore convenient to introduce the notation:

$$
F_{\mathrm{S}}\Big|_a^b \equiv \frac{\gamma_{\mathrm{S}}\cdot\int_a^b \mathrm{d}m_{K\pi}\left|g_{K_0^*}(m_{K\pi})\right|^2}{\gamma_{\mathrm{S}}\cdot\int_a^b \mathrm{d}m_{K\pi}\left|g_{K_0^*}(m_{K\pi})\right|^2 + (\gamma_{\mathrm{P0}}+\gamma_{\mathrm{PT}})\cdot\int_a^b \mathrm{d}m_{K\pi}\left|f_{892}(m_{K\pi})\right|^2}, \qquad (6.36)
$$

for brevity, the units of a and b will always be MeV/c^2 and will be dropped for the remainder of this thesis. The parameter of interest for this analysis is $F_{\mathrm{S}}|_{644}^{1200}$. Angular analyses are typically performed in the window of $\pm 100\,\mathrm{MeV}/c^2$ around the nominal $K_1^*(892)^0$ mass, in this notation the F_{S} parameter of concern to these analyses is $F_{\mathrm{S}}|_{796}^{996}$.

The S-wave-corrected differential branching fraction is given by,

$$
\begin{aligned}
\frac{\mathrm{d}\mathcal{B}\,[B^0 \to K_1^*(892)^0\mu\mu]}{\mathrm{d}q^2} = &\frac{(1-\langle F_{\mathrm{S}}|_{644}^{1200}\rangle)}{\Delta q^2}\cdot\frac{\langle n_{K\pi\mu\mu}\rangle}{n_{J/\psi K^*}}\cdot\frac{\epsilon_{J/\psi K^*}}{\langle\epsilon_{K\pi\mu\mu}\rangle}\\
&\times \mathcal{B}\big[B^0 \to J/\psi\, K^{*0}\big]\cdot\mathcal{B}[J/\psi \to \mu\mu].
\end{aligned} \qquad (6.37)
$$

In this equation, n_f (ϵ_f) is the number of signal events returned by the fit (efficiency) for the decay $B^0 \to f$. The parentheses $\langle\rangle$ are used to denote quantities measured in a q^2 bin, with bin width

$$
\Delta q^2 = {q^2}_{\mathrm{max.}} - {q^2}_{\mathrm{min.}}. \qquad (6.38)
$$

It is important, for the above formula to hold, that $\mathcal{B}\,[B^0 \to J/\psi\, K^{*0}]$ does not distinguish S-wave from P-wave, *i.e.* that F_{S} for $B^0 \to J/\psi\, K^{*0}$ decays need not be taken into account. In this case, a simple fit to the $m_{K\pi\mu\mu}$ spectrum of $B^0 \to J/\psi\, K^{*0}$ candidate events can be used rather than the 3D fit needed to separate the P- and S-wave components. The normalisation branching fraction is discussed as an external input in the following section.

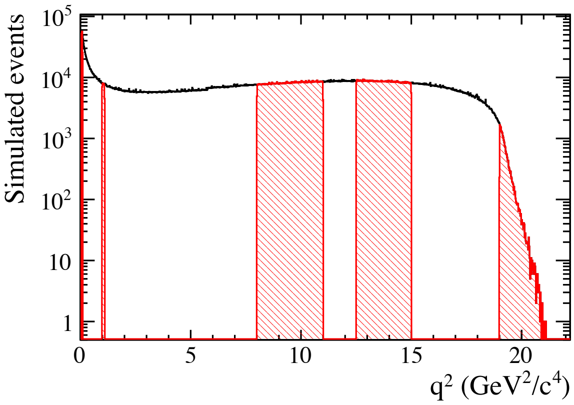


Figure 6.8: The generator-level (gen.) simulated q^2 distribution shown on a logarithmic scale. The red shaded areas are removed by the charmonia vetoes (Section 4.2.1) or excluded by the binning scheme. In order to compute the total branching fraction, a factor to account for this, f_{q^2}, is computed.

The total branching fraction is then given by

$$\mathcal{B}\left[B^0 \to K_1^*(892)^0 \mu\mu\right] = f_{q^2} \cdot \sum_{q^2 \text{ bins}} \Delta q^2 \cdot \left\langle \frac{\mathrm{d}\mathcal{B}\left[B^0 \to K_1^*(892)^0 \mu\mu\right]}{\mathrm{d}q^2} \right\rangle \tag{6.39}$$

$$\begin{aligned} &= f_{q^2} \cdot \frac{\epsilon_{J/\psi\, K^*}}{n_{J/\psi\, K^*}} \cdot \mathcal{B}\left[B^0 \to J/\psi\, K^{*0}\right] \cdot \mathcal{B}[J/\psi \to \mu\mu] \\ &\times \sum_{q^2 \text{ bins}} (1 - \langle F_{\mathrm{S}}|_{644}^{1200} \rangle) \frac{\langle n_{K\pi\mu\mu} \rangle}{\langle \epsilon_{K\pi\mu\mu} \rangle}. \end{aligned} \tag{6.40}$$

The correction factor, f_{q^2}, accounts for the events that are removed due to the q^2 vetoes, described in Section 4.2.1, and that the final q^2 bin terminates at $19\,\mathrm{GeV}^2/c^4$ (as will be discussed in Section 6.5.1). This factor is taken from the gen. level simulation sample and is found to be $f_{q^2} = 1.5322 \pm 0.0002$ which is slightly larger than the value quoted in Ref. [88][4] due to the exclusion of $B^0 \to \phi(1020) K^{*0}$ events that is performed in the present analysis. Figure 6.8 shows the q^2 distribution of the gen. sample with the vetoed regions shaded in red. The sum of the events that remain after the vetoes accounts is $1/f_{q^2} \approx 65\%$ of the total.

External inputs

The differential (and total) branching fraction requires two external inputs: the branching fraction of the normalisation channel, $\mathcal{B}\,[B^0 \to J/\psi\, K^{*0}]$, and the branching fraction of the decay $J/\psi \to \mu^+\mu^-$.

The PDG average for $\mathcal{B}\,[B^0 \to J/\psi\, K^{*0}]$ takes values from analyses which make the assumption that the $\Upsilon(4S)$ decays equally to B^+ and B^0. For the analysis of Ref [28, 88] the most precise values from Ref. [87] were taken, translated into partial widths and, assuming isospin of the $B \to J/\psi\, K^*$ decays, translated back into branching fractions. This procedure avoids the assumption of equal production at the expense of a slightly larger uncertainty due to the normalisation channel.

[4]Ref. [88] used $f_{q^2} = 1.50$, with no uncertainty, given in the same notation as above.

The normalisation branching fraction is therefore taken from Ref. [28]. Which is calculated from the most precise single measurement [87] and does not distinguish between P- and S-wave components,

$$\mathcal{B}\left[B^0 \to J/\psi\, K^{*0}\right] = (1.331 \pm 0.025 \pm 0.084) \times 10^{-3}. \tag{6.41}$$

This value is consistent with the PDG average, but less precise. The remaining branching fraction is taken from the PDG average, it is

$$\mathcal{B}\left[J/\psi \to \mu^+\mu^-\right] = (5.961 \pm 0.033)\%. \tag{6.42}$$

The BaBar measurement of the control channel branching fraction made no distinction between P- and S-wave components and selected the K^{*0} using the window $m_{K\pi} \in [796, 996]\,\text{MeV}/c^2$ [87]. Therefore the normalisation channel yield, $n_{J/\psi\, K^*}$, must be taken from the fit to the $m_{K\pi\mu\mu}$ spectrum for this narrower $m_{K\pi}$ window. By contrast, the 3D fit to the window $m_{K\pi} \in [644, 1200]\,\text{MeV}/c^2$ yields $n_{\text{sig}} = n_{K\pi\mu\mu}$. The ratio of efficiencies accounts for this difference in $m_{K\pi}$ window used to select the $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to K^+\pi^-\mu^+\mu^-$ candidates.

6.4 Parameter count

The full combined model with the LASS (Isobar) description of the S-wave has 27 (29) parameters. Not all of these parameters are allowed to float freely in the fit.

The background model contains five parameters,

$$\{p_0,\, n,\, k_1,\, k_2\} \cup \{m_{th}\}\,. \tag{6.43}$$

The kinematic threshold, m_{th} is fixed to the value of Eqn. 6.30.

The signal model contains 22 parameters,

$$\begin{aligned}\vec{\mathbf{s}} = &\{\gamma_{\text{S}},\, \gamma_{\text{P0}},\, \gamma_{\text{PT}},\, x_1,\, x_2\} \cup \{a, r\}\\ &\cup \{m_{892},\, \Gamma_{892},\, m_{1430},\, \Gamma_{1430},\, d\}\\ &\cup \{S_3,\, A_{\text{FB}},\, S_9\} \cup \{\vec{\alpha},\, m_{B^0}\}\end{aligned} \tag{6.44}$$

which are referred to as the 'angular distribution parameters', the 'LASS parameters', the '$m_{K\pi}$ lineshape parameters', the 'nuisance parameters' and the 'double-CB parameters' respectively.

The angular distribution parameters have an arbitrary scale, governed by the integral of the $m_{K\pi}$ lineshapes which are not individually normalised. As in Ref. [90], d is fixed to $1.6\,[\text{GeV}/c]^{-1}$. The parameter d is found to be uncorrelated to other parameters in high statistics pseudodata fits and in fits to the control channel when they are allowed to float. The systematic uncertainty incurred by the choice of fixed d is described in Section 7.3.2.

By contrast, the LASS parameters a and r are found to be rather correlated to each other although uncorrelated to other parameters. The systematic uncertainty due to fixed values of a and r is included in the discussion in Section 7.3.1.

In the case of the Isobar model:

$$\begin{aligned} \vec{\mathbf{s}} \mapsto \vec{\mathbf{s}} + \{\alpha_{800},\, \delta\beta_{800},\, m_{800},\, \Gamma_{800}\} \\ - \{a,\, r\}\,. \end{aligned} \tag{6.45}$$

That is, that the LASS parameters are replaced by parameters related to the $K_0^*(800)^0$ component.

The nuisance parameters are only present due to the shape of the acceptance in $\cos\theta_\ell$ and ϕ. The model is very weakly dependent on them. For the $B^0 \to K^+\pi^-\mu^+\mu^-$ (control channel $B^0 \to J/\psi\, K^+\pi^-$) fits, the nuisance parameters are locked to the values given in Ref. [29] (zero) and varied to assess the systematic bias due to this choice. The effect of these nuisance parameters is discussed in Section 7.3.9. The Crystal Ball parameters are fixed to the values found in a fit to the $m_{K\pi\mu\mu}$ spectrum of $B^0 \to J/\psi\, K^+\pi^-$ events, in Table 6.1.

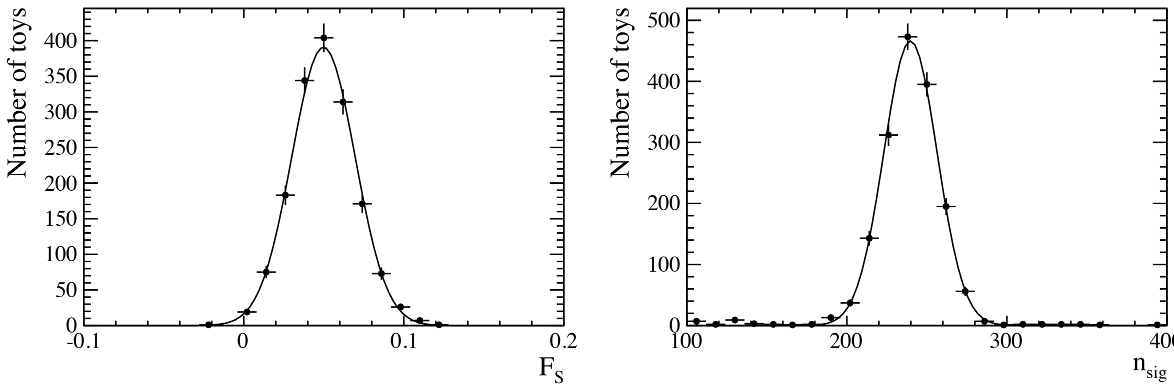
From inspection of Eqn. 2.157, the overall scale of the model is still undetermined. This freedom allows one angular distribution parameter to be fixed by normalisation convention. This is chosen to be $\gamma_{\mathrm{P}0}$. As a cross check, γ_{S} is fixed instead and has no effect on the fit result if an S-wave component is present. A second cross check is to fix $\gamma_{\mathrm{P}0}$ to a number a factor of 100 larger, a rescaling of the remaining angular distribution parameters is observed, which also has no effect on the fitted minimum in any other parameters as expected. One area where this choice does have some effect is in the coverage, as will be discussed in the following section. The choice to allow γ_{S} to float and explore unphysical (negative) values improves the coverage of the derived F_{S} parameter.

6.5 Validation using pseudodata

Fits to pseudodata are employed to check the model for degeneracies and to assess the sensitivity to F_{S} and the signal yield. Initially, fits to subsamples of the control channel, $B^0 \to J/\psi\, K^+\pi^-$, are used to obtain realistic starting values for parameters. Following this initial fit, pseudodata samples are generated with varying number of pseudodata per sample in order to assess the sensitivity of the model and ensure there are no large biases. Pseudodata are generated from the full 3D fit model (Eqn 6.2) using the accept-reject Monte Carlo method. In addition to validation of the model, this study drives the choice of binning scheme in q^2 which dictates the size of the real data samples that are fitted.

After performing this study, the fit is found to give an unbiased estimation for the parameters of interest with samples of 200 signal pseudodata per dataset. For a realistic signal-to-background ratio, this sample size corresponds to $100 - 200$ background pseudodata. The number of data generated for each sample is Poisson fluctuated about the total.

Figure 6.9: The fitted value of the parameters of interest. These correspond to the first two pull distributions of Fig. 6.10.



The pull of each of the parameters (and the derived value and pull of $F_S|_{796}^{996}$) is calculated for each fit result.

The resulting histograms of the fitted value of F_S and n_{sig} for 5000 pseudoexperiments are shown in Fig. 6.9. These two parameters are the parameters of interest. The spread of the values correspond to the expected precision on $F_S|_{796}^{996}$ and the branching fraction.

For toys where F_S is small, the coverage is improved by allowing the parameter γ_S to explore negative (unphysical) values. The other floating parameters ensure that the total pdf remains positive. Negative values of γ_S only occur in a small minority of fits. This can provide an unphysical error interval, but avoids the need to correct the coverage.

The histograms of pulls for the signal and background parameters are shown in Figs 6.10 and 6.11 respectively. The histograms and the result of making a fit of a Gaussian to these pull distributions are presented in Table 6.2.

Some of the background parameters have non-Gaussian pull distributions. This bias is restricted to the background parameters. The worst of these is n (the exponent of the $m_{K\pi}$ background model) and this parameter is seen to be uncorrelated to the parameters from which F_S is calculated. Furthermore no significant bias is observed in F_S, since the uncertainty calculation for the derived quantities takes the correlations into account. This effect, that the background parameters are broadly uncorrelated to the signal parameters is a consequence of including the background region of $m_{K\pi\mu\mu}$ in the fit.

6.5.1 Defining a q^2 binning scheme

A secondary result of fits to pseudodata, is that it drives the choice of the q^2 binning scheme employed to extract the result.

Fitting data samples with $O(200)$ signal events are found to provide an acceptable coverage of the signal parameters of interest, whereas fits to $O(100)$ are not. This is a balance of precision on F_S, for which more data are better, and fitting the finest q^2 binning scheme possible to provide more information about the shape of the differential branching

Figure 6.10: The results of pseudodata sample fits presented as histograms of pulls of the signal model parameters. Each sample has a number of data which is Poisson-fluctuated about 300 events. The non-Gaussian behaviour in some of the S-wave parameters is not at a level where it has an effect on the parameters of interest (F_S and n_{sig}).

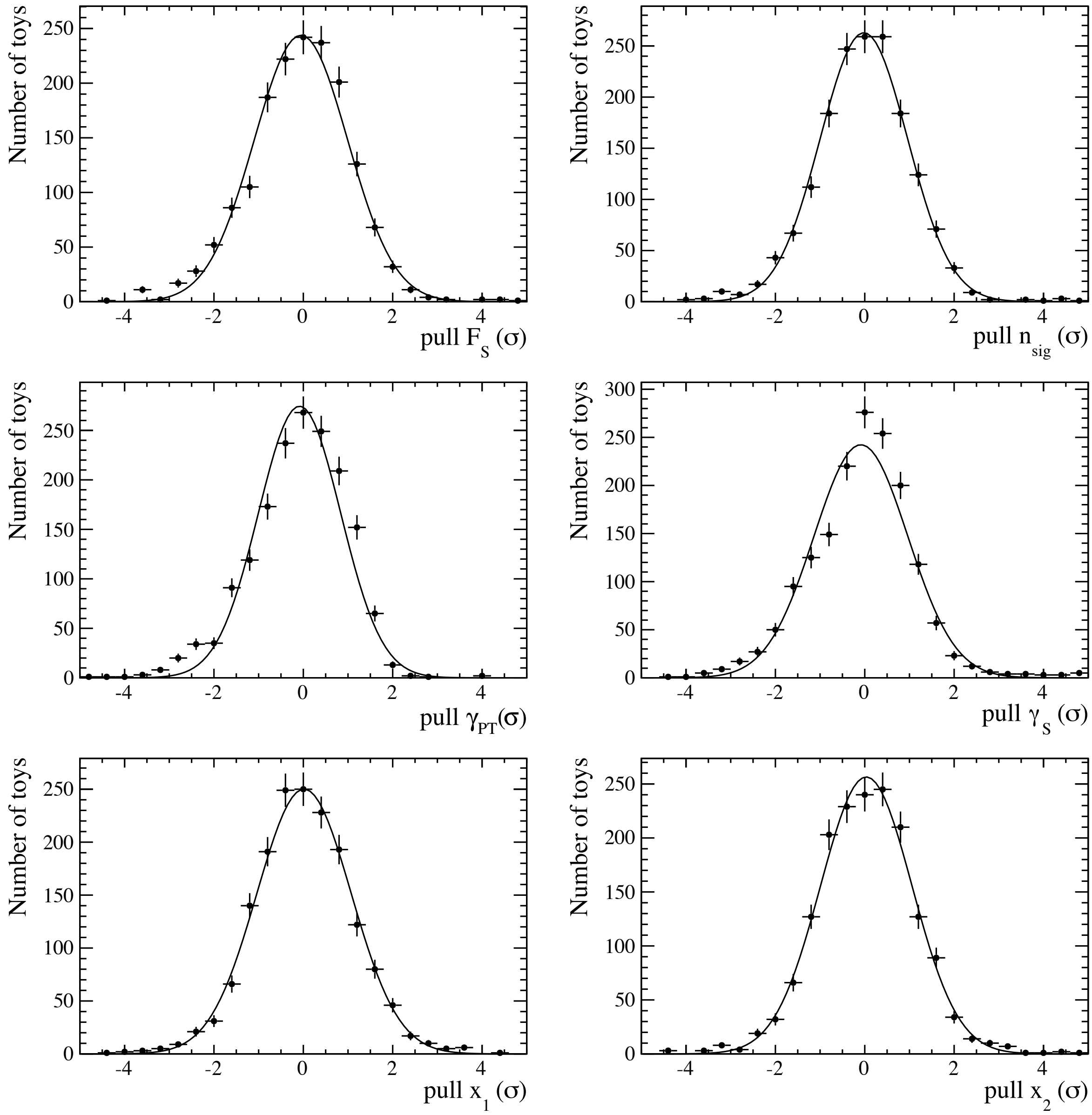
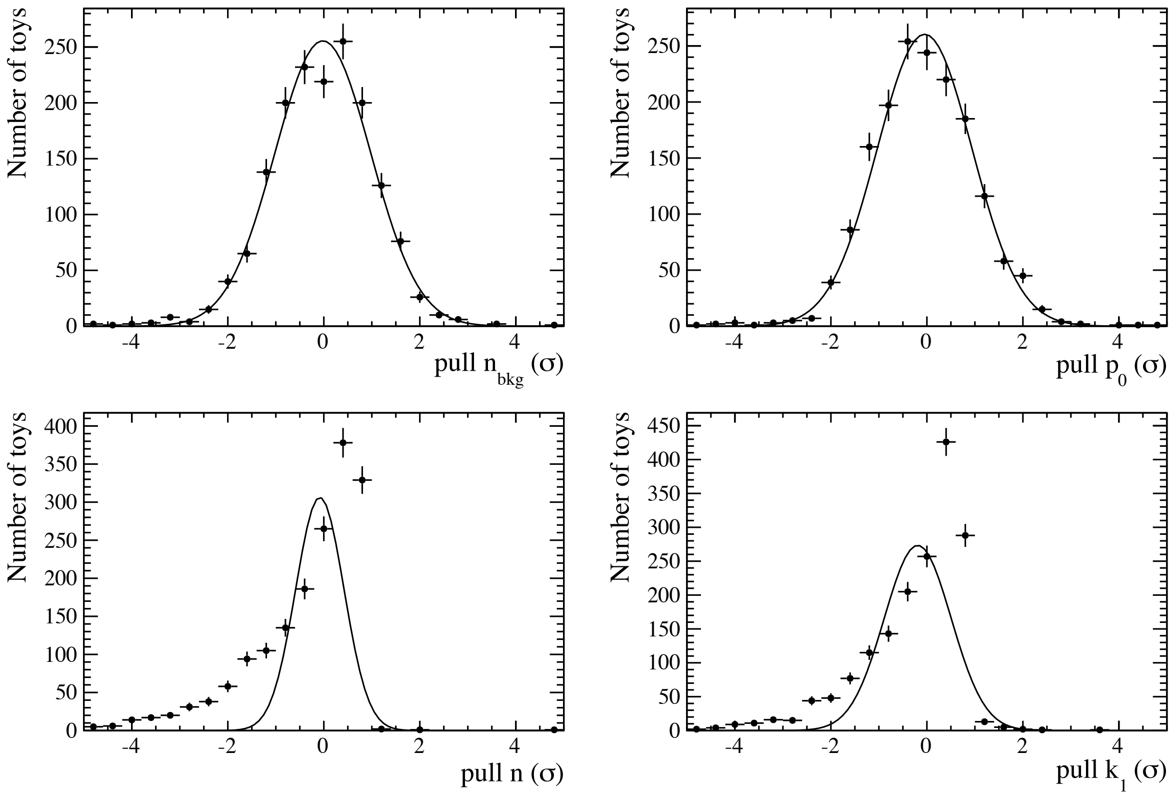


Table 6.2: The pulls for signal and background parameters in the full 3D model, shown as the 'raw' and fitted mean and widths corresponding to Figs 6.10 and 6.11. The parameters of interest are F_S, and the signal yield, n_{sig} which are used in calculating the differential branching fraction.

parameter	raw mean	RMS	fitted mean	fitted σ
F_S	-0.08 ± 0.03	1.11 ± 0.02	-0.05 ± 0.03	1.05 ± 0.02
n_{sig}	-0.03 ± 0.03	1.05 ± 0.02	-0.02 ± 0.03	0.98 ± 0.02
x_1	0.01 ± 0.03	1.10 ± 0.02	0.02 ± 0.03	1.06 ± 0.02
x_2	0.03 ± 0.03	1.09 ± 0.02	0.03 ± 0.03	1.02 ± 0.02
γ_S	-0.07 ± 0.03	1.15 ± 0.02	-0.09 ± 0.03	1.05 ± 0.02
γ_{PT}	-0.11 ± 0.03	1.05 ± 0.02	-0.08 ± 0.03	0.92 ± 0.02
n_{bkg}	-0.05 ± 0.03	1.05 ± 0.02	-0.02 ± 0.03	1.00 ± 0.02

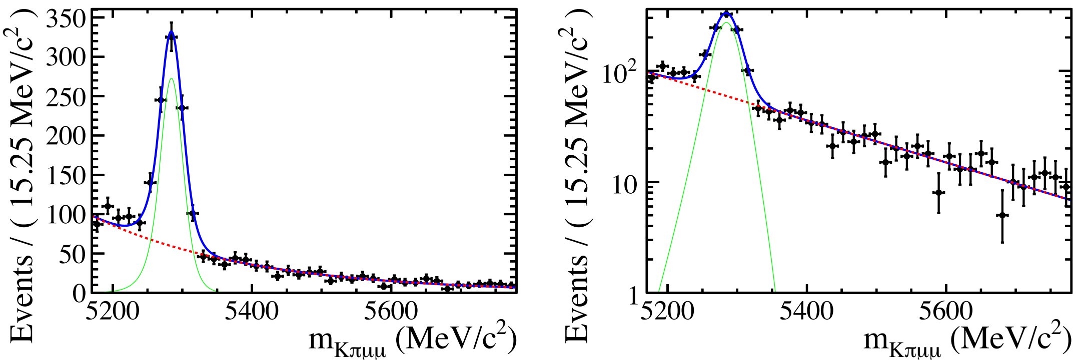
Figure 6.11: The results of pseudodata sample fits presented as histograms of pulls of the background model parameters. The lower two plots are the $m_{K\pi}$ background parameter, n, and $\cos\theta_K$ background parameter, k_1 which display non-Gaussian behaviour. These two parameters are uncorrelated to all signal parameters and this behavior has no effect on the parameters of interest. The upper two plots, particularly the background yield are correlated (as expected) to the differential branching fraction, however these parameters show Gaussian behavior.



q^2 bin $(\,\mathrm{GeV}^2/c^4)$	n_{sig}	n_{bkg}
$[0.1, 0.98]$	373 ± 21	135 ± 14
$[1.1, 2.5]$	240 ± 19	282 ± 20
$[2.5, 4.0]$	195 ± 19	442 ± 24
$[4.0, 6.0]$	367 ± 24	643 ± 30
$[6.0, 8.0]$	403 ± 26	733 ± 31
$[11.0, 12.5]$	349 ± 23	454 ± 25
$[15.0, 17.0]$	446 ± 25	389 ± 24
$[17.0, 19.0]$	256 ± 19	275 ± 20
$[1.1, 6.0]$	802 ± 36	1367 ± 43
$[15.0, 19.0]$	703 ± 31	664 ± 31

Table 6.3: The fitted signal yield, n_{sig}, of $B^0 \to K^+\pi^-\mu^+\mu^-$ events and background yield n_{bkg} in bins of q^2. The binning scheme follows the convention of Ref. [29] and provides at least 200 signal events per bin, for which the coverage is found to be acceptable and the fit is unbiased. The fitted spectrum for the region $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$ is shown in Fig. 6.12. The remaining fitted distributions are included in Appendix E.2.

Figure 6.12: Fit to the $m_{K\pi\mu\mu}$ spectrum for $B^0 \to K^+\pi^-\mu^+\mu^-$ decays shown on linear (left) and logarithmic (right) scales for the theoretically clean region $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$. The set of shape parameters for the double Crystal Ball, $\vec{\alpha}$, are fixed to those of Table 6.1. The yields extracted from this fit corresponds to the ninth row in Table 6.3. The remaining fitted spectra are given in Appendix E.2.



fraction. The binning scheme that is chosen is identical to that of Ref. [29] and provides at least 200 signal events per bin. Table 6.3 shows the fitted signal and background yields, n_{sig} and n_{bkg}, respectively. One should note that the purity is low since the background region of $m_{K\pi\mu\mu}$ is included. Figure 6.12 shows fits to $m_{K\pi\mu\mu}$ for the region $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$, as described in Section 6.1. The remainder of the fitted distributions which result in the yields in Table 6.3 are included in Appendix E.2.

6.6 Validation using control channel data

Fits under a variety of configurations are performed on $B^0 \to J/\psi\, K^+\pi^-$ data events. The default configuration of the model is the full 3D model (Eqn 6.2) with the LASS description of the S-wave component. The alternative Isobar model, is also fitted in the 3D configuration. In order to check the effect of the acceptance, the 2D fit is also performed

Table 6.4: Fitted parameter values from the 3D fit to control channel $B^0 \to J/\psi K^+\pi^-$ data. The S-wave component is described by the LASS model which is the default model choice. The angular distribution parameters have an arbitrary scale, governed by the integral of the $m_{K\pi}$ lineshapes which are not individually normalised.

parameter	fitted value	units
$F_S\|_{644}^{1200}$	12.2 ± 0.12	%
$F_S\|_{796}^{996}$	5.19 ± 0.056	%
n_{bkg}	41356 ± 272	(events)
n_{nsig}	395540 ± 654	(events)
γ_S	1.07 ± 0.01	(arb.)
γ_{PT}	0.0422 ± 0.0002	(arb.)
x_1	-0.2418 ± 0.002	(arb.)
x_2	0.071 ± 0.002	(arb.)
m_{892}	894.84 ± 0.06	MeV/c^2
Γ_{892}	48.7 ± 0.2	MeV/c^2
n	1.42 ± 0.02	–
p_0	-4.76 ± 0.04	$10^{-4}[\text{MeV}/c^2]^{-1}$
k_1	0.1 ± 0.1	–
k_2	0.1 ± 0.2	–

with no acceptance function. In all cases, values of $F_S|_{796}^{996}$ are found in the range $4-5\%$.

Some mismodelling is observed in the approximate region $\cos\theta_K \in [-0.8, -0.5]$ (at a level below the required statistical uncertainty for the statistics of fits to $B^0 \to K^+\pi^-\mu^+\mu^-$). The mismodelling is also seen around the same region of $\cos\theta_K$ by the CMS analysis [95]. This mismodelling was investigated and found to be consistent with a contribution from a state similar to the $Z(4430)^-$. The $Z(4430)^-$ is a candidate for a tetraquark state recently observed in the decay mode $B^0 \to \psi(2S)K^+\pi^-$ [90]. In the case of the control channel, this contribution would be of the form $B^0 \to X^-K^+$ with $X^- \to J/\psi\pi^-$ which is claimed to be observed by Belle [96]. The results of some studies performed in collaboration with others are given in Appendix D. The full analysis of the $B^0 \to J/\psi K^+\pi^-$ decay mode would require a more precise description of many other components. This is beyond the scope of this analysis which only requires that the model be valid for the statistics of the $B^0 \to K^+\pi^-\mu^+\mu^-$ data sample. An analysis within the LHCb collaboration is ongoing.

Projections of the fitted 3D model with the LASS description of the S-wave are shown in Fig. 6.13. The values of the fitted parameters are given in Table 6.4. This fit finds $F_S|_{796}^{996} = (5.19 \pm 0.06)\%$ and $F_S|_{644}^{1200} = (12.2 \pm 0.1)\%$. The Isobar model is also fitted to this data and the resulting fitted projections are shown in Fig. 6.14. The fitted parameters for this Isobar model fit are given in Table 6.5.

Figure 6.13: Projections of the 3D fit to control channel $B^0 \to J/\psi K^+\pi^-$ data. The S-wave component is described by the LASS model which is the default choice. Some mismodelling is observed in the region $\cos\theta_K \in [-0.8, -0.5]$. This is understood to be consistent with a contribution from a state similar to the $Z(4430)^-$ (see Appendix D).

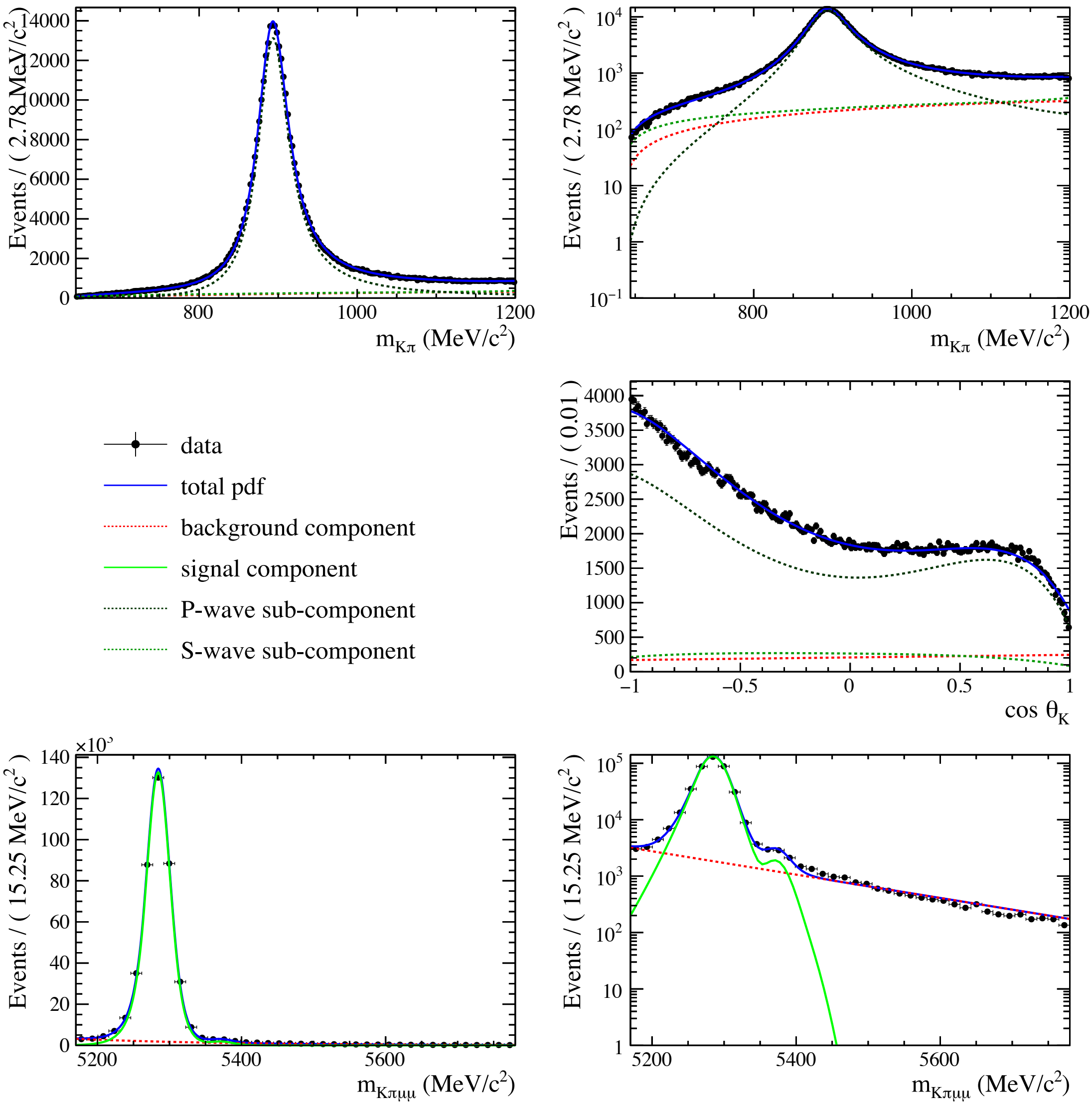
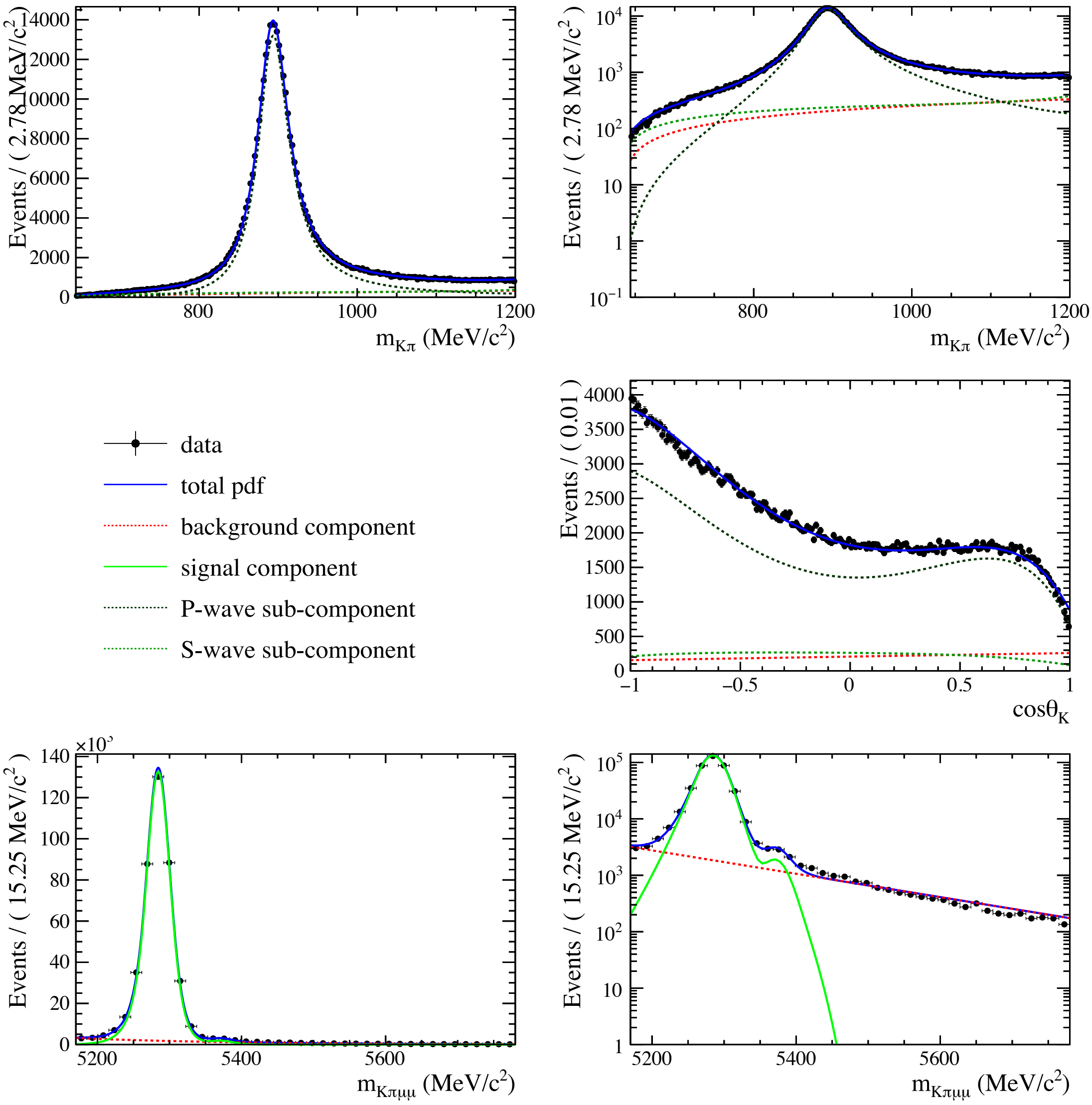


Table 6.5: Fitted parameter values from the 3D fit to control channel $B^0 \to J/\psi K^+\pi^-$ data. The S-wave component is described by the Isobar model which is used as an alternative model to the LASS in order to estimate the systematic uncertainty from the choice of model. The angular distribution parameters have an arbitrary scale, governed by the integral of the $m_{K\pi}$ lineshapes which are not individually normalised. In this fit the parameter γ_{P0} was allowed to float in place of γ_{S}.

parameter	fitted value	units
$F_{\mathrm{S}}\|_{644}^{1200}$	11.8 ± 0.16	%
$F_{\mathrm{S}}\|_{796}^{996}$	5.05 ± 0.063	%
n_{bkg}	41431 ± 290	(events)
n_{sig}	395464 ± 662	(events)
γ_{P0}	6931 ± 452	(arb.)
γ_{PT}	5739 ± 383	(arb.)
x_1	$-\,567 \pm 461$	(arb.)
x_2	15636 ± 614	(arb.)
α_{800}	0.69 ± 0.01	(arb.)
$\delta\beta_{800}$	3.18 ± 0.06	rad.
m_{892}	895.09 ± 0.06	MeV/c^2
Γ_{892}	48.5 ± 0.2	MeV/c^2
n	1.61 ± 0.03	–
p_0	$-\,4.77 \pm 0.04$	$10^{-4}[\,\mathrm{MeV}/c^2]^{-1}$
k_1	0.1 ± 0.2	–
k_2	0.2 ± 0.6	–

Figure 6.14: Projections of the 3D fit to control channel $B^0 \to J/\psi K^+\pi^-$ data. The S-wave component is described by the Isobar model which is used as an alternative model to the LASS in order to estimate the systematic uncertainty from the choice of model. As with Fig 6.13, some mismodelling is observed in the region $\cos\theta_K \in [-0.8, -0.5]$. This is understood to be consistent with a contribution from a state similar to the $Z(4430)^-$.



Chapter 7

Results

The fit is performed in all q^2 bins, resulting in the differential branching fraction measurement shown in Fig. 7.1. The results for the fractional size of the S-wave component ($F_{\rm S}$) are shown in Fig. 7.2. Both results are given in Table 7.1 which also shows the total systematic uncertainty described in Section 7.3. Multiplying the differential branching fractions by the bin widths, summing, and accounting for the vetoed regions as described in Section 6.3, yields the total branching fraction which is found to be

$$\mathcal{B}\left[B^0 \to K_1^*(892)^0\mu^+\mu^-\right] = (1.058^{+0.017}_{-0.016} \pm 0.013 \pm 0.070) \times 10^{-6}, \tag{7.1}$$

where the first uncertainty is statistical, the second systematic, and the third is due to the uncertainty on the branching fraction of the normalisation channel, $\mathcal{B}\left[B^0 \to J/\psi\, K^{*0}\right] \cdot \mathcal{B}\left[J/\psi \to \mu^+\mu^-\right]$.

Individual projections of the fitted distributions are shown in Figure 7.3 for the bin $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$ and in Figures 7.4 and 7.5 for the bins $q^2 \in [2.5, 4]\,\mathrm{GeV}^2/c^4$ and $q^2 \in [4, 6]\,\mathrm{GeV}^2/c^4$ respectively. The remainder are included in Appendix E.3.

7.1 Comparison to theoretical predictions

Theoretical predictions for the differential branching fraction at low q^2 are given in Ref. [23]. This reference provides predictions for the first four bins of q^2 where the uncertainties due to parametric sources of uncertainty, form-factor uncertainty, and an estimate of hadronic effects are separated out. From private communication with the authors of Ref. [23], an additional prediction was provided in the ranges $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$, $q^2 \in [15, 19]\,\mathrm{GeV}^2/c^4$, and additional information was given regarding the treatment of the three sources of uncertainty which are not precisely Gaussian, and therefore the total error is not exactly the quadratic sum of the three components. The value at high q^2 is calculated by the authors of Ref. [23] using the form factors given in Ref. [25]. These predictions are given in Table 7.2, and shown on Fig 7.1 as blue bands.

Figure 7.1: Results for the differential branching fraction of $B^0 \to K_1^*(892)^0\mu^+\mu^-$ in bins of q^2. The vertical error bars are divided into the statistical component alone (before the crossbar) and the extra contribution due to systematic sources of uncertainty and the normalisation channel (full error bar). The horizontal uncertainty denotes the q^2 bin. The blue shaded areas correspond to the SM predictions from Refs [23, 25] which are summarised in Table 7.2.

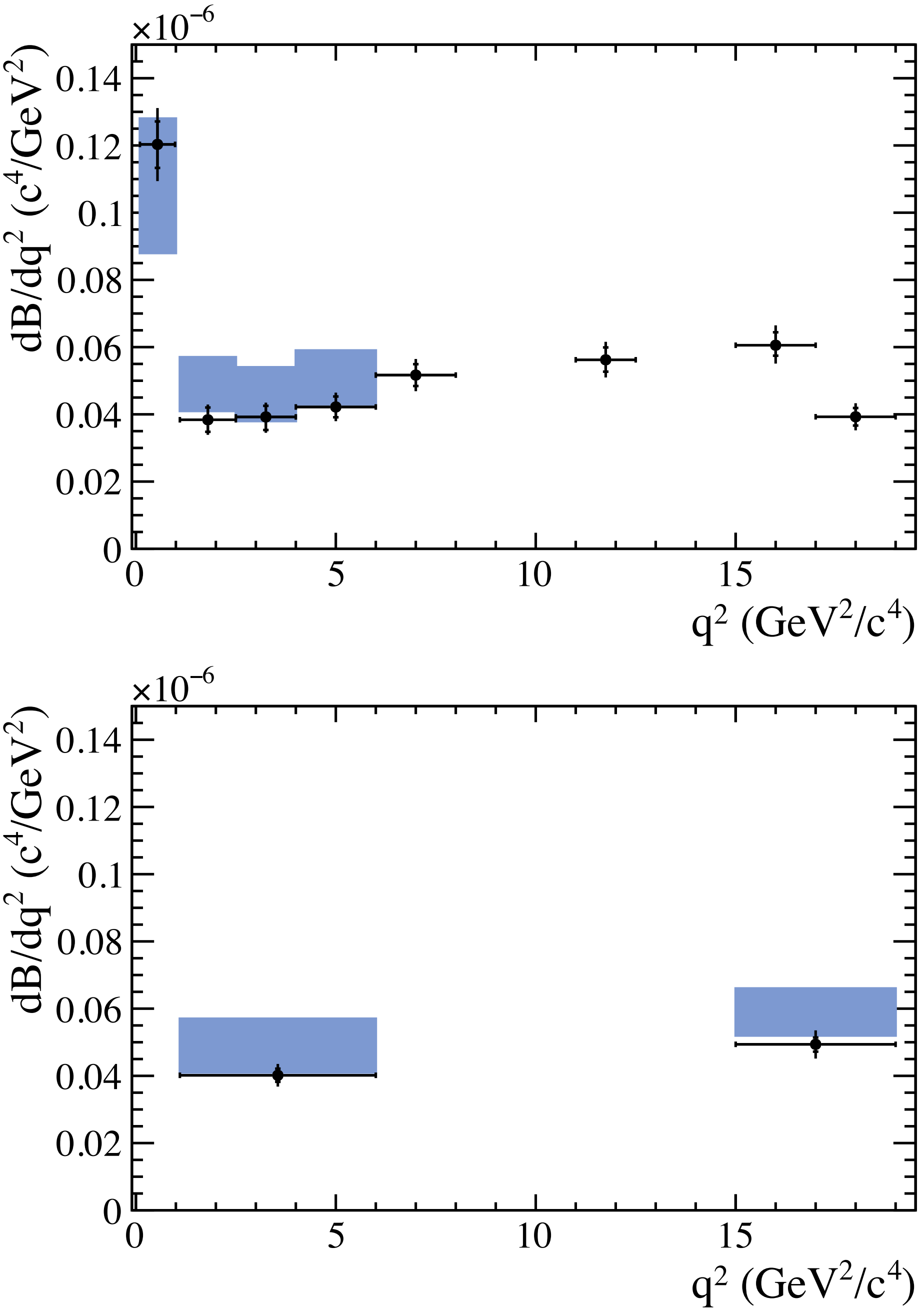


Figure 7.2: Results for the fraction of S-wave in bins of q^2. The left- (right-) hand figure corresponds to the range $m_{K\pi} \in [796, 996]\,\text{MeV}/c^2$ ($m_{K\pi} \in [644, 1200]\,\text{MeV}/c^2$). The vertical error bars are divided into the statistical component alone (before the crossbar) and the extra contribution due to systematic uncertainty. The horizontal uncertainty denotes the q^2 bin. The systematic contribution is small in all bins except for high-q^2 bins for $F_S|_{644}^{1200}$ (rightmost points on the right-hand plot) where it is just visible.

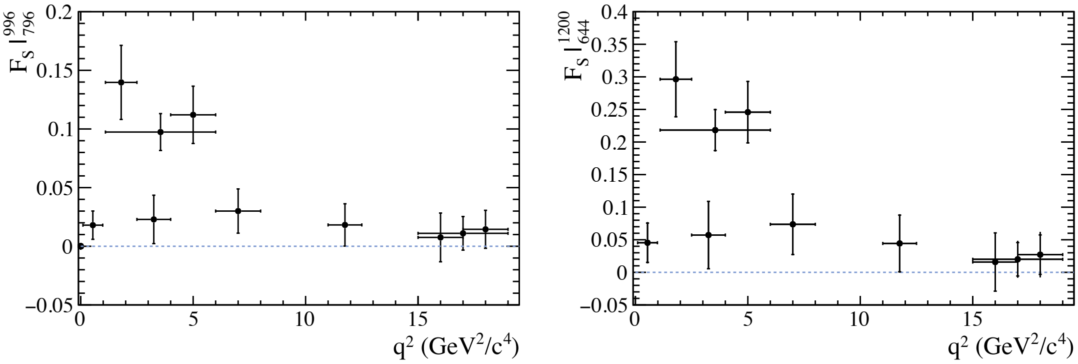


Table 7.1: Table of results of the 3D fit to extract F_S and the differential branching fraction. In all columns, the first uncertainty is statistical and the second systematic. In the final column, the third error is the systematic due to the branching fraction of the normalisation channel.

q^2 bin (GeV^2/c^4)	$F_S\|_{796}^{996}$	$F_S\|_{644}^{1200}$	$10^7 \times \text{d}\mathcal{B}/\text{d}q^2$ ($c^4/\,\text{GeV}^2$)
$[0.1, 0.98]$	$0.018^{+0.012}_{-0.012} \pm 0.0085$	$0.045^{+0.030}_{-0.030} \pm 0.0123$	$1.203^{+0.071}_{-0.069} \pm 0.030 \pm 0.080$
$[1.1, 2.5]$	$0.140^{+0.032}_{-0.032} \pm 0.0092$	$0.296^{+0.058}_{-0.058} \pm 0.0123$	$0.384^{+0.038}_{-0.035} \pm 0.010 \pm 0.025$
$[2.5, 4.0]$	$0.023^{+0.021}_{-0.021} \pm 0.0086$	$0.057^{+0.051}_{-0.051} \pm 0.0120$	$0.392^{+0.036}_{-0.034} \pm 0.008 \pm 0.026$
$[4.0, 6.0]$	$0.112^{+0.024}_{-0.024} \pm 0.0078$	$0.246^{+0.047}_{-0.047} \pm 0.0120$	$0.422^{+0.033}_{-0.030} \pm 0.011 \pm 0.028$
$[6.0, 8.0]$	$0.030^{+0.019}_{-0.019} \pm 0.0072$	$0.074^{+0.046}_{-0.046} \pm 0.0135$	$0.517^{+0.033}_{-0.032} \pm 0.010 \pm 0.034$
$[11.0, 12.5]$	$0.018^{+0.018}_{-0.018} \pm 0.0062$	$0.044^{+0.044}_{-0.044} \pm 0.0132$	$0.562^{+0.036}_{-0.035} \pm 0.013 \pm 0.037$
$[15.0, 17.0]$	$0.007^{+0.018}_{-0.018} \pm 0.0053$	$0.016^{+0.039}_{-0.039} \pm 0.0111$	$0.606^{+0.035}_{-0.033} \pm 0.021 \pm 0.040$
$[17.0, 19.0]$	$0.015^{+0.017}_{-0.017} \pm 0.0089$	$0.028^{+0.031}_{-0.031} \pm 0.0188$	$0.393^{+0.026}_{-0.025} \pm 0.017 \pm 0.026$
$[1.1, 6.0]$	$0.097^{+0.016}_{-0.016} \pm 0.0084$	$0.218^{+0.031}_{-0.031} \pm 0.0122$	$0.402^{+0.020}_{-0.019} \pm 0.008 \pm 0.027$
$[15.0, 19.0]$	$0.011^{+0.015}_{-0.015} \pm 0.0068$	$0.020^{+0.026}_{-0.026} \pm 0.0122$	$0.494^{+0.020}_{-0.022} \pm 0.017 \pm 0.033$

Figure 7.3: The result of the full 3D fit to the data for the bin $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$ which is the sum of the second, third and fourth bins in Fig. 7.1. This bin is the theoretically favoured region away from the photon pole and before the $c\bar{c}$ effects are thought to become sizeable.

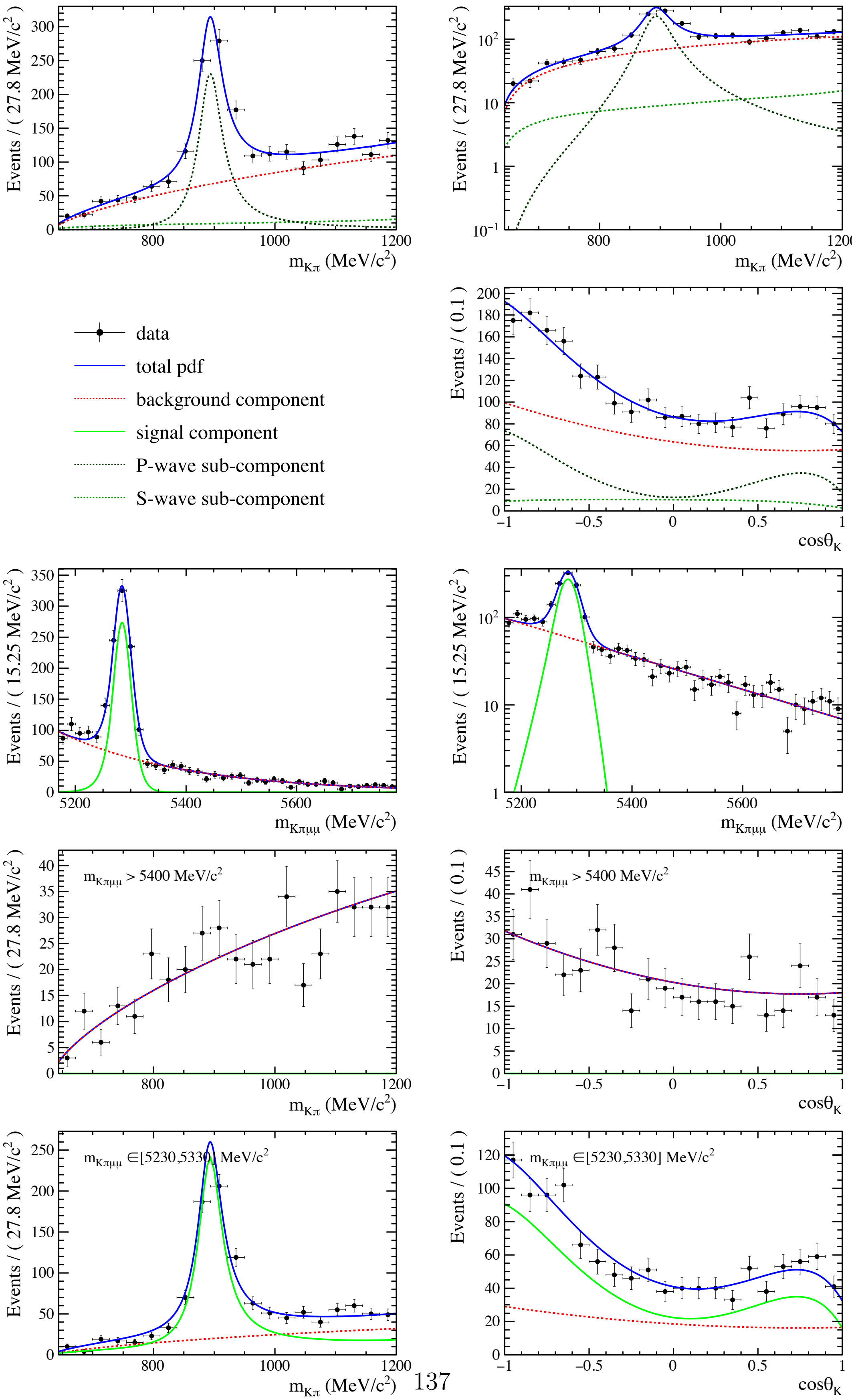


Figure 7.4: The result of the full 3D fit to the data for the bin $q^2 \in [2.5, 4]\,\mathrm{GeV}^2/c^4$ which is the third bin in Figs 7.2 and 7.1.

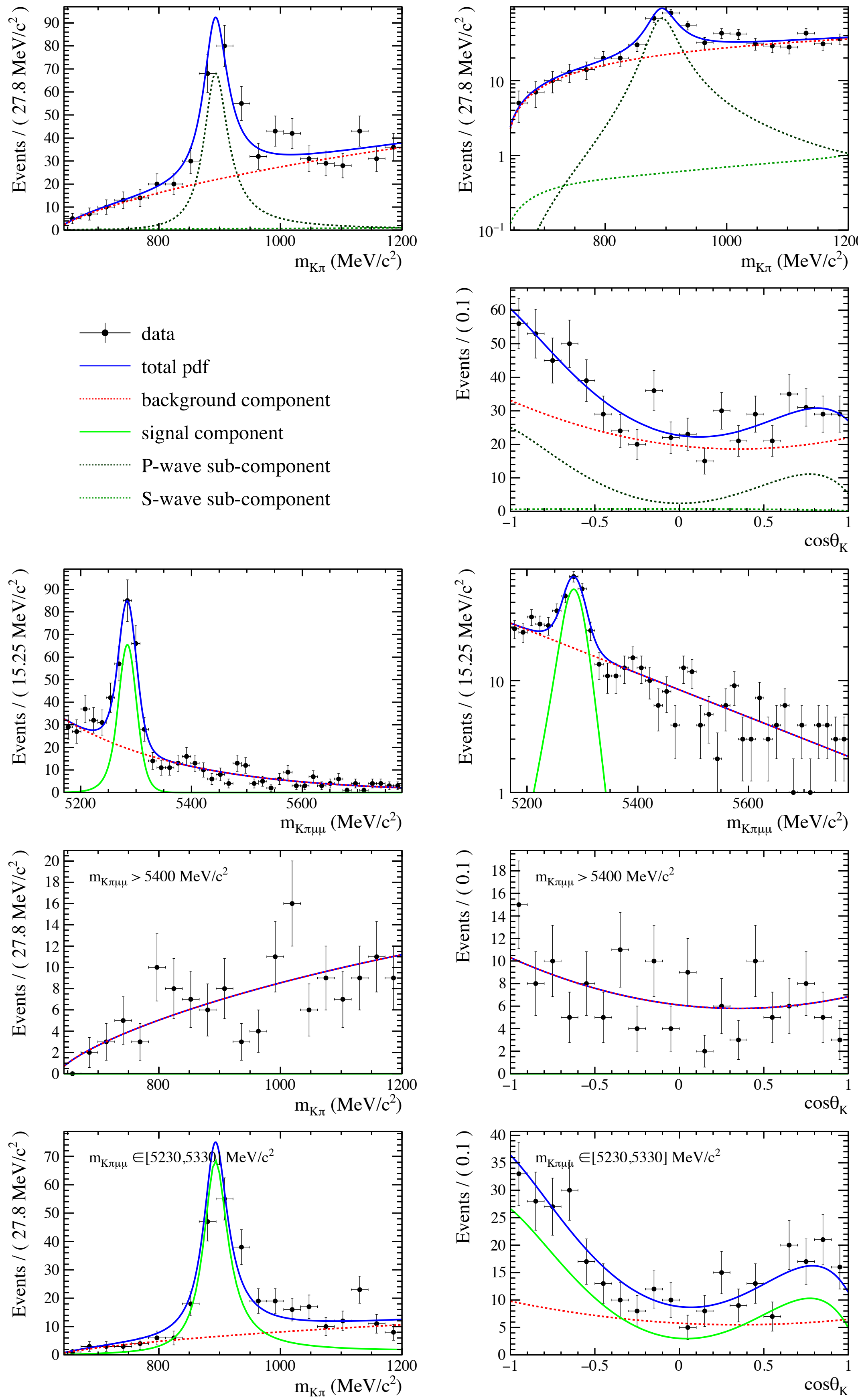


Figure 7.5: The result of the full 3D fit to the data for the bin $q^2 \in [4, 6]\,\text{GeV}^2/c^4$ which is the fourth bin in Figs 7.2 and 7.1.

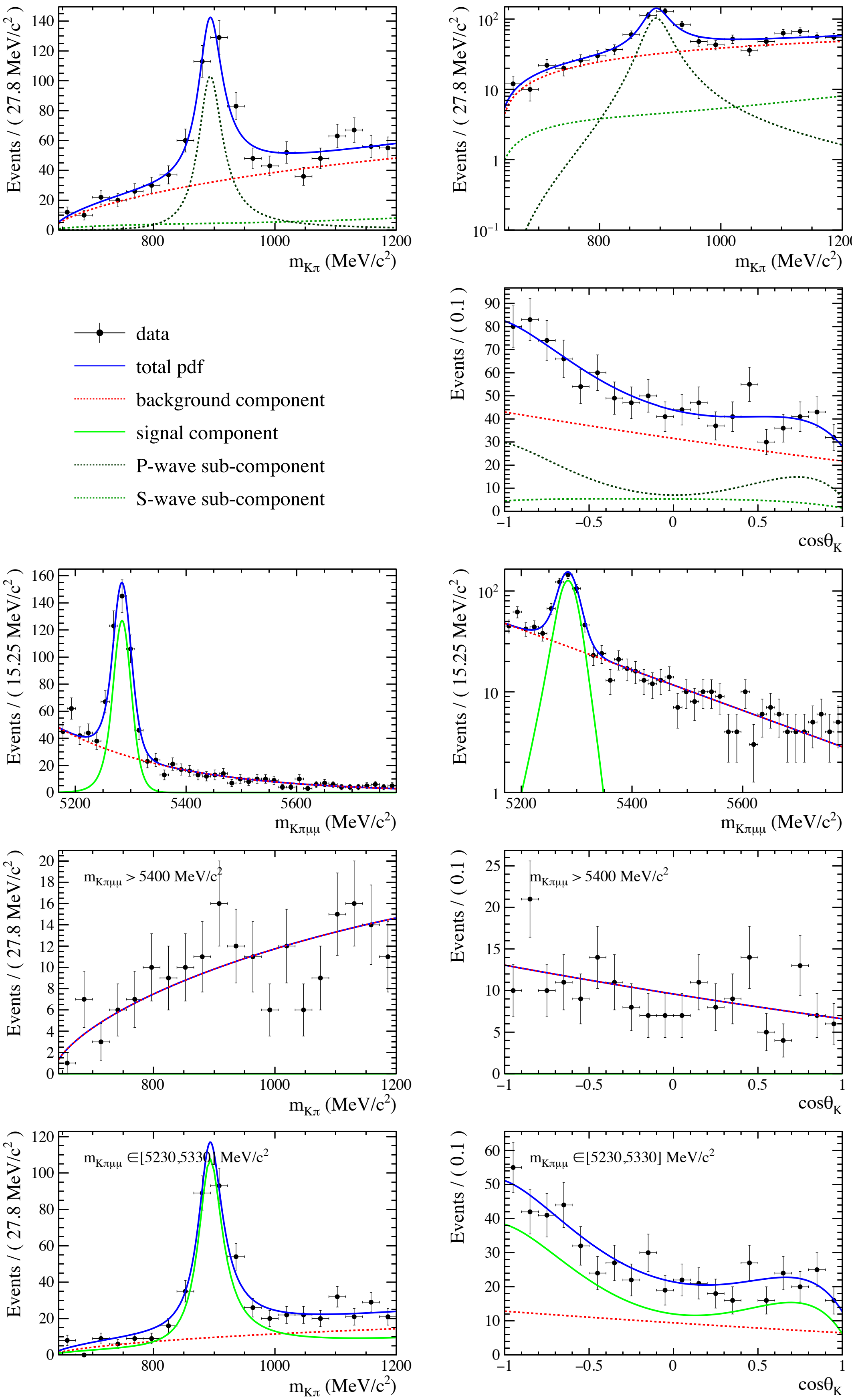


Table 7.2: Theoretical predictions for the differential branching fraction in the SM. The central value, and projected uncertainties for the first four bins are taken from Ref. [23], the value for $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$ and the total errors are from a private communication with the authors of Ref. [23]. The value at high q^2 is computed by the authors of Ref. [23] using the form factors from Ref. [25].

q^2 range (GeV^2/c^4)	prediction $10^7 \times \mathrm{d}\mathcal{B}/\mathrm{d}q^2$ (c^4/GeV^2)	total uncertainty
$[0.1, 1.0]$	$1.08 \pm 0.07 \pm 0.15 \pm 0.06$	± 0.20
$[1.1, 2.5]$	$0.49 \pm 0.07 \pm 0.07 \pm 0.02$	± 0.08
$[2.5, 4.0]$	$0.46 \pm 0.06 \pm 0.06 \pm 0.02$	± 0.08
$[4.0, 6.0]$	$0.51 \pm 0.06 \pm 0.06 \pm 0.03$	± 0.08
$[1.1, 6.0]$	$0.49 \pm 0.03 \pm 0.06 \pm 0.02$	± 0.08
$[15.0, 19.0]$	0.59	± 0.07

As can be seen in Fig. 7.1, the results are compatible with the SM within uncertainties. There is some tension with the prediction at high q^2 for the region $q^2 \in [15, 19]\,\mathrm{GeV}^2/c^4$. However this prediction is taken purely from form factors calculated at using the lattice QCD method [25], if instead one combines the form factors from the LCSR and lattice results in order to provide predictions, then the SM prediction in this region becomes $(0.54 \pm 0.06) \times 10^{-7} c^4/\mathrm{GeV}^2$ which is more compatible with the result of this analysis. The results across q^2, are low with respect to the predictions. As discussed in Section 2.4, this is consistent with other $b \to s\ell^+\ell^-$ decay processes.

7.2 Cross checks of the result

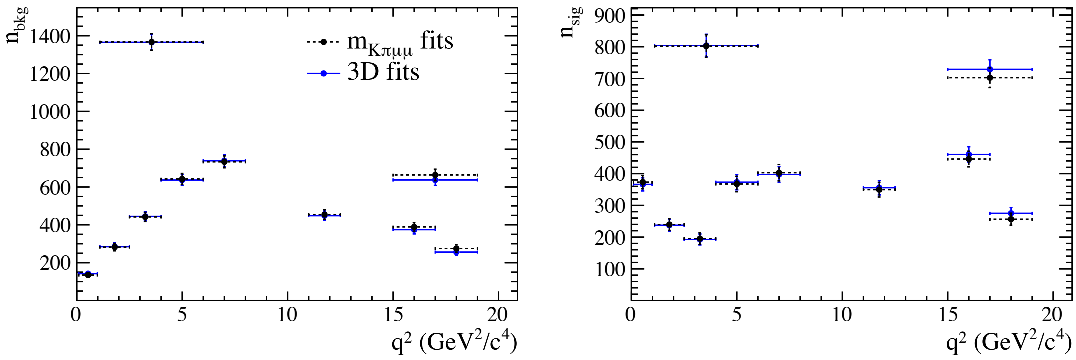
Several cross checks are performed on the result presented above. They are described in this section.

Since the background is described by a threshold model in $m_{K\pi}$ and a polynomial model in $\cos\theta_K$, there is the possibility that signal S-wave, which has a similar $m_{K\pi}$-dependence and is linear in $\cos\theta_K$, might be mistaken for background. However, Fig. 7.6 shows that this is not the case. The signal yields obtained by fitting the $m_{K\pi\mu\mu}$ spectrum (given in Table 6.3) are completely comparable with the yield parameters from the 3D fit.

From Eqn. 2.149 it is possible to extract the fraction of longitudinal polarised amplitudes (F_{L}) from the fitted parameters of the 3D fit. Good agreement is found with the values reported in Ref. [29], as shown in Fig. 7.7. Whilst the angular analysis described in Ref. [29] was performed on the same Run I data sample, the present analysis employs a different selection, with a wider window of $m_{K\pi}$, and does not include the angles $\cos\theta_\ell$ and ϕ. It is therefore to be expected that these values should not agree perfectly, in particular as the extra angular information is known to provide a somewhat orthogonal fit.

The fast variation between the first and second bins is understood in terms of the

Figure 7.6: Plots of the background (left) and signal (right) yields from fits of $m_{K\pi\mu\mu}$ alone (black points, Eqn. 6.3) overlaid with the yields from the 3D fit (blue points). Uncertainties are statistical only. The values are compatible, which implies that there is no exchange of S-wave signal with background in the 3D fit.



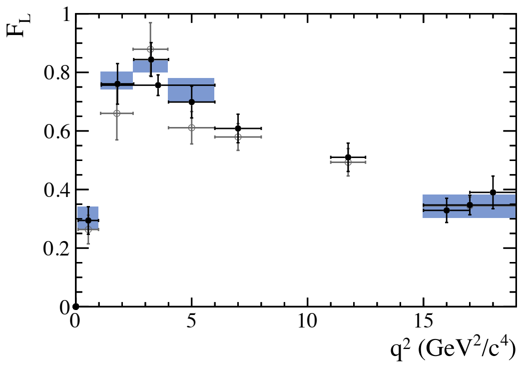
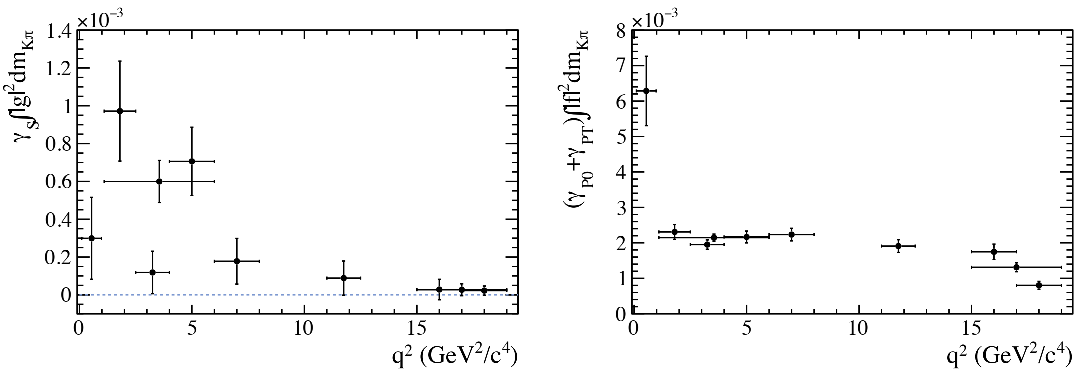


Figure 7.7: A comparison of the F_{L} angular parameter extracted from the 3D fit (black points showing only the statistical uncertaintiy), with the values reported in Ref. [29] (grey points) and the SM predictions (shaded areas) from Ref. [23]. Good agreement is observed.

Figure 7.8: Individual components that enter the calculation of $F_{\mathrm{S}}|_{644}^{1200}$. The first bin is expected to contain a dominant contribution from the photon pole which contributes to the P-wave and does not contribute to the S-wave. The left- (right-) hand figure corresponds to the S-wave (P-wave) amplitude given by Eqn. 7.2 (7.3) which enter into the calculation of F_{S} in Eqn. 6.36. The uncertainties shown are statistical only.



photon pole in the P-wave state. This can be seen by plotting the individual terms entering into the F_{S} calculation,

$$\gamma_{\mathrm{S}} \cdot \int_a^b \left|g_{K^*_0}^{\mathrm{LASS}}(m_{K\pi})\right|^2 \mathrm{d}m_{K\pi}, \tag{7.2}$$

which is the squared magnitude of the total S-wave amplitude, and

$$(\gamma_{\mathrm{P0}} + \gamma_{\mathrm{PT}}) \cdot \int_a^b \left|f_{892}(m_{K\pi})\right|^2 \mathrm{d}m_{K\pi}, \tag{7.3}$$

which is the squared magnitude of the total P-wave amplitude. These terms are shown in Eqn. 6.36. Integrating over the window $m_{K\pi} \in [644, 1200]\,\mathrm{MeV}/c^2$, these are shown in Fig. 7.8. Since the lowest q^2 bin is dominated by the photon pole, which is not present in the S-wave decay, the P-wave component is expected to dominate there and F_{S} will be much smaller in the first bin than in subsequent low-q^2 bins.

The remaining variation (between the second, third, and fourth bins) is not altered by even aggressive simplifications to the acceptance. One such simplification is to exclude all dimensions except for $\cos\theta_K$ from the angular efficiency function so that

$$\epsilon(q^2,\, m_{K\pi},\, \cos\theta_\ell,\, \cos\theta_K,\, \phi) \mapsto \epsilon(\cos\theta_K). \tag{7.4}$$

This test also has the effect of excluding any possible effect of the nuisance parameters. The possibility that the fit has landing in a local minimum is tested by randomising the parameters' starting values in repeated fits. These randomised fits are repeated at least twenty times for each of the second, third, and fourth q^2 bins. Furthermore, if the pattern were caused by a local minimum, then changing the model should have an effect. The pattern is not altered by performing a fit with the Isobar model, nor upon varying the $m_{K\pi}$-lineshape parameters. The effect of any remaining specific backgrounds is small, as will be shown in Section 7.3.6. The hypothesis of a statistical fluctuation is explored. Unfortunately, the assessment of a fluctuation effecting the F_{S} parameter is rather difficult as there are no real predictions for the shape of F_{S} in q^2. It is often asserted that F_{S} should behave roughly as F_{L}. Indeed in Ref. [30] the S-wave amplitudes are shown to be similar to the longitudinal P-wave amplitudes. The assertion that $F_{\mathrm{S}} \sim F_{\mathrm{L}}$ is used in a rather coarse way to assess the level of this variation. A spline interpolation is performed on the F_{L} distribution which is then scaled, by minimising the χ^2, to the distribution of $F_{\mathrm{S}}|_{644}^{1200}$. The χ^2/ndf of the spline fit to the $F_{\mathrm{S}}|_{644}^{1200}$ data is 13.9/6 which has a corresponding probability of 3.4%. The spline interpolation of F_{L} and the distribution of $F_{\mathrm{S}}|_{644}^{1200}$ is shown in Fig. 7.9.

The differential branching fraction has recently been measured by the CMS collaboration [95], and previously by the BaBar [97], Belle [98], and CDF [99] collaborations. Figure 7.10 shows a comparison of the results presented in this thesis with these values. Several different binning schemes are used across the different measurements. However, broadly speaking, the results are in good agreement.

Figure 7.9: An assessment of the variation in $F_{\rm S}|_{644}^{1200}$ as a statistical effect. The $F_{\rm L}$ shape is extracted using a spline interpolation (left), this is scaled to the distribution of $F_{\rm S}|_{644}^{1200}$ (right) which yields a χ^2 probability of 3.4%.

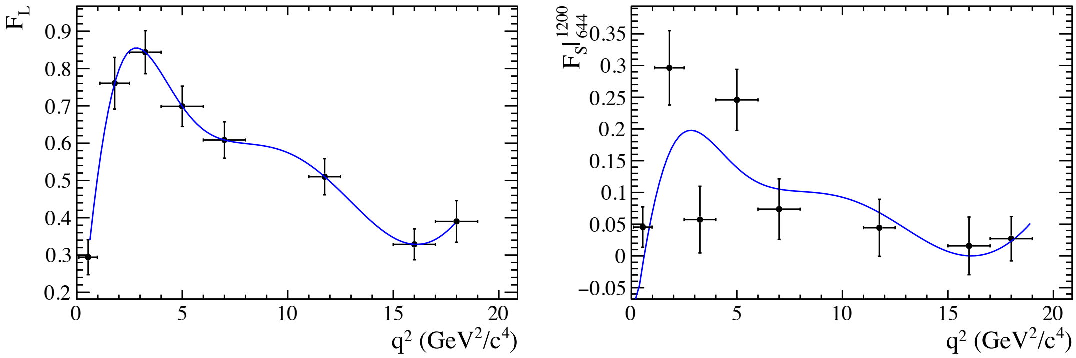
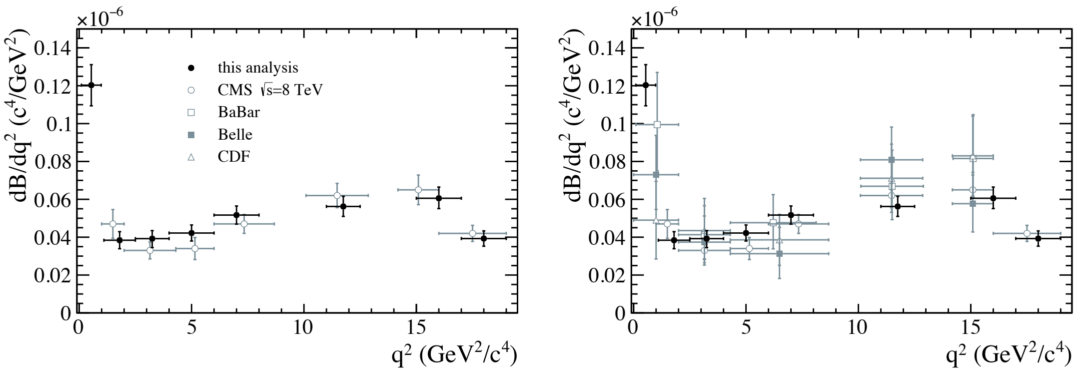


Figure 7.10: A comparison to previous measurements of the differential branching fraction. The left-hand figure shows the recent results of the CMS analysis (open circles) [95] compared to the results presented in this thesis. The right-hand plot shows the same, in addition to the most recent BaBar [97], Belle [98] results (open and filled squares respectively) as well as results from CDF (open triangles) [99]. All the results are in agreement.



7.3 Systematic uncertainties

The measurement of the (differential) branching fraction presented in this thesis is dominated by the the uncertainty on the branching fraction of the normalisation channel, $\mathcal{B}[B^0 \to J/\psi K^{*0}]$. This uncertainty is considered separately from the systematic uncertainties presented in this section. In both the measurement of the branching fraction and F_S, the statistical uncertainty dominates over the systematic.

For the branching fraction measurement, the dominant systematic uncertainty is due to the model dependence of the simulation, which affects the ratio of efficiencies. The leading sources of systematic uncertainty on the measurement of F_S is the modelling of the S-wave lineshape and residual differences between data and simulation which affects the determination of the angular efficiency. In general, the uncertainty on F_S is a small effect, when propagated to the branching fraction.

A common strategy is employed to assess the size of various modelling effects and several of the effects related to the description of the angular efficiency. This strategy is to generate very large pseudodata samples with a realistic alternative choice of model (or efficiency function). The pseudodata are then fitted with the nominal model and the alternative fit, the difference between the fitted values of F_S and the signal yields under each configuration is recorded. This is repeated for around 50 samples (or pseudoexperiments), and the size of the systematic bias is taken as the average of the differences observed in each pseudoexperiment. High statistics are used in order to exclude any statistical effect, typically $O(10^6)$ pseudodata per sample is used. Effects that are small require a larger sample size in order that the be resolved above the statistical precision.

The remainder of this section is a discussion of the assessment of the size of these effects. A summary is provided in Tables 7.3–7.5.

7.3.1 S-wave lineshape

The dominant source of systematic uncertainty on F_S is the choice of model used to describe the S-wave lineshape. It is also possible that this model choice might have a biasing effect on the signal yield, n_{sig}, since this parameter floats in the same fit. However the variation in n_{sig} that is observed to occur is well below the statistical uncertainty of even very large samples of pseudodata. With the inclusion of $m_{K\pi\mu\mu}$ in the fit, which provides a strong constraint on the yields, this secondary effect can be safely neglected. The S-wave shape choice is therefore considered as a standalone uncertainty on F_S which is directly propagated to the branching fraction according to the derivative of formula given in Eqn 6.36.

Several variations of the S-wave lineshape are tested. The average of the observed differences (per pseudoexperiment) is taken as the systematic variation due to the specific model choice. With several possible model variations to assess the same systematic effect,

Table 7.3: A summary of the systematic uncertainties applicable to F_{S}. The columns are split by q^2 range in units of GeV^2/c^4. The S-wave model choice is the dominant uncertainty in most bins, the second largest being the effect of residual differences between data and simulation due to the presence of S-wave in the control channel. The first five rows are propagated to the branching fraction neglecting secondary corrlation, for example to the signal yield. The last four rows are propagated taking correlations into account. Continued in Table 7.4.

		[0.1, 0.98]		[1.1, 2.5]		[2.5, 4.0]	
	§	$F_{\mathrm{S}}\vert_{796}^{996}$	$F_{\mathrm{S}}\vert_{644}^{1200}$	$F_{\mathrm{S}}\vert_{796}^{996}$	$F_{\mathrm{S}}\vert_{644}^{1200}$	$F_{\mathrm{S}}\vert_{796}^{996}$	$F_{\mathrm{S}}\vert_{644}^{1200}$
S-wave model	7.3.1	0.0065	0.0022	0.0070	0.0013	0.0066	0.0017
simulation stats	7.3.4	0.0006	0.0013	0.0004	0.0009	0.0004	0.0009
meson radius	7.3.2	0.0010	0.0018	0.0010	0.0018	0.0010	0.0019
$B(m_{K\pi})$	7.3.7	0.0009	0.0017	0.0012	0.0021	0.0006	0.0011
re-sampling	7.3.8	0.0033	0.0071	0.0034	0.0064	0.0035	0.0076
nuisances	7.3.9	0.0000	0.0000	0.0009	0.0016	0.0002	0.0003
$\Lambda_b^0 \to pK^-\mu^+\mu^-$	7.3.6	0.0001	0.0002	0.0001	0.0001	0.0001	0.0002
$B_s^0 \to \phi(1020)\mu^+\mu^-$	7.3.6	0.0002	0.0004	0.0002	0.0003	0.0002	0.0004
kaon mismodelling	7.3.3	0.0024	0.0053	0.0021	0.0047	0.0017	0.0038
pion mismodelling	7.3.3	0.0035	0.0077	0.0040	0.0087	0.0037	0.0080
sum in quad.		**0.0085**	**0.0123**	**0.0092**	**0.0123**	**0.0086**	**0.0120**

		[4.0, 6.0]		[6.0, 8.0]		[11.0, 12.5]	
	§	$F_{\mathrm{S}}\vert_{796}^{996}$	$F_{\mathrm{S}}\vert_{644}^{1200}$	$F_{\mathrm{S}}\vert_{796}^{996}$	$F_{\mathrm{S}}\vert_{644}^{1200}$	$F_{\mathrm{S}}\vert_{796}^{996}$	$F_{\mathrm{S}}\vert_{644}^{1200}$
S-wave model	7.3.1	0.0055	0.0015	0.0037	0.0004	0.0009	0.0011
simulation stats	7.3.4	0.0004	0.0009	0.0004	0.0008	0.0007	0.0015
meson radius	7.3.2	0.0011	0.0021	0.0011	0.0023	0.0011	0.0031
$B(m_{K\pi})$	7.3.7	0.0007	0.0013	0.0007	0.0013	0.0007	0.0016
re-sampling	7.3.8	0.0032	0.0069	0.0035	0.0075	0.0034	0.0071
nuisances	7.3.9	0.0000	0.0001	0.0003	0.0005	0.0003	0.0007
$\Lambda_b^0 \to pK^-\mu^+\mu^-$	7.3.6	0.0002	0.0003	0.0001	0.0003	0.0001	0.0002
$B_s^0 \to \phi(1020)\mu^+\mu^-$	7.3.6	0.0003	0.0005	0.0003	0.0006	0.0002	0.0004
kaon mismodelling	7.3.3	0.0023	0.0052	0.0032	0.0072	0.0028	0.0061
pion mismodelling	7.3.3	0.0035	0.0077	0.0038	0.0081	0.0039	0.0084
sum in quad.		**0.0078**	**0.0120**	**0.0072**	**0.0135**	**0.0062**	**0.0132**

Table 7.4: Continued from Table 7.3. A summary of the systematic uncertainties applicable to F_{S}. The columns are split by q^2 range in units of GeV^2/c^4. The S-wave model choice is the dominant uncertainty in most bins, the second largest being the effect of residual differences between data and simulation due to the presence of S-wave in the control channel. The first five rows are propagated to the branching fraction neglecting secondary corrlation, for example to the signal yield. The last four rows are propagated taking correlations into account.

		$[15.0, 17.0]$		$[17.0, 19.0]$	
	§	$F_{\mathrm{S}}\|_{796}^{996}$	$F_{\mathrm{S}}\|_{644}^{1200}$	$F_{\mathrm{S}}\|_{796}^{996}$	$F_{\mathrm{S}}\|_{644}^{1200}$
S-wave model	7.3.1	0.0031	0.0065	0.0080	0.0169
simulation stats	7.3.4	0.0004	0.0008	0.0007	0.0015
meson radius	7.3.2	0.0012	0.0025	0.0003	0.0005
$B(m_{K\pi})$	7.3.7	0.0016	0.0033	0.0024	0.0045
re-sampling	7.3.8	0.0036	0.0075	0.0031	0.0068
nuisances	7.3.9	0.0003	0.0006	0.0002	0.0004
$\Lambda_b^0 \to pK^-\mu^+\mu^-$	7.3.6	0.0000	0.0000	0.0000	0.0000
$B_s^0 \to \phi(1020)\mu^+\mu^-$	7.3.6	0.0002	0.0005	0.0000	0.0000
kaon mismodelling	7.3.3	0.0004	0.0008	0.0001	0.0002
pion mismodelling	7.3.3	0.0012	0.0024	0.0004	0.0009
sum in quad.		**0.0053**	**0.0111**	**0.0089**	**0.0188**

		$[1.1, 6.0]$		$[15.0, 19.0]$	
	§	$F_{\mathrm{S}}\|_{796}^{996}$	$F_{\mathrm{S}}\|_{644}^{1200}$	$F_{\mathrm{S}}\|_{796}^{996}$	$F_{\mathrm{S}}\|_{644}^{1200}$
S-wave model	7.3.1	0.0063	0.0015	0.0054	0.0113
meson radius	7.3.4	0.0011	0.0020	0.0007	0.0012
simulation stats	7.3.2	0.0008	0.0019	0.0005	0.0012
$B(m_{K\pi})$	7.3.7	0.0006	0.0012	0.0021	0.0038
re-sampling	7.3.8	0.0034	0.0074	0.0034	0.0072
nuisances	7.3.9	0.0001	0.0002	0.0003	0.0006
$\Lambda_b^0 \to pK^-\mu^+\mu^-$	7.3.6	0.0004	0.0007	0.0001	0.0002
$B_s^0 \to \phi(1020)\mu^+\mu^-$	7.3.6	0.0006	0.0011	0.0003	0.0005
kaon mismodelling	7.3.3	0.0019	0.0042	0.0007	0.0016
pion mismodelling	7.3.3	0.0037	0.0080	0.0003	0.0005
sum in quad.		**0.0084**	**0.0122**	**0.0068**	**0.0142**

Table 7.5: A summary of the systematic uncertainties applicable to the differential branching fraction. The columns are split by q^2 range in units of GeV^2/c^4. All values are in units of $10^{-7}c^4/\,\mathrm{GeV}^2$.

	§	$[0.1, 0.98]$	$[1.1, 2.5]$	$[2.5, 4.0]$	$[4.0, 6.0]$	$[6.0, 8.0]$
ϵ/ϵ model	7.3.5	0.0173	0.0039	0.0026	0.0019	0.0065
kaon mismodelling	7.3.3	0.0111	0.0038	0.0030	0.0037	0.0018
pion mismodelling	7.3.3	0.0153	0.0068	0.0047	0.0064	0.0050
re-sampling	7.3.8	0.0151	0.0042	0.0049	0.0070	0.0054
$\Lambda_b^0 \to pK^-\mu^+\mu^-$	7.3.6	0.0007	0.0002	0.0003	0.0005	0.0005
$B_s^0 \to \phi(1020)\mu^+\mu^-$	7.3.6	0.0011	0.0003	0.0003	0.0005	0.0006
F_{S} uncorrelated	–	0.0045	0.0019	0.0012	0.0017	0.0016
sum in quad.		**0.0301**	**0.0099**	**0.0080**	**0.0105**	**0.0102**

		$[11.0, 12.5]$	$[15.0, 17.0]$	$[17.0, 19.0]$	$[1.1, 6.0]$	$[15.0, 19.0]$
ϵ/ϵ model	7.3.5	0.0034	0.0091	0.0086	0.0024	0.0082
kaon mismodelling	7.3.3	0.0079	0.0153	0.0110	0.0028	0.0115
pion mismodelling	7.3.3	0.0067	0.0082	0.0039	0.0051	0.0031
re-sampling	7.3.8	0.0066	0.0065	0.0063	0.0049	0.0061
$\Lambda_b^0 \to pK^-\mu^+\mu^-$	7.3.6	0.0003	0.0001	0.0000	0.0009	0.0003
$B_s^0 \to \phi(1020)\mu^+\mu^-$	7.3.6	0.0005	0.0006	0.0001	0.0010	0.0005
F_{S} uncorrelated	–	0.0024	0.0048	0.0071	0.0015	0.0060
sum in quad.		**0.0130**	**0.0212**	**0.0173**	**0.0082**	**0.0168**

the largest effect observed is taken as the size of the systematic uncertainty. The systematic uncertainty obtained is combined with other uncertainties by summation in quadrature. Since it is difficult to assess whether this method of obtaining the uncertainty returns the width of the normally distributed variation, the uncertainty is also combined assuming a uniformly distributed variation as a test. The two combinations are indistinguishable for all of the values of the differential branching fraction and $F_{\mathrm{S}}|_{644}^{1200}$ in all q^2 bins and for $F_{\mathrm{S}}|_{796}^{996}$ at high q^2. For $F_{\mathrm{S}}|_{796}^{996}$ measured in the low-q^2 bins, the test indicates a possible slight over-estimation of the total systematic uncertainty which is ignored.

The effect of using the LASS parameters given by Eqns 6.22 and 6.23 is assessed by the alternative parameters given by Eqns 6.24 and 6.25. Since these LASS parameters are correlated with eachother the combinations are also tested. This choice contributes the dominant systematic effect at low q^2 at around 0.6% for $F_{\mathrm{S}}|_{796}^{996}$ and in the range $1-2\%$ for $F_{\mathrm{S}}|_{644}^{1200}$. The alternative Isobar model is considered with factors $(q_0/m_{K\pi})^{L_{K\pi}}$ and $(q_0/q)^{L_{K\pi}}$. The latter contributes at higher q^2, whereas the former has a smaller effect than both other alternatives.

7.3.2 Meson radius parameter

Another modelling choice was to fix the meson radius parameter, d, as described in Section 6.2.2. This parameter is present in the P-wave lineshape, described by the function $f_{892}(m_{K\pi})$ (Eqn. 6.15). The size of the effect is assessed by generating a pseudodata sample using the alternative values of $d = 3.0, 5.0\,[\mathrm{GeV}/c]^{-1}$ and refitting. The effect is small compared to the effect of the S-wave shape choice, described above, and is given in the third row of Tables 7.3 and 7.4. As with the S-wave modelling choice, there is negligible effect on the yields and the uncertainty is therefore only assigned to F_{S} and propagated to the (differential) branching fraction.

7.3.3 Residual differences between data and simulation due to the presence of an S-wave component in the control channel data

A systematic effect arises due to the presence of a genuine S-wave component in the control channel data from which corrections to the simulation were derived. This has an effect on the event weight corrections described in Section 5.1.2. The simulated control channel data is pure P-wave events, whereas the sFit weighted control channel data contains both P- and S-wave events. This will affect the kinematic properties of the hadrons.

The size of this effect is assessed by reweighting the hadron momentum distributions in simulation to reproduce the sFit weighted $B^0 \to J/\psi\, K^{*0}$ data. In order to minimise the effect of genuine S-wave, which is not present in the simulation, the weights are derived in a very narrow window of $\pm 20\,\mathrm{MeV}/c^2$ around the $K_1^*(892)^0$. The reweighting is performed in momentum and p_{T} of each hadron. The 2D distributions of data and simulation, and the correction weight are shown in Fig 7.11 for the pion momenta. A similar plot for the kaon momenta is given in Fig. C.6 in Appendix C.3. The resulting projections of momentum before and after the correction are also included in Appendix C.3. The common strategy of large pseudodata samples is employed: the samples are generated with an acceptance function (derived from the reweighted simulation) and refit with the nominal model and the same model with the new acceptance function. The resulting differences observed in F_{S} are given in the eighth and ninth rows of Tables 7.3 and 7.4. In general, this effect is below level of systematic uncertainty assigned to F_{S} due to the S-wave modelling uncertainty. For the differential branching fraction, these weights are also applied to the reco. sample which is used to form the ratio of efficiencies. The uncertainty on the branching fraction is propagated from the correlated variations observed in the ratio of efficiencies and the value of $F_{\mathrm{S}}|_{644}^{1200}$ for each q^2 bin. This is given in the second and third rows of Table 7.5. The possible correlation between the two reweightings is ignored when combining to obtain the total systematic uncertainty. The size of the correlation is estimated by applying both weights and repeating the above procedure. The values of the correlations are below the level that the effect the total systematic uncertainty.

7.3.4 Statistical uncertainty of the simulated data sample used to extract efficiency function

Section 5.3 described the procedure to extract a functional description of the angular efficiency from simulated data. This procedure is affected by statistical fluctuations in the sample of simulation from which the function is extracted. In order to assess the size of this effect, the covariance matrix, formed of the variances between estimators, is calculated following Eqn. 5.10. The covariance matrix is used to draw fluctuated sets of coefficients corresponding to a fluctuated efficiency function. These fluctuated sets of coefficients are treated as alternatives to the nominal efficiency function. Each fluctuated efficiency function is used, with the nominal model, to generate a large sample of pseudodata. The pseudodata sample from a fluctuated efficiency function is refit with the fluctuated model, and with the nominal model (with the nominal efficiency function). This procedure is repeated for 100 fluctuations. The resulting systematic bias is found to be small relative to the effect of modelling; except at high q^2, where it is of a comparable size. This effect contributes less than 0.1% uncertainty on F_{S} in either $m_{K\pi}$ range, the exact values are presented in the second row of Tables 7.3 and 7.4. No significant effect is observed on the yields in these pseudoexperiments therefore only the uncertainty on $F_{\mathrm{S}}|_{644}^{1200}$ is propagated to the branching fraction.

Figure 7.11: Extra correction weights (bottom centre) are derived for disagreement between the pion momenta in data (left) and simulation (right) for sFit weighted $B^0 \to J/\psi\, K^{*0}$ (shown on a normalised z-scale) with a narrow window around the $K_1^*(892)^0$.

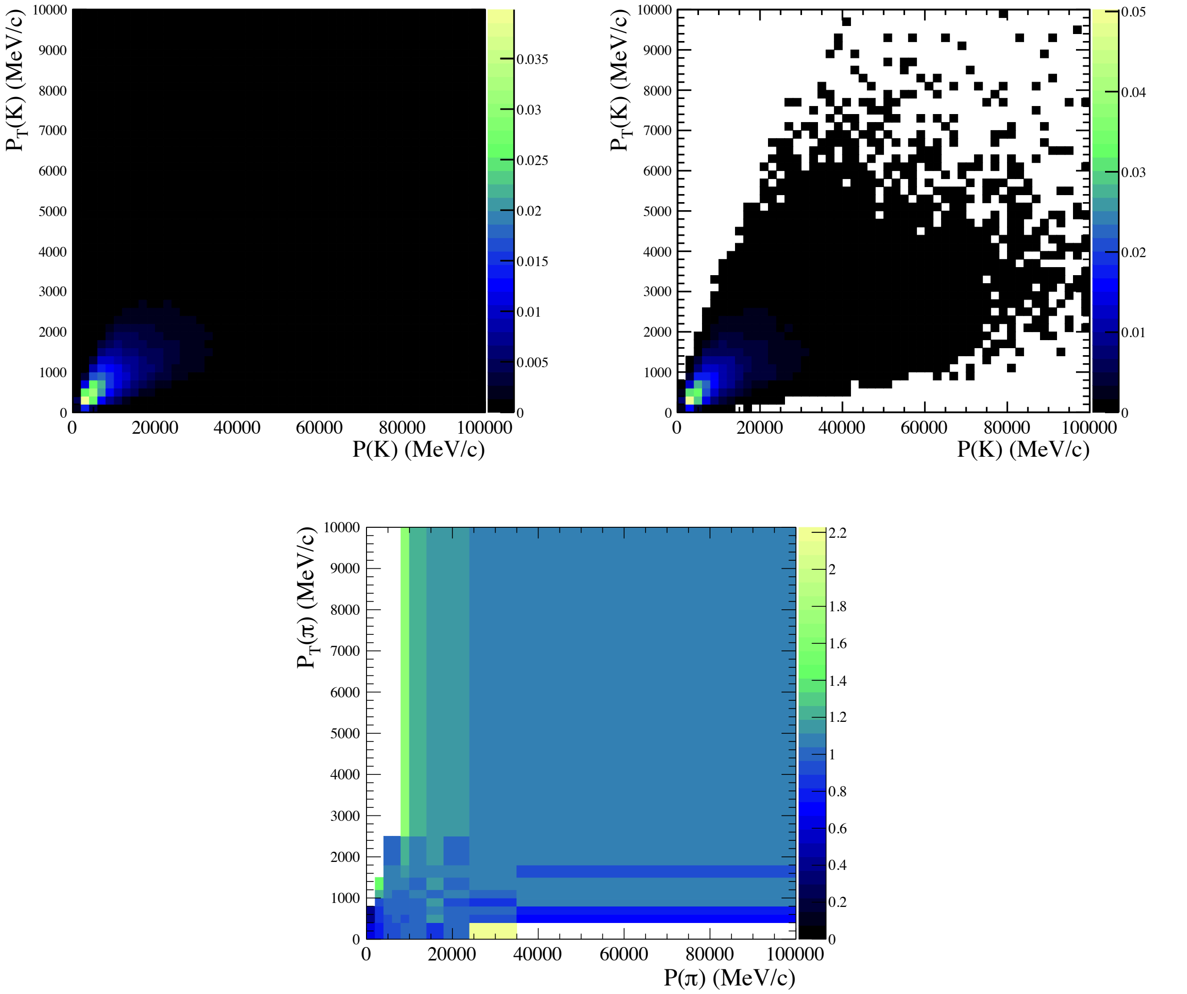
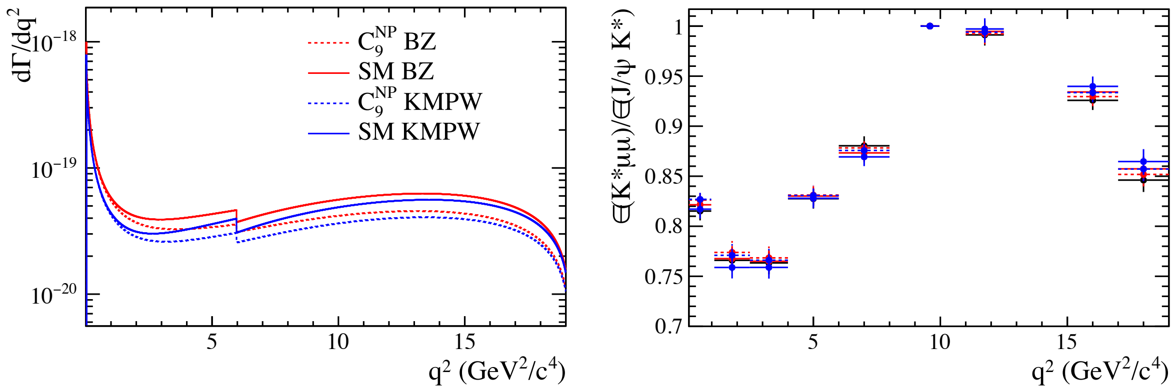


Figure 7.12: Different model predictions for the (unnormalised) differential rate, $\mathrm{d}\Gamma/\mathrm{d}q^2$ (left). The form factors are taken from BZ [24] and KMPW [101] (according to the authors' initials), amplitudes are found with the EOS software package [100] for two models: the SM and a model where $C_9^{\mathrm{NP}} = -1.5$. The functions are used to estimate the systematic uncertainty in the binned ratio of efficiencies, shown in the right-hand plot, and described in the text. For the purposes of estimating this systematic uncertainty: only the shape information in the differential rate is used. The left-hand figure is shown unnormalised in order to demonstrate that a modified C_9^{NP} predicts a somewhat lower rate which is consistent with the picture presented in Section 2.4.



7.3.5 Model dependence of the ratio of efficiencies and veto correction factor

The ratio of efficiencies, and the veto correction factor, f_{q^2}, (introduced in Eqn. 6.39) are taken from simulation which is integrated over the angles in order to provide the q^2 spectrum. As such, there is a model-dependent assumption that the angular distribution of the data agrees with the SM angular distribution which is simulated. In order to test the possible effect of this model dependence, the simulation samples are weighted to reproduce alternative models. Weights are derived from the functional forms of alternative model descriptions of the differential rate, $\mathrm{d}\Gamma/\mathrm{d}q^2$. The functions, shown to the left of Fig. 7.12, are extracted using the EOS open source software package [100]. Two sets of form factors are used from Refs. [24, 101], in two model scenarios: the SM and a realistic NP scenario where $C_9^{\mathrm{NP}} = -1.5$. The original gen. sample is used to derive four sets of weights which reproduce each of these four curves. The gen. and reco. samples are both weighted and the ratio of efficiencies, shown to the right of Fig. 7.12, is recomputed in each case. For each q^2 bin, the largest difference between each of the weighted values and the nominal is taken as the systematic uncertainty due to the model dependence of integrating the simulation in this way.

The veto correction factor is also recomputed and the largest variation is ± 0.01 *cf.* the uncertainty due to the statistics of the gen. sample, ± 0.0002. This is treated as correlated to the uncertainty assigned to the ratio of efficiencies when propagated to the total branching fraction.

7.3.6 Remaining contribution from specific background processes

Several sources of specific background processes are listed in Table 4.3. Whilst all contamination is $< 2\%$, it is possible that the dominant background processes, $\Lambda_b^0 \to pK^-\mu^+\mu^-$ (which can contribute as described in Section 4.2.2) and $B_s^0 \to \phi(1020)\mu^+\mu^-$ (described in Section 4.2.3) may be present in the q^2 spectrum in a way that conspires to have an effect in a local region of q^2. In the case of $\Lambda_b^0 \to pK^-\mu^+\mu^-$, the q^2 spectrum is not known (and the simulated sample therefore relies on a phase-space model). The effect of a contribution from either of these processes is assessed using data.

The strategy employed is: remove the vetoes that are normally employed to reject the process in question and remove the BDT requirement, employ a selection to remove signal $B^0 \to K^+\pi^-\mu^+\mu^-$ and other background events, take the distribution of these events in data as 'template' shapes, and make an assessment of the contribution using these templates. The template shapes are implemented as 3D histograms (of the $m_{K\pi\mu\mu}$, $m_{K\pi}$, and $\cos\theta_K$ dimensions) from which a random samples of pseudodata background events are drawn. The effect is quantified, as with systematics due to modelling effects, with large samples of pseudodata generated from the nominal model except that it is the data which is given an alternative, rather than the model. For every high statistics pseudodata sample, the sample is refit with the model, then a sample of the specific background component is generated at the proportion according to Table 4.3 and the proportional size of that bin in the q^2 spectrum. These background events are added to the original pseudodata sample, and this pseudodata-plus-background sample is refit. The mean of the differences in the parameters of interest for each sample, is taken as the systematic variation.

Candidate $\Lambda_b^0 \to pK^-\mu^+\mu^-$ events are selected if

$$(5575 < m_{(\pi\to p)K\mu\mu} < 5665)\,\mathrm{MeV}/c^2, \tag{7.5}$$

and

$$\mathrm{DLL}_{p\pi}(\pi) > 20. \tag{7.6}$$

The doubly misidentified candidates, where the proton is misidentified as a "kaon" track and where the kaon is also misidentified as the "pion" track, are not used. Candidate $B_s^0 \to \phi(1020)\mu^+\mu^-$ events are selected if

$$(5317 < m_{(\pi\to K)K\mu\mu} < 5417)\,\mathrm{MeV}/c^2, \tag{7.7}$$

$$(1010 < m_{(\pi\to K)K} < 1030)\,\mathrm{MeV}/c^2, \tag{7.8}$$

and

$$\mathrm{DLL}_{K\pi}(\pi) > 0. \tag{7.9}$$

It is important that the above criteria remove the genuine signal $B^0 \to K^+\pi^-\mu^+\mu^-$ component. This is necessary, since injecting extra signal into the large pseudodata samples

Figure 7.13: The q^2 spectrum for all $\Lambda_b^0 \to pK^-\mu^+\mu^-$ candidates selected in data (left) and the Λ_b^0 hypothesis four-body invariant mass ($m_{(\pi\to p)K\mu\mu}$) spectrum for $\Lambda_b^0 \to J/\psi\, pK^-$ candidates (right). Away from the J/ψ, there are only few candidates for each q^2 bin.

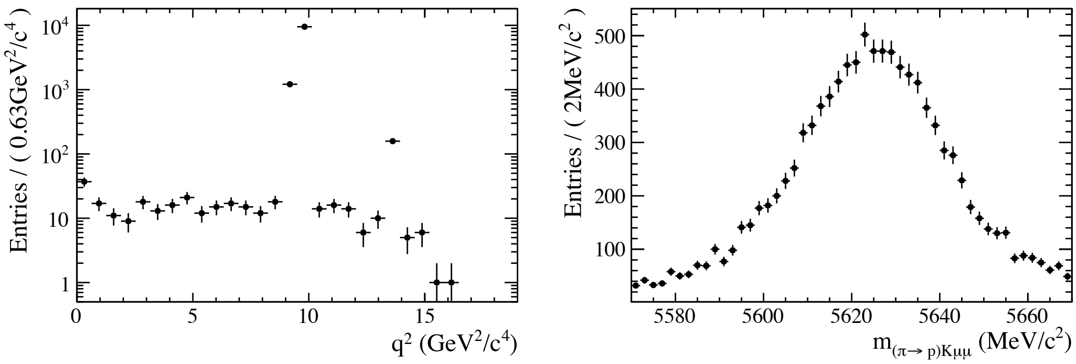
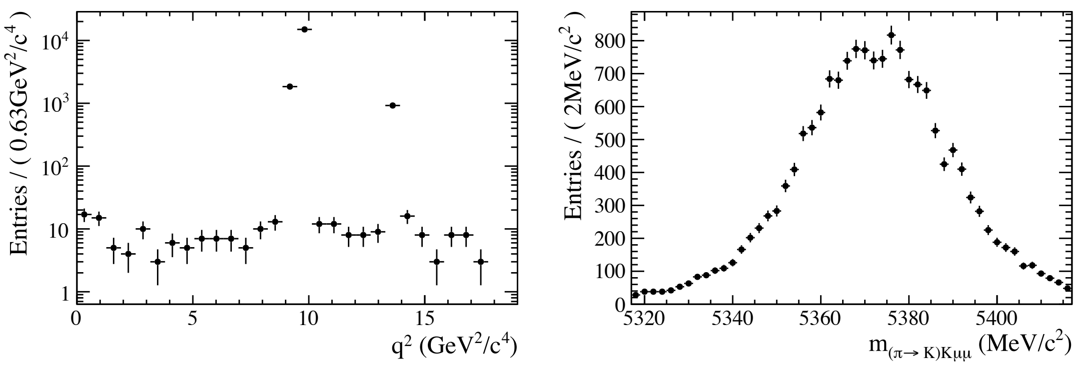


Figure 7.14: The q^2 spectrum for all $B_s^0 \to \phi(1020)\mu^+\mu^-$ candidates selected in data (left) and the B_s^0 hypothesis four-body invariant mass ($m_{(\pi\to K)K\mu\mu}$) spectrum for $B_s^0 \to J/\psi\,\phi(1020)$ candidates (right). Away from the J/ψ, there are only few candidates for each q^2 bin.



under the guise of 'background' will result in an overestimate for the uncertainty on the fitted signal yield and therefore the branching fraction.

For the q^2 bins given in Table 6.3, there are few candidates of either specific process per q^2 bin. Rather than use template shapes from each q^2 bin, where the samples have low statistics, the resonant J/ψ mode for each decay, $\Lambda_b^0 \to J/\psi\, pK^-$ and $B_s^0 \to J/\psi\,\phi(1020)$, are used. The J/ψ modes are selected by the addition of the window $q^2 \in [9.22, 9.96]\,\mathrm{GeV}^2/c^4$. The proportion of background for each q^2 bin is scaled to the proportion of data candidates in that bin (using the q^2 spectra of the data candidates) and relative to the 1.2% (0.75%) of signal for the Λ_b^0 (B_s^0) decay given in Table 4.3. Figure 7.13 (7.14) shows the Λ_b^0 (B_s^0) candidates selected by the requirements in Eqns 7.5 and 7.6 (7.7–7.9) and the q^2 spectrum for all candidates.

The effect observed on F_S is rather small, however it is given in the sixth and seventh rows of Tables 7.3 and 7.4. This effect observed on the yields is also small. The uncertainty on the signal yield and $F_\mathrm{S}|_{644}^{1200}$ is propagated to the branching fraction accounting for the correlation and given in the fifth and sixth rows of Table 7.5.

7.3.7 Background modelling

In Eqn. 6.31, an alternative model for the background in $m_{K\pi}$ is given. This model is used to assess the systematic bias which arises from the choice of the simple background model given in Eqn 6.29. Realistic values for the parameters $\{a,\, b,\, c\}$ are taken from fits of the $m_{K\pi}$ dimension in the background sideband in $m_{K\pi\mu\mu}$. These values are then used to generate a large pseudodata sample of the full 3D model, but with $B(m_{K\pi})$ replaced by $B_{\text{alt.}}(m_{K\pi})$. As with other modelling effects the resulting difference observed in the parameters of interest is taken as the systematic uncertainty. The resulting uncertainties are shown the fourth row of Tables 7.3 and 7.4. Very little effect is observed in the yields and therefore the uncertainty is assigned to F_{S} only and propagated to the branching fraction as uncorrelated.

For $\cos\theta_K$, the alternative model used is to extend the quadratic function to include a cubic term. Reasonable values for this higher order model are similarly taken from fits to the background sideband region of $m_{K\pi\mu\mu}$. Large pseudodata samples are fit with this alternative model, and with the nominal model, as above. In the worst q^2 bin, which is $q^2 \in [1.1, 2.5]\,\text{GeV}^2/c^4$, the observed variation is an order of magnitude smaller than the smallest of the systematic uncertainties; elsewhere the variation is smaller still. Therefore no systematic uncertainty is assigned.

7.3.8 Re-sampling

The re-sampling procedure for correcting the DLL variables in simulation is described in Section 5.1.1. This procedure only preserves the correlations in simulation, for the η, N_{tracks}, and p_{T} variables. The size of the systematic uncertainty due to breaking these correlations is estimated by deriving an acceptance correction where the selection is applied with the original (‘*un-re-sampled*’) DLL variables and uncorrected BDT. That is, the use of the red distributions in Fig. 5.1. This alternative efficiency function is treated as an alternative similarly to the procedure described in Section 7.3.4, above. The alternative efficiency function is used to generate a large sample of pseudodata. The pseudodata sample from the alternative efficiency function is refit with the fluctuated model, and with the nominal model (with the nominal efficiency function). The resulting systematic bias is generally found small relative to the effect of modelling.

7.3.9 Nuisance parameters

The effect of fixing the parameters $\{S_3,\, A_{\text{FB}},\, S_9\}$ is assessed by performing the fit with them set to zero, and set to the value $\pm 1\sigma$ of the quoted uncertainty, in whichever direction is furthest from zero in Ref. [29]. Note that the latter is rather aggressive, but results in a negligible difference in the fitted values of F_{S}. Fitting with the terms fixed to zero also has very little effect except in the bin $q^2 \in [1.1, 2.5]\,\text{GeV}^2/c^4$where the effect is small but should not be neglected. All values are therefore included in the fifth row of Tables 7.3 and 7.4 for completeness. As with many other effects, the variation is only observed in F_{S} and therefore propagated directly to the branching fraction as a component of the seventh row of Table 7.5.

7.3.10 Other subdominant effects and checks

Various other effects could introduce a systematic bias. The effects discussed in this section are considered and no systematic uncertainty is assigned.

The mismodelling of the tracking efficiency in simulation is often considered as a source of systematic uncertainty, however this effect has been shown to contribute at less than the percent level in previous work [1] and is therefore not considered here. Similarly, the effect of the invariant mass resolution which is similarly neglected.

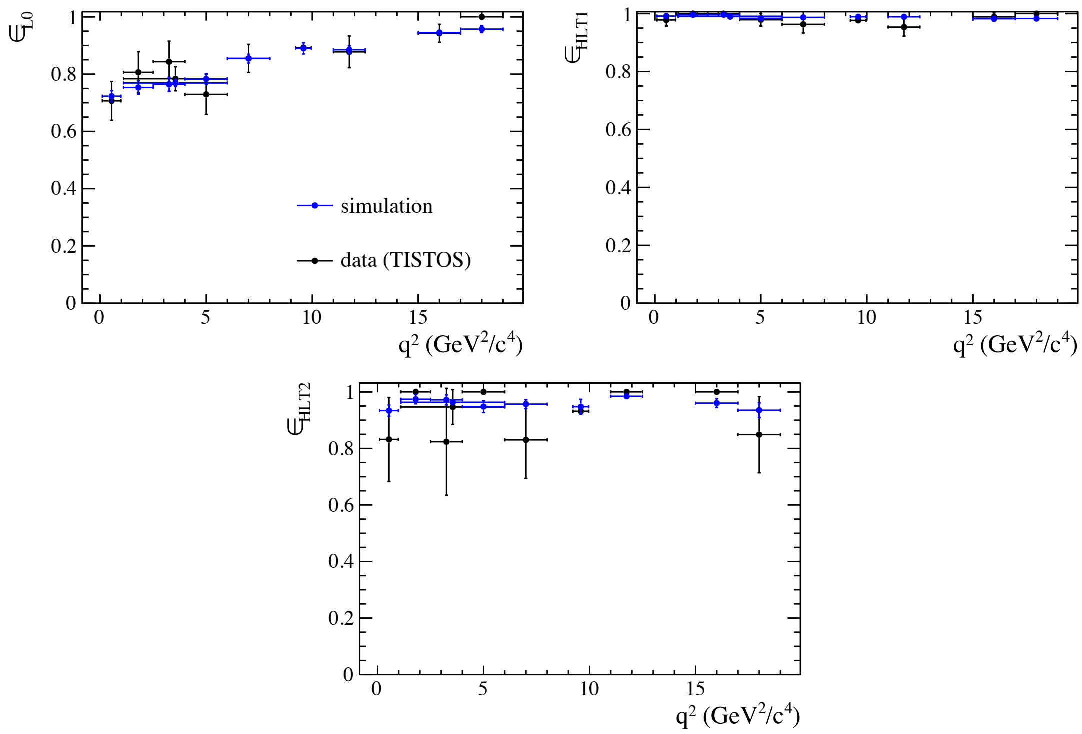
Agreement between data and simulation due to the modelling of the trigger

Any systematic uncertainty related to the mismodelling of the trigger efficiency in the simulation is assessed by comparing the trigger efficiencies in simulation and data. In data, it is possible to extract the trigger efficiency, with relatively poor precision, using events that were trigger independently of signal (TIS). The data used is selected with the full selection chain, as described in Chapter 4, with the exception of the trigger requirements. The background component in the data is removed using the sFit technique, as described in Section 4.3.3. Using the independent sample of events that were selected but were TIS, the efficiency is given by the ratio of the number of events that were both TIS and trigger on signal (TOS) to the total number of events that were TOS,

$$\epsilon_{\text{trigger}} = \frac{n_{\text{TIS\&TOS}}}{n_{\text{TIS}}}. \tag{7.10}$$

The efficiency distributions are shown for data and simulation in Fig 7.15. These distributions are compatible within the statistical uncertainty, therefore no systematic uncertainty is assigned. In some cases in the data, there are no TIS events in the sample therefore the bin reports an efficiency of unity, these bins should be ignored.

Figure 7.15: The efficiency of, the L0 trigger (top left), HLT1 (top right), and HLT2 (bottom centre) in data and simulation. The trigger efficiency in data is extracted using the method described in the text. The distributions are compatible to within uncertainties and therefore no systematic uncertainty is assigned.



Chapter 8

Conclusions and outlook

This thesis presents the world's most precise measurement of the differential and total branching fraction of $B^0 \to K_1^*(892)^0\mu^+\mu^-$. The analysis makes the first measurement of the S-wave pollution using a partial angular analysis and modelling of the $K\pi$ invariant mass spectrum. This procedure allows the branching fraction of the pure P-wave component to be measured. Previous analyses have only ever extracted the sum of the P- and S-wave components and compared to the theoretical predictions for the P-wave.

The measured differential branching fraction for all q^2 bins is in shown Fig. 7.1 and given in Table 7.1. The results are below SM predictions although consistent within uncertainties. In the theoretically favoured region $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$, the differential branching fraction is found to be

$$\frac{\mathrm{d}\mathcal{B}\,[B^0 \to K_1^*(892)^0\mu^+\mu^-]}{\mathrm{d}q^2} = (0.402^{+0.020}_{-0.019} \pm 0.008 \pm 0.027) \times 10^{-7}, \tag{8.1}$$

where the first uncertainty is statistical, the second is systematic and the third due to the precision on the branching fraction of the normalisation channel. This measurement can be compared to the SM prediction of $(0.49 \pm 0.08) \times 10^{-7}$ [23].

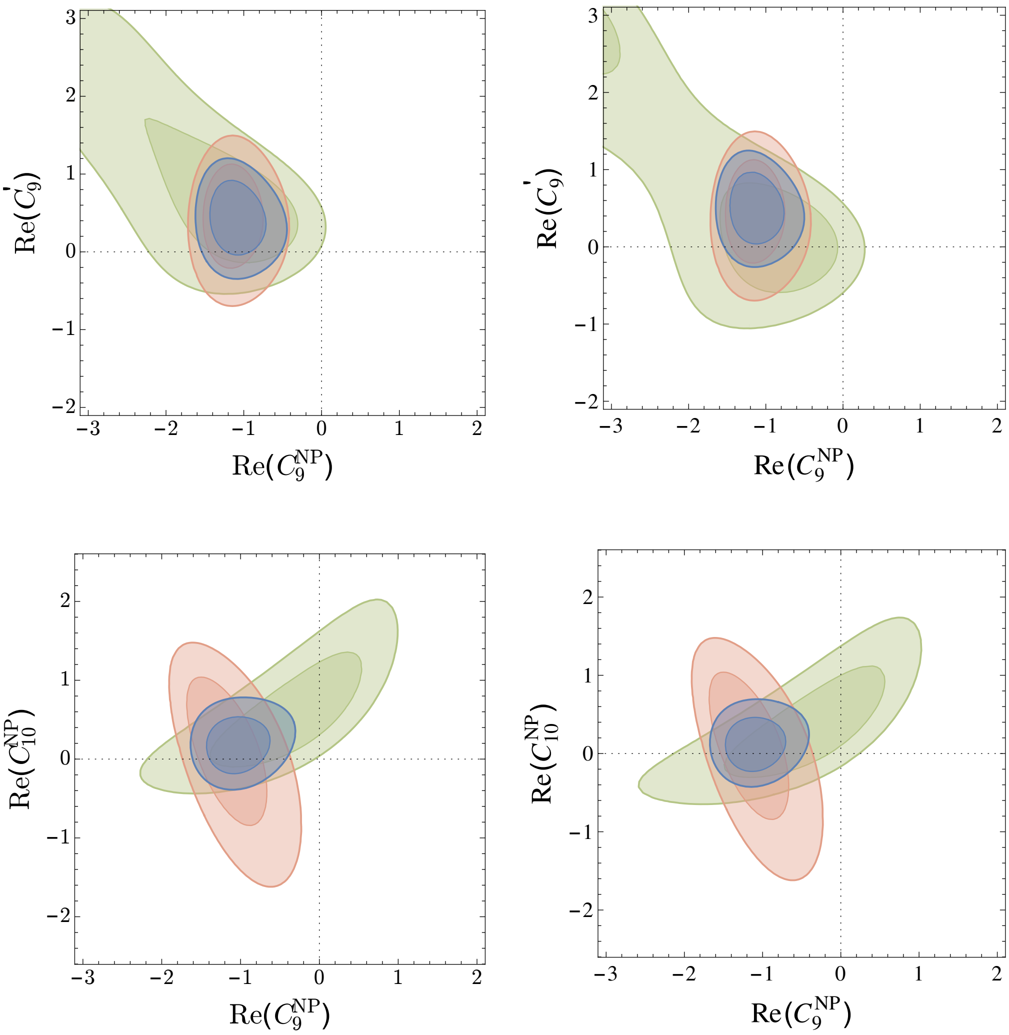
Although the differential branching fraction for $q^2 \in [1.1, 6]\,\mathrm{GeV}^2/c^4$ is within 1σ of the prediction, the measurement exhibits the same trend as measurements of other $b \to s\ell^+\ell^-$ processes, such as the branching fractions of $B_s^0 \to \phi(1020)\mu^+\mu^-$ [32], $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K_S^0\mu^+\mu^-$, and $B^+ \to K^{*+}\mu^+\mu^-$ [28] which are all lower than SM predictions. This is intriguing as it could be explained by a NP contribution to the Wilson coefficient $C_9^{(\prime)}$ which describes a left- (right-) handed vector coupling for the $b \to s\ell^+\ell^-$ transition. The NP contribution makes C_9 smaller than predicted in the SM. This modified C_9 could be explained by, for example, a Z' with a mass of $m_{Z'} \sim 7\,\mathrm{TeV}/c^2$ which is not excluded by direct searches [35]. A low value of C_9 is consistent with the 3.7σ anomaly in a quantity derived from the angular distribution of $B^0 \to K_1^*(892)^0\mu^+\mu^-$ decays [27, 29], and the 2.6σ deviation from unity in the ratio of branching fractions of $B^+ \to K^+\mu^+\mu^-$ to $B^+ \to K^+e^+e^-$ decays, R_K, (see Eqn 2.93) [43]. Whilst a consistent NP explanation of the various anomalies is viable, the SM hypothesis is by no means excluded [47, 48].

Figure 8.1 shows the result of a global fit to the Wilson Coefficients from Ref. [29] and an updated analysis with measurements of the differential branching fraction of $B_s^0 \to \phi(1020)\mu^+\mu^-$ decays [32] and the measurements presented in this thesis. The tension between the best fit point and the SM, since publication of Ref. [48], is somewhat reduced. This is due, in part to a slightly smaller tension in the differential branching fraction of $B^0 \to K_1^*(892)^0\mu^+\mu^-$ than was reported in the previous measurement [26] which is superceded by the analysis of this thesis. Whilst the differential branching fraction of $B_s^0 \to \phi(1020)\mu^+\mu^-$ is in tension for the region $q^2 \in [1, 6]\,\mathrm{GeV}^2/c^4$, the overall agreement of the angular observables for this process with the SM predictions is good [32]. This also contributes to the slightly reduced tension.

The LHC is currently running, and the LHCb collaboration will update these measurements with several times the dataset size. However, for the angular observables in $B^0 \to K_1^*(892)^0\mu^+\mu^-$ decays, and the branching fractions, there is still considerable debate in the theory community as to whether the theoretical uncertainties are as small as claimed. In particular, the size of charm-quark-loop corrections to the predictions for the Wilson coefficients (and therefore to SM predictions of observables) is still the subject of intense debate [39]. A future measurement of the relative phase between the J/ψ and $\psi(2S)$ resonances may help in testing the understanding of this source of theoretical uncertainty [39].

Conversely, the lepton universality observable, R_K, should not be sensitive to charm-loop effects. Future measurements to test lepton universality such as measurements of R_ϕ and R_{K^*} (defined analogously to Eqn 2.93) are planned at LHCb, and may help to clarify the situation. Given lepton non-universality, lepton flavour violation becomes well motivated [45]. Searches for lepton flavour violating processes of the type $B \to K\ell\ell'$ are also under study.

Figure 8.1: Global fits to the Wilson coefficients shown as planes of $\mathfrak{Re}\left(C_9^{\mathrm{NP}}\right) - \mathfrak{Re}\left(C_9^{\prime\,\mathrm{NP}}\right)$ (upper) and $\mathfrak{Re}\left(C_9^{\mathrm{NP}}\right) - \mathfrak{Re}\left(C_{10}^{\mathrm{NP}}\right)$ (lower) performed before (left) and after (right) the inclusion of the results presented in this thesis and the branching fraction of $B_s^0 \to \phi(1020)\mu^+\mu^-$ from Ref. [32]. The darker and lighter contours correspond to 1σ and 2σ best fit regions respectively. The blue shaded area shows the global best fit point. The red (green) area shows the individual constraint from angular observables (branching fractions).



References

[1] C. Parkinson, *The angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay at LHCb*, Ph.D. thesis, Imperial College, London, June, 2013, CERN-THESIS-2013-167. [Cited on pages 4, 80, 87, 110, and 154.]

[2] Y. Grossman and F. Tanedo, *Just A Taste: Lectures on Flavour Physics*, `http://www.lepp.cornell.edu/~pt267/files/notes/FlavorNotes.pdf`. Accessed: 2014-09-05. [Cited on page 21.]

[3] M. Schumann, *Dark Matter 2014*, EPJ Web Conf. **96** (2015) 01027, `arXiv:1501.01200`. [Cited on page 21.]

[4] **Particle Data Group**, K. A. Olive *et al.*, *Review of particle physics*, Chin. Phys. **C38** (2014) 090001. [Cited on pages 17, 21, 34, 42, 76, 78, 113, 114, 116, and 118.]

[5] **ATLAS collaboration**, G. Aad *et al.*, *Search for squarks and gluinos with the ATLAS detector in final states with jets and missing transverse momentum using $\sqrt{s} = 8$ TeV proton–proton collision data*, JHEP **09** (2014) 176, `arXiv:1405.7875`, latest results available at `http://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults`; **CMS collaboration**, S. Chatrchyan *et al.*, *Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s}$= 8 TeV*, JHEP **06** (2014) 055, `arXiv:1402.4770`, latest results available at `http://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS`. [Cited on page 22.]

[6] C.-D. Lü and W. Wang, *Analysis of $B \to K_J^*(\to K\pi)\mu^+\mu^-$ in the higher kaon resonance region*, Phys. Rev. **D85** (2012) 034014, `arXiv:1111.1513`. [Cited on pages 23, 48, and 53.]

[7] A. V. Manohar and M. B. Wise, *Heavy quark physics*, Cambridge University Press, Cambridge, United Kindom, 2000. Chapter 1. isbn: 9780521037570. [Cited on page 25.]

[8] T. Mannel, *Effective Field Theories in Flavor Physics*, Springer-Verlag, Berlin, Germany, 2004. Chapters 2.2, 3.1, 4, and 5.1.4, doi: 10.1007/b62268. isbn: 9783540219316. [Cited on pages 25, 36, and 38.]

[9] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Westview Press, Colorado, United States, 1995. Chapters 17.1, 17.2, and 20. isbn: 9780201503975. [Cited on page 25.]

[10] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8** (1964) 214. [Cited on page 25.]

[11] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579; S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264; A. Salam, *Weak and Electromagnetic Interactions*, Proceedings, The Nobel Symposium, Stockholm, Sweden **C680519** (1968) 367. [Cited on page 25.]

[12] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321; P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508; G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585. [Cited on page 26.]

[13] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531; S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) 1285; M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652. [Cited on page 33.]

[14] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945. [Cited on page 34.]

[15] **CKMfitter collaboration**, J. Charles *et al.*, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. **C41** (2005) 1, `arXiv:hep-ph/0406184`, updated results and plots available at `http://ckmfitter.in2p3.fr`; **UTfit collaboration**, M. Bona *et al.*, *The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters from Lattice QCD: A Reappraisal after the Measurements of Δm_s and $\mathcal{B}[B \to \tau\nu_\tau]$*, JHEP **0610** (2006) 081, `arXiv:hep-ph/0606167`, updated results and plots available at `http://www.utfit.org`. [Cited on page 34.]

[16] W. Altmannshofer *et al.*, *Symmetries and Asymmetries of $B \to K^*\mu^+\mu^-$ Decays in the Standard Model and Beyond*, JHEP **0901** (2009) 019, `arXiv:0811.1214`. [Cited on pages 38, 39, 40, 41, 48, 49, 51, and 52.]

[17] F. Kruger and J. Matias, *Probing new physics via the transverse amplitudes of $B^0 \to K^{*0}(\to K^-\pi^+)\ell^+\ell^-$ at large recoil*, Phys. Rev. **D71** (2005) 094009, `arXiv:hep-ph/0502060`. [Cited on pages 38 and 49.]

[18] D. van Dyk, *The Decays $\bar{B} \to \bar{K}^{(*)}\ell^+\ell^-$ at Low Recoil and their Constraints on New Physics*, Ph.D. thesis, Technische Universität Dortmund, July, 2012, `http://eldorado.tu-dortmund.de/bitstream/2003/29514/1/Dissertation.pdf`. [Cited on pages 38, 39, and 49.]

[19] **CMS and LHCb collaborations**, V. Khachatryan *et al.*, *Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature **522** (2015) 68, `arXiv:1411.4413`. [Cited on pages 39 and 44.]

[20] M. Beneke, T. Feldmann, and D. Seidel, *Exclusive radiative and electroweak $b \to d$ and $b \to s$ penguin decays at NLO*, Eur. Phys. J. **C41** (2005) 173, `arXiv:hep-ph/0412400`. [Cited on page 40.]

[21] J. Matias, F. Mescia, M. Ramon, and J. Virto, *Complete Anatomy of $\bar{B}_d \to \bar{K}^{*0}(\to K\pi)\ell^+\ell^-$ and its angular distribution*, JHEP **1204** (2012) 104, `arXiv:1202.4266`. [Cited on pages 40, 41, 48, 49, and 51.]

[22] T. Blake, T. Gershon, and G. Hiller, *Rare b hadron decays at the LHC*, `arXiv:1501.03309`. [Cited on pages 15, 40, and 41.]

[23] A. Bharucha, D. M. Straub, and R. Zwicky, *$B \to V\ell^+\ell^-$ in the Standard Model from Light-Cone Sum Rules*, `arXiv:1503.05534`. [Cited on pages 41, 43, 133, 134, 139, 140, and 156.]

[24] P. Ball and R. Zwicky, *$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form-factors from light-cone sum rules reexamined*, Phys. Rev. **D71** (2005) 014029, `arXiv:hep-ph/0412079`. [Cited on pages 41 and 150.]

[25] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, *Lattice QCD calculation of form factors describing the rare decays $B \to K^*\ell^+\ell^-$ and $B_s \to \phi\ell^+\ell^-$*, Phys. Rev. **D89** (2014) 094501, `arXiv:1310.3722`. [Cited on pages 41, 133, 134, and 139.]

[26] **LHCb collaboration**, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$*, JHEP **08** (2013) 131, `arXiv:1304.6325`. [Cited on pages 42, 43, 53, 110, and 157.]

[27] **LHCb collaboration**, R. Aaij *et al.*, *Measurement of form-factor-independent observables in the decay $B^0 \to K^{*0}\mu^+\mu^-$*, Phys. Rev. Lett. **111** (2013) 191801, `arXiv:1308.1707`. [Cited on pages 42, 43, 53, 110, and 156.]

[28] **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`. [Cited on pages 42, 43, 110, 122, 123, and 156.]

[29] **LHCb collaboration**, R. Aaij *et al.*, *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay*, arXiv:1512.04442, submitted to JHEP. [Cited on pages 17, 42, 43, 93, 124, 128, 139, 140, 153, 156, 157, 181, 183, and 184.]

[30] D. Bečirević and A. Tayduganov, *Impact of $B \to K_0^*\ell^+\ell^-$ on the New Physics search in $B \to K^*\ell^+\ell^-$ decay*, Nucl. Phys. **B868** (2013) 368, arXiv:1207.4004. [Cited on pages 42 and 141.]

[31] S. Descotes-Genon, T. Hurth, J. Matias, and J. Virto, *Optimizing the basis of $B \to K^*\ell^+\ell^-$ observables in the full kinematic range*, JHEP **1305** (2013) 137, arXiv:1303.5794. [Cited on pages 43, 49, 52, and 53.]

[32] **LHCb collaboration**, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$*, JHEP **09** (2015) 179, arXiv:1506.08777. [Cited on pages 43, 45, 156, 157, and 158.]

[33] S. Descotes-Genon, J. Matias, and J. Virto, *Understanding the $B \to K^*\mu^+\mu^-$ Anomaly*, Phys. Rev. **D88** (2013) 074002, arXiv:1307.5683. [Cited on page 43.]

[34] W. Altmannshofer and D. M. Straub, *New physics in $B \to K^*\mu\mu$?*, Eur. Phys. J. **C73** (2013) 2646, arXiv:1308.1501. [Cited on pages 43 and 44.]

[35] R. Gauld, F. Goertz, and U. Haisch, *An explicit Z'-boson explanation of the $B \to K^*\mu^+\mu^-$ anomaly*, JHEP **1401** (2014) 069, arXiv:1310.1082. [Cited on pages 43 and 156.]

[36] S. Jäger and J. Martin Camalich, *On $B \to V\ell\ell$ at small dilepton invariant mass, power corrections, and new physics*, JHEP **1305** (2013) 043, arXiv:1212.2263. [Cited on pages 43 and 49.]

[37] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *On the impact of power corrections in the prediction of $B \to K^*\mu^+\mu^-$ observables*, JHEP **1412** (2014) 125, arXiv:1407.8526. [Cited on page 43.]

[38] S. Jäger and J. Martin Camalich, *Reassessing the discovery potential of the $B \to K^*\ell^+\ell^-$ decays in the large-recoil region: SM challenges and BSM opportunities*, arXiv:1412.3183. [Cited on page 43.]

[39] 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. [Cited on pages 43, 45, and 157.]

[40] **LHCb collaboration**, R. Aaij *et al.*, *Observation of a resonance in $B^+ \to K^+\mu^+\mu^-$ decays at low recoil*, Phys. Rev. Lett. **111** (2013) 112003, arXiv:1307.7595. [Cited on page 44.]

[41] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, *Quark flavor transitions in $L_\mu - L_\tau$ models*, Phys. Rev. **D89** (2014) 095033, arXiv:1403.1269. [Cited on page 44.]

[42] G. Hiller and F. Kruger, *More model independent analysis of $b \to s$ processes*, Phys. Rev. **D69** (2004) 074020, arXiv:hep-ph/0310219. [Cited on page 44.]

[43] **LHCb collaboration**, R. Aaij *et al.*, *Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays*, Phys. Rev. Lett. **113** (2014) 151601, arXiv:1406.6482. [Cited on pages 44 and 156.]

[44] G. Hiller and M. Schmaltz, *R_K and future $b \to s\ell\ell$ BSM opportunities*, Phys. Rev. **D90** (2014) 054014, arXiv:1408.1627. [Cited on page 44.]

[45] S. L. Glashow, D. Guadagnoli, and K. Lane, *Lepton Flavor Violation in B Decays?*, Phys. Rev. Lett. **114** (2015) 091801, arXiv:1411.0565. [Cited on pages 44 and 157.]

[46] D. Ghosh, M. Nardecchia, and S. A. Renner, *Hint of Lepton Flavour Non-Universality in B Meson Decays*, arXiv:1408.4097. [Cited on page 44.]

[47] W. Altmannshofer and D. M. Straub, *New physics in $b \to s$ transitions after LHC run 1*, arXiv:1411.3161. [Cited on pages 44 and 156.]

[48] W. Altmannshofer and D. M. Straub, *Implications of $b \to s$ measurements*, Proceedings, 50th Rencontres de Moriond on EW Interactions and Unified Theories La Thuile, Italy (2015) arXiv:1503.06199. [Cited on pages 15, 18, 44, 45, 46, 156, and 157.]

[49] U. Egede *et al.*, *New observables in the decay mode $\overline{B}^0 \to \overline{K}^{*0}\ell^+\ell^-$*, JHEP **0811** (2008) 032, arXiv:0807.2589. [Cited on pages 49 and 52.]

[50] U. Egede *et al.*, *The $\bar{B}_d \to \bar{K}^{*0}\mu^+\mu^-$ decay*, Nucl. Phys. Proc. Suppl. **209** (2010) 143, arXiv:1011.0920. [Cited on pages 49 and 52.]

[51] T. Blake, U. Egede, and A. Shires, *The effect of S-wave interference on the $B^0 \to K^{*0}\ell^+\ell^-$ angular observables*, JHEP **1303** (2013) 027, arXiv:1210.5279. [Cited on page 53.]

[52] **LHCb collaboration**, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005. [Cited on pages 15, 57, 61, 62, 64, and 65.]

[53] **LHCb collaboration**, *$b\bar{b}$ Production Angle Plots webpage*, http://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html. Accessed: 2014-11-13. [Cited on pages 15 and 58.]

[54] **LHCb collaboration**, *Detector Geometry webpage*, http://lhcb-geom.web.cern.ch/lhcb-geom/. Accessed: 2014-09-23. [Cited on pages 15 and 59.]

[55] **LHCb collaboration**, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352. [Cited on pages 15, 59, and 61.]

[56] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808. [Cited on page 61.]

[57] J. Van Tilburg, *Track simulation and reconstruction in LHCb*, Ph.D. thesis, Vrije U. Amsterdam, Amsterdam, September, 2005, CERN-THESIS-2005-040. [Cited on pages 15 and 62.]

[58] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759. [Cited on pages 15, 63, and 64.]

[59] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249. [Cited on pages 15, 19, 65, and 66.]

[60] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055. [Cited on pages 15, 19, 66, 67, 68, 69, 70, and 71.]

[61] 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. [Cited on pages 66, 67, 68, 69, and 70.]

[62] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, arXiv:1211.1346. [Cited on page 67.]

[63] R. E. Kalman, *A new approach to linear filtering and prediction problems*, J. Fluids Eng. **82** (1960) 35. [Cited on page 69.]

[64] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444. [Cited on page 69.]

[65] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984. [Cited on pages 69 and 82.]

[66] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning and an application to boosting*, Jour. Comp. and Syst. Sc. **55** (1997) 119. [Cited on pages 69, 82, and 83.]

[67] 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. [Cited on pages 69 and 70.]

[68] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047; M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023. [Cited on page 72.]

[69] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, `arXiv:hep-ph/0603175`. [Cited on page 72.]

[70] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820`. [Cited on page 72.]

[71] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152; **EvtGen working group**, J. Back, *et al.* , *Evtgen project homepage*, `http://evtgen.warwick.ac.uk/`. Accessed: 2015-06-05. [Cited on page 72.]

[72] 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`. [Cited on page 72.]

[73] **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. [Cited on page 72.]

[74] M. Needham, *Identification of Ghost Tracks using a Likelihood Method*, LHCb-2008-026. [Cited on page 74.]

[75] **LHCb collaboration**, R. Aaij *et al.*, *Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region*, Phys. Lett. **B694** (2010) 209, `arXiv:1009.2731`. [Cited on page 76.]

[76] **LHCb collaboration**, R. Aaij *et al.*, *Study of the kinematic dependences of Λ_b^0 production in pp collisions and a measurement of the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ branching fraction*, JHEP **08** (2014) 143, `arXiv:1405.6842`. [Cited on page 76.]

[77] A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, `arXiv:physics/0703039`. [Cited on page 82.]

[78] M. Pivk and F. R. Le Diberder, *sPlot: a statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, `arXiv:physics/0402083`. [Cited on page 85.]

[79] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, Ph.D. thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02. [Cited on pages 85 and 172.]

[80] S. Coquereau and M. Chrząszcz. Private communication; S. Coquereau, *Study of the $B^0 \to K^{*0}\mu^+\mu^-$ decay with the LHCb experiment: angular analysis and measurement of the ratio $R_{K^{*0}}$*, Ph.D. thesis, Université Pierre et Marie Curie, Paris, September, 2015, Submission pending, to appear. [Cited on pages 87, 90, and 92.]

[81] **LHCb collaboration**, R. Aaij *et al.*, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment*, Phys. Rev. Lett. **111** (2013) 101805, arXiv:1307.5024. [Cited on page 87.]

[82] L. Gavardi, *Search for lepton flavour violation in τ decays at the LHCb experiment*, Master's thesis, Milan Bicocca University, November, 2013. CERN-THESIS-2013-259. [Cited on pages 87 and 90.]

[83] **LHCb collaboration**, R. Aaij *et al.*, *Search for the lepton flavour violating decay $\tau^- \to \mu^-\mu^+\mu^-$*, JHEP **02** (2015) 121, arXiv:1409.8548. [Cited on page 90.]

[84] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *A Comparative study of the decays $B \to (K, K^*)\ell^+\ell^-$ in standard model and supersymmetric theories*, Phys. Rev. **D61** (2000) 074024, arXiv:hep-ph/9910221. [Cited on page 98.]

[85] F. James, *Statistical methods in experimental physics*, World Scientific, Singapore, 2006. Chapter 8.2. isbn: 9789812567956. [Cited on pages 101 and 103.]

[86] F. Beaujean, M. Chrząszcz, N. Serra, and D. van Dyk, *Extracting Angular Observables without a Likelihood and Applications to Rare Decays*, arXiv:1503.04100. [Cited on pages 101 and 181.]

[87] **BaBar Collaboration**, B. Aubert *et al.*, *Measurement of branching fractions and charge asymmetries for exclusive B decays to charmonium*, Phys. Rev. Lett. **94** (2005) 141801, arXiv:hep-ex/0412062. [Cited on pages 109, 122, and 123.]

[88] P. H. Owen, *Measurement of branching fractions, isospin asymmetries and angular observables in exclusive electroweak penguin decays*, Ph.D. thesis, Imperial College, London, July, 2014, CERN-THESIS-2014-057. [Cited on pages 110 and 122.]

[89] J. Blatt and V. Weisskopf, *Theoretical Nuclear Physics*, John Wiley & Sons, New York, 1952; F. Von Hippel and C. Quigg, *Centrifugal-barrier effects in resonance partial decay widths, shapes, and production amplitudes*, Phys. Rev. **D5** (1972) 624. [Cited on page 113.]

[90] **LHCb collaboration**, R. Aaij *et al.*, *Observation of the resonant character of the $Z(4430)^-$ state*, Phys. Rev. Lett. **112** (2014) 222002, arXiv:1404.1903. [Cited on pages 18, 113, 114, 115, 123, 129, 181, 182, 183, and 184.]

[91] **Belle collaboration**, K. Chilikin *et al.*, *Experimental constraints on the spin and parity of the $Z(4430)^+$*, Phys. Rev. **D88** (2013) 074026, arXiv:1306.4894. [Cited on page 114.]

[92] **LASS collaboration**, D. Aston *et al.*, *A Study of $K^-\pi^+$ Scattering in the Reaction $K^-p \to K^-\pi^+n$ at 11 GeV/c*, Nucl. Phys. **B296** (1988) 493. [Cited on pages 114 and 115.]

[93] **BaBar collaboration**, B. Aubert *et al.*, *Search for the $Z(4430)^-$ at BaBar*, Phys. Rev. **D79** (2009) 112001, arXiv:0811.0564. [Cited on page 115.]

[94] **BaBar collaboration**, B. Aubert *et al.*, *Evidence for $D^0 - \bar{D}^0$ Mixing*, Phys. Rev. Lett. **98** (2007) 211802, arXiv:hep-ex/0703020. [Cited on page 117.]

[95] **CMS**, V. Khachatryan *et al.*, *Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV*, arXiv:1507.08126; M. Dinardo, *Angular analysis of the decay $B^0 \to K^{*0}\mu\mu$ with CMS*, https://indico.cern.ch/event/356420/session/3/contribution/235. Slide 17 shows control channel fits. Accessed: 2015-08-04. [Cited on pages 17, 18, 129, 141, 142, 181, and 182.]

[96] **Belle collaboration**, K. Chilikin *et al.*, *Observation of a new charged charmoniumlike state in $\overline{B}^0 \to J/\psi K^-\pi^+$ decays*, Phys. Rev. **D90** (2014), no. 11 112009, arXiv:1408.6457. [Cited on pages 129, 181, and 184.]

[97] **BaBar collaboration**, J. P. Lees *et al.*, *Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays $B \to K^{(*)}\ell^+\ell^-$*, Phys. Rev. **D86** (2012) 032012, arXiv:1204.3933. [Cited on pages 17, 141, and 142.]

[98] **Belle collaboration**, J.-T. Wei *et al.*, *Measurement of the Differential Branching Fraction and Forward-Backword Asymmetry for $B \to K^{(*)}\ell^+\ell^-$*, Phys. Rev. Lett. **103** (2009) 171801, arXiv:0904.0770. [Cited on pages 17, 141, and 142.]

[99] **CDF collaboration**, T. Aaltonen *et al.*, *Measurement of the Forward-Backward Asymmetry in the $B \to K^{(*)}\mu^+\mu^-$ Decay and First Observation of the $B^0\_s \to \phi\mu^+\mu^-$ Decay*, Phys. Rev. Lett. **106** (2011) 161801, arXiv:1101.1028. [Cited on pages 17, 141, and 142.]

[100] **EOS working group**, F. Beaujean *et. al*, *EOS project homepage*, http://project.het.physik.tu-dortmund.de/eos/. Last accessed: 2015-06-02; C. Bobeth, G. Hiller, and D. van Dyk, *The Benefits of $\bar{B} \to \bar{K}^*\ell^+\ell^-$ Decays at Low Recoil*, JHEP **1007** (2010) 098, arXiv:1006.5013. [Cited on page 150.]

[101] 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 **1009** (2010) 089, arXiv:1006.4945. [Cited on page 150.]

[102] U. Egede, M. Patel, and K. A. Petridis, *Method for an unbinned measurement of the q^2 dependent decay amplitudes of $\overline{B}^0 \to \overline{K^{*0}}\mu^+\mu^-$ decays*, JHEP **06** (2015) 084, arXiv:1504.00574. [Cited on pages 18, 157, 181, 182, and 183.]

Appendices

Appendix A

Supplementary figures related to the selection

The figures shown in this appendix are supplementary to those shown in Chapter 4.

Figure A.1 shows the $\cos\theta_\ell$ and ϕ distributions for simulated events that are not excluded by the full selection. This figure is supplementary to Fig. 4.1.

Figure A.2 shows the $m_{(K\to p)(\pi\to K)\mu\mu}$ and $\mathrm{DLL}_{K\pi}(\pi)$ plane for simulated signal and phase-space $\Lambda_b^0\to\Lambda^*(1520)^0\mu^+\mu^-$ where the p is reconstructed as the "K^+" candidate and the K^- is reconstructed as the "π^-". This figure is analogous to Fig. 4.3 and depicts the veto detailed in Eqns 4.6 and 4.7.

Figure A.3 shows the relative efficiency to select simulated signal events and reject the misidentified and doubly misidentified tracks originating from the $\Lambda_b^0\to pK^-\mu^+\mu^-$ decays.

Figure A.1: Distributions of simulated events shown as stacked histograms and scaled to the number of events expected in 3 fb^{-1}. The chanels shown are: $B^0 \to K^{*0}\mu^+\mu^-$ (signal) in **cyan**; signal mode with hadron swaps in **pink**; $B_s^0 \to \phi\mu\mu$ in **green**; $B^+ \to K^+\mu^+\mu^-$ in **red**; and samples of $\Lambda_b^0 \to pK^-\mu^+\mu^-$ and $\Lambda_b^0 \to \Lambda^*(1520)^0\mu^+\mu^-$ in black.

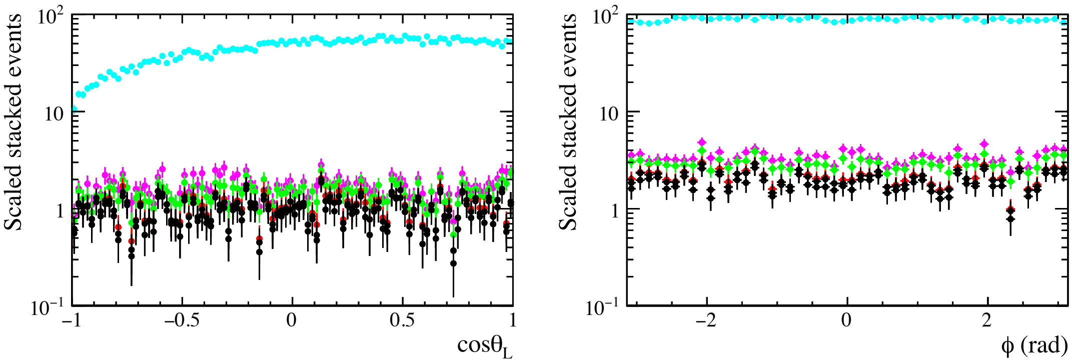


Figure A.2: Planes of Λ_b^0 candidate mass ($m_{(K\to p)(\pi\to K)\mu\mu}$) and pion candidate $\mathrm{DLL}_{K\pi}$ for simulated physics signal (left) and simulated phase-space $\Lambda_b^0 \to \Lambda^*(1520)^0\mu^+\mu^-$ events (right). The red shaded box shows the area that is excluded by the veto (Eqns 4.6 and 4.7).

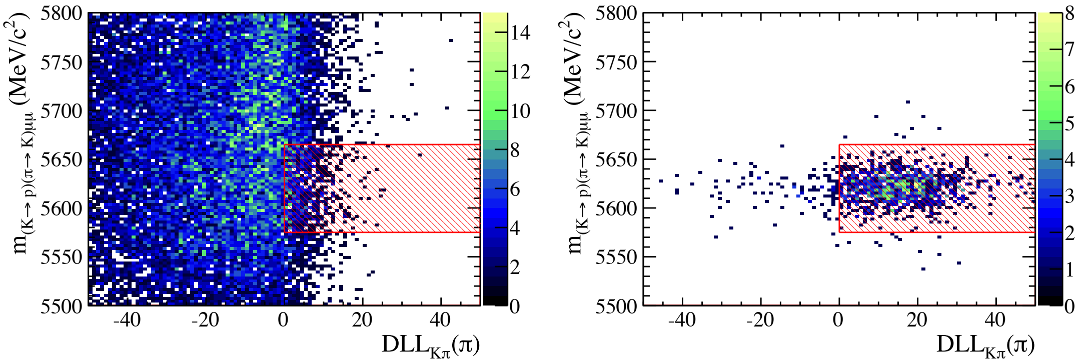
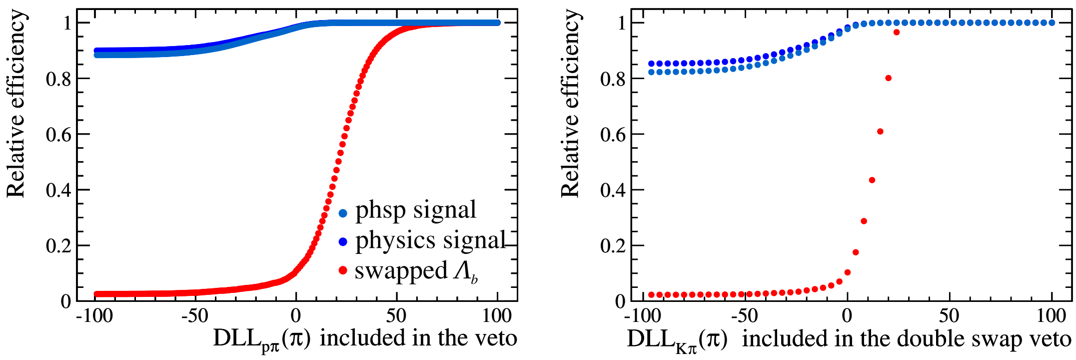


Figure A.3: The relative efficiency to select signal (blue) and reject misidentified $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays (red). Two samples are used to determine the signal efficiency. One simulated with a physics model (dark blue) and one simulated with a phsp model. The left-hand plot shows the effect of varying the cut at zero in Eqn. 4.5 for events which also fulfil Eqn. 4.4 and are candidate Λ_b^0 decays to be removed. The right-hand plot shows similar but for varying the cut in Eqn. 4.7 for the veto designed against doubly misidentified $\Lambda_b^0 \to pK^-\mu^+\mu^-$ events (*double swaps*).



Appendix B

Supplementary description of various functions

This appendix gives further information concerning various functions used in this thesis.

B.1 Crystal Ball function

The Crystal Ball (CB) function [79][1] is a Gaussian core with an asymmetric power-law tail. The function is used to model the $m_{K\pi\mu\mu}$ spectra. It is introduced in Sections 4.3.3 and 6.1. The function is defined

$$F_{\mathrm{CB}} = F_{\mathrm{CB}}(m_{K\pi\mu\mu};\, m_B, \sigma, \alpha, n) \tag{B.1}$$

$$\equiv \frac{1}{N_{\mathrm{CB}}} \begin{cases} \exp\left(-\dfrac{(m_{K\pi\mu\mu} - m_B)^2}{2\sigma^2}\right), & \text{for } \dfrac{m_{K\pi\mu\mu} - m_B}{\sigma} > -\alpha \\ X \cdot \left(Y - \dfrac{m_{K\pi\mu\mu} - m_B}{\sigma}\right)^{-n}, & \text{for } \dfrac{m_{K\pi\mu\mu} - m_B}{\sigma} \leqslant -\alpha \end{cases} \tag{B.2}$$

$$X = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right), \tag{B.3}$$

$$Y = \frac{n}{|\alpha|} - |\alpha| \tag{B.4}$$

and is normalised by the factor,

$$N_{\mathrm{CB}} = \sigma(U + V); \tag{B.5}$$

$$U = \sqrt{\frac{\pi}{2}} \left(1 + \mathrm{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right); \tag{B.6}$$

$$V = \frac{n}{|\alpha|} \cdot \frac{1}{n-1} \cdot \exp\left(= \frac{|\alpha|^2}{2}\right). \tag{B.7}$$

[1] Appendix E of [79]

B.2 Triangle function and phase-space momentum

The phase-space momentum for a decay of the type $M \to m_1 m_2$ is given by

$$\mathcal{P}(M, m_1, m_2) = \begin{cases} \frac{\sqrt{\lambda(M^2, m_1^2, m_2^2)}}{2M} & \lambda(M^2, m_1^2, m_2^2) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{B.8}$$

where the *triangle function* is given by,

$$\lambda(a, b, c) \equiv a^2 + b^2 + c^2 - 2ab - 2bc - 2ac. \tag{B.9}$$

In Eqns 6.15, 6.16, 6.18 and 6.21, the notation p, p_R, k, and k_R is used to denote various phase-space momenta. The K^* in the rest frame of the B is given by:

$$p \equiv \mathcal{P}(m_{B^0}, m_{K\pi}, \sqrt{q^2}), \tag{B.10}$$

and p_R is defined as above with the value of $m_{K\pi}$ set to that of the mean (pole mass) of resonance. For example, when evaluating Eqn. 6.15 for the $K_1^*(892)^0$: $p_R = p_{892} = \mathcal{P}(m_{B^0}, 896, \sqrt{q^2})$. The kaon momentum in the rest frame of the K^* is:

$$k \equiv \mathcal{P}(m_{K\pi}, m_K, m_\pi) \tag{B.11}$$

and similarly, k_R is obtained from the above with $m_{K\pi}$ at the mean of the resonance.

Appendix C

Supplementary information and figures related to the angular efficiency

The figures shown in this appendix are supplementary to those shown in Chapter 5 and Section 7.3.3. They either show the effect of (re)weighting simulation (Figs C.1, and C.7 or the efficiency projections for dimensions over which the integral is performed (Figs C.2–C.5). A brief proof of Eqn. 5.8 is also provided.

C.1 Figures supplementary to Chapter 5

Using the event weights a comparison of the track isolation distributions is shown Figure C.1. This figure is supplementary to Section 5.1.3 and Fig. 5.3.

Figure C.2 shows the 1D projections of the simulated efficiency and MoM results for $\cos\theta_\ell$ and ϕ. This figure is supplementary to Fig. 5.6.

Figures C.3–C.5 show 2D projections of the simulated efficiency and the integrated polynomial for all combinations containing $\cos\theta_\ell$ or ϕ.

C.2 Proof of Equation 5.8

The orthogonality condition for the Legendre polynomials is

$$\int_{-1}^{+1} P_m(x)P_n(x)\mathrm{d}x = \frac{2}{2n+1}\delta_{mn}. \tag{5.5}$$

The efficiency function is expanded as

$$\epsilon = \sum_{ghijk} c_{ghijk} P_g(m_{K\pi}{}') P_h(q^{2\prime}) P_i(\cos\theta_\ell) P_j(\cos\theta_K) P_k(\phi'). \tag{5.6}$$

Figure C.1: Comparison plots for isolation variables on linear (left) and logarithmic (right) scales. Uncorrected simulation (red) is weighted and re-sampled (blue) and compared to sFit weighted data (black points)

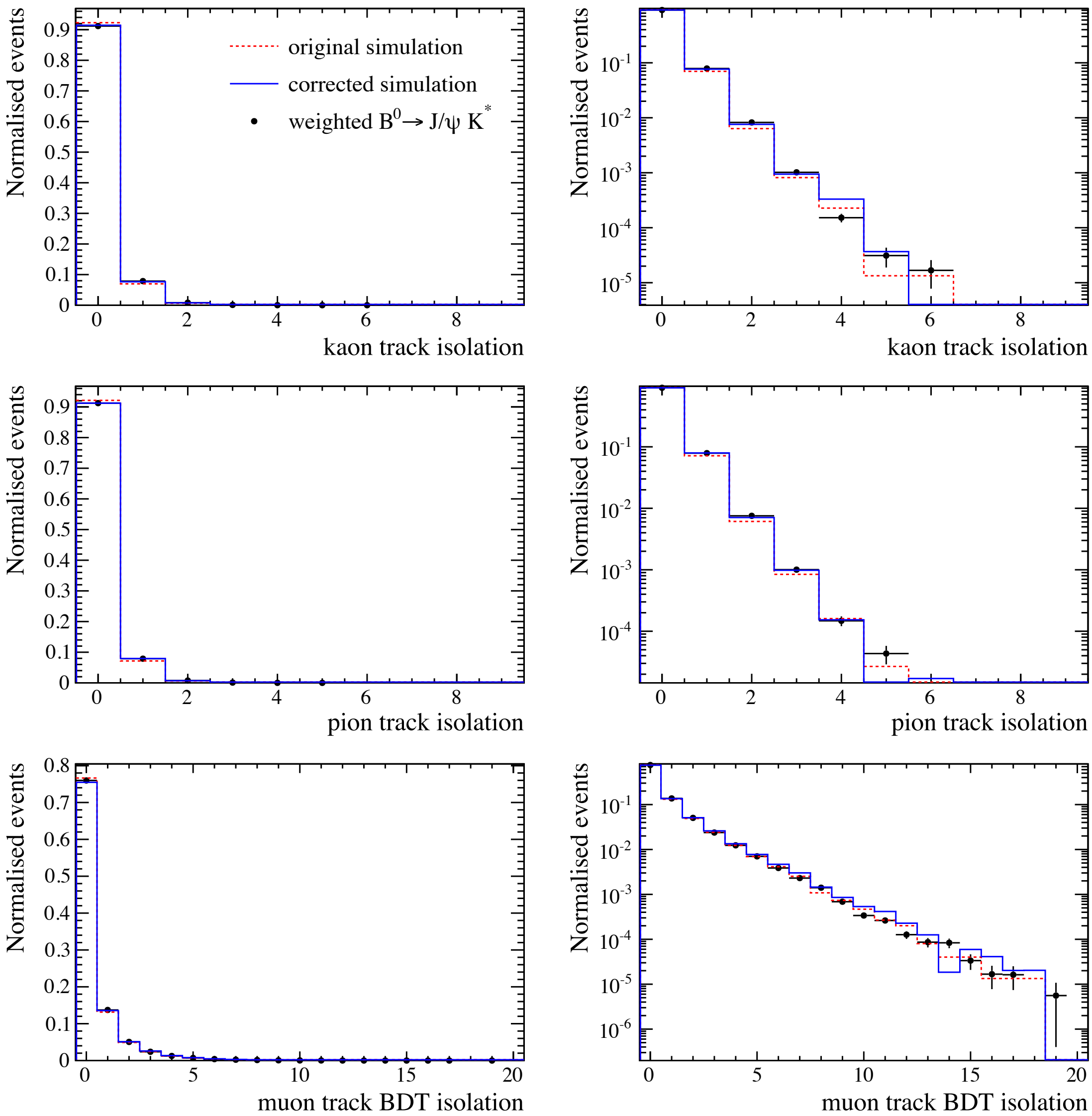
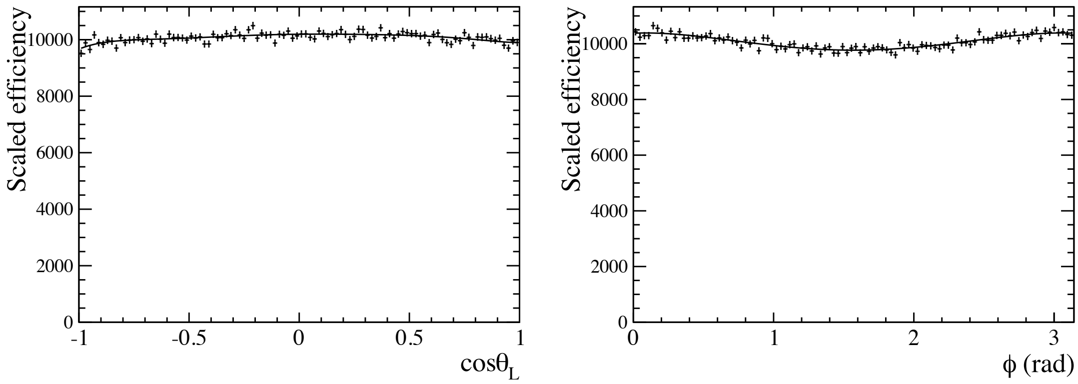


Figure C.2: One-dimensional projections of the simulated efficiency distribution with the polynomial function returned by the Legendre method of moments (MoM) integrated and overlaid. Supplementary to Fig. 5.6.



Multiplying Eqn. 5.6 through by $P_m(m_{K\pi}{}')$ and integrating over $m_{K\pi}{}'$,

$$\int P_m(m_{K\pi}{}')\cdot\epsilon\cdot \mathrm{d}m_{K\pi}{}' = \sum_{ghijk}\left[c_{ghijk}\int P_g(m_{K\pi}{}')P_m(m_{K\pi}{}')\mathrm{d}m_{K\pi}{}' \right.$$
$$\left. P_h(q^{2\prime})P_i(\cos\theta_\ell)P_j(\cos\theta_K)P_k(\phi')\right] \tag{C.1}$$
$$= \frac{2}{2m+1}\sum_{hijk}\left[c_{mhijk}P_h(q^{2\prime})P_i(\cos\theta_\ell)P_j(\cos\theta_K)P_k(\phi')\right]. \tag{C.2}$$

Repeating the process of multiplying by a term $P_n(x)$ and integrating for each variable yields

$$\int\!\!\!\int\!\!\!\int\!\!\!\int\!\!\!\int P_m(m_{K\pi}{}')P_n(q^{2\prime})P_o(\cos\theta_\ell)P_p(\cos\theta_K)P_q(\phi')\,\epsilon\,\mathrm{d}m_{K\pi}{}'\mathrm{d}q^{2\prime}\mathrm{d}\cos\theta_\ell\mathrm{d}\cos\theta_K\mathrm{d}\phi'$$
$$= \frac{2}{2m+1}\frac{2}{2n+1}\frac{2}{2o+1}\frac{2}{2p+1}\frac{2}{2q+1}c_{mnopq}, \tag{C.3}$$

which may be trivially rearranged to give Eqn. 5.8.

C.3 Figures supplementary to Section 7.3.3

Figure C.6 is supplementary to Fig 7.11 and shows the correction weights applied to the simulated events due to mismodelling the kaon momenta. Figure C.7 shows the effect of reweighting simulated data based on the distributions of hadron momenta shown in Figs C.6 and 7.11. The effect of mismodelled hadron kinematics contributes a relatively large portion of the systematic uncertainty, although considerably smaller than the uncertainty due to the limited statistics of the data.

Figure C.3: Two-dimensional projections of the sampled efficiency (left) and polynomial function returned by the Legendre method of moments (MoM) (right). The z-axis in both plots shows the projected efficiency scaled to the number of entries in the offline selected simulated sample, and is therefore arbitrary. This Fig, together with Figs C.4 and C.5 are supplementary to Fig. 5.7.

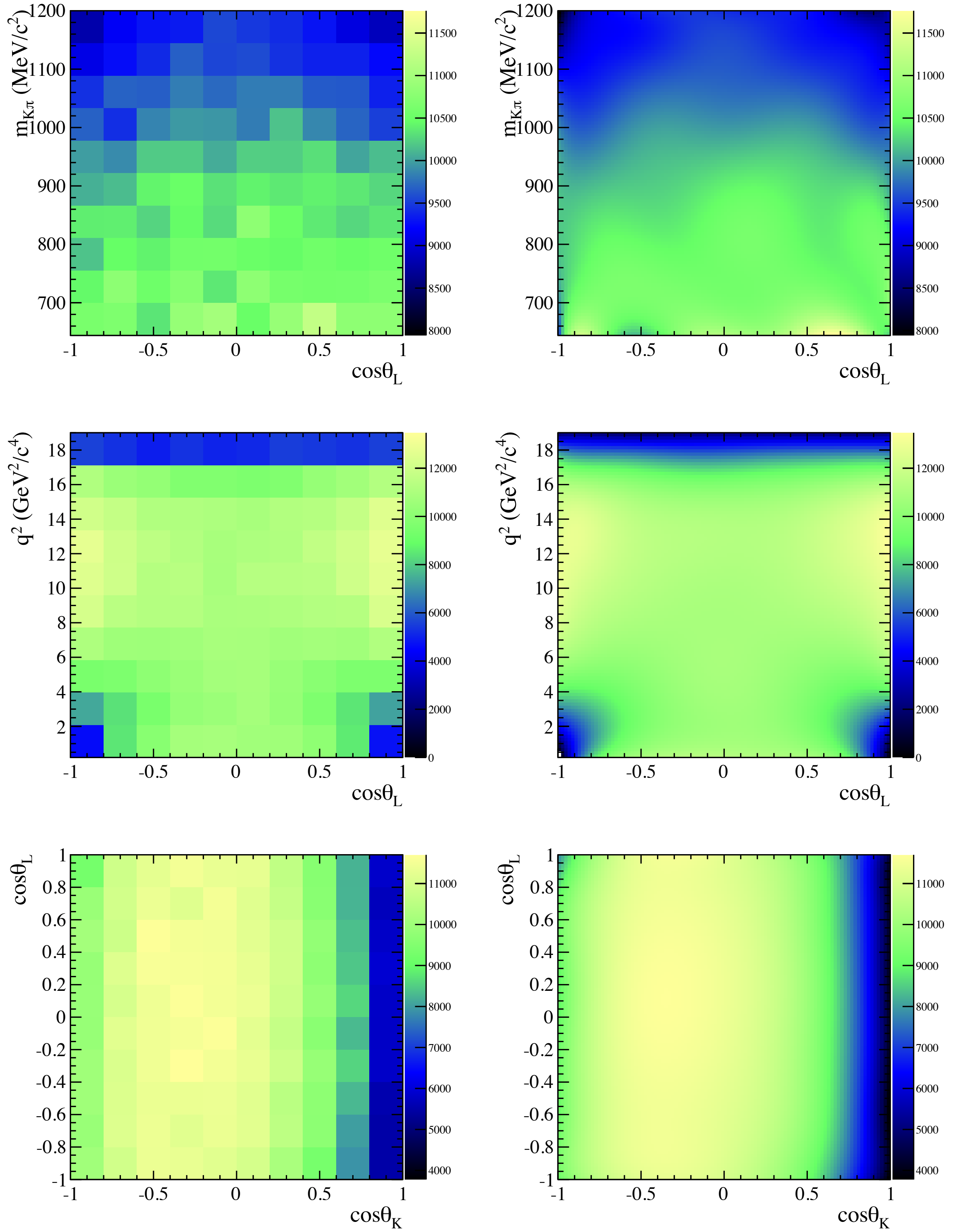


Figure C.4: Continued from Fig C.3

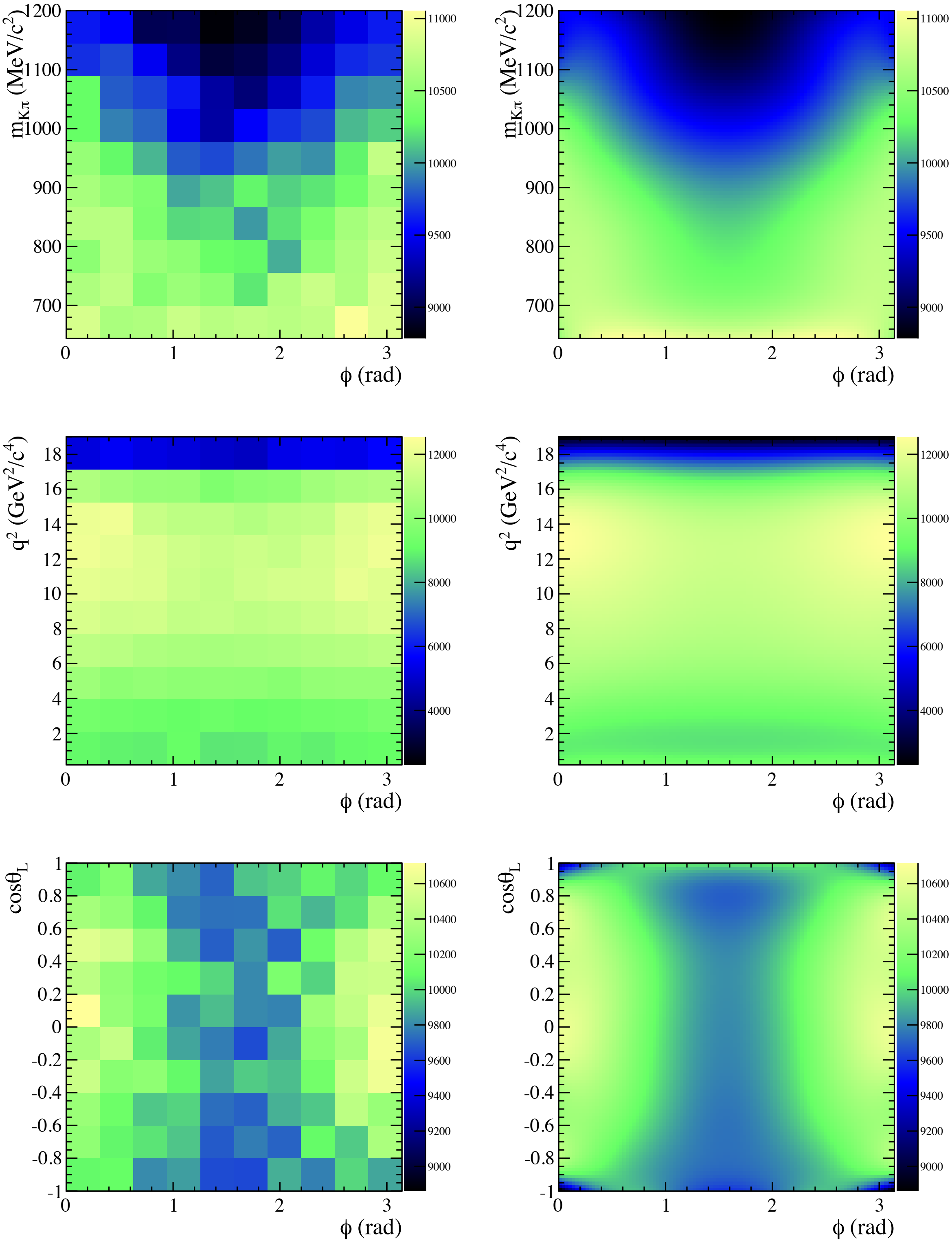


Figure C.5: Continued from Fig C.4

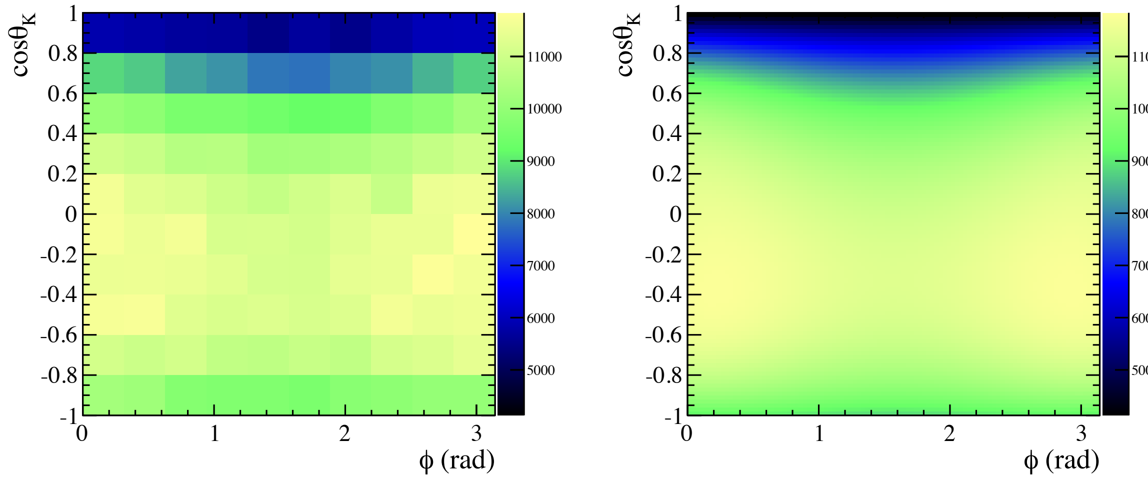


Figure C.6: Supplementary to Fig. 7.11 in Section 7.3.3. Extra correction weights (bottom centre) are derived for disagreement between the kaon momenta in data (left) and simulation (right) for sFit weighed $B^0 \rightarrow J/\psi\, K^{*0}$ with a narrow window around the $K_1^*(892)^0$.

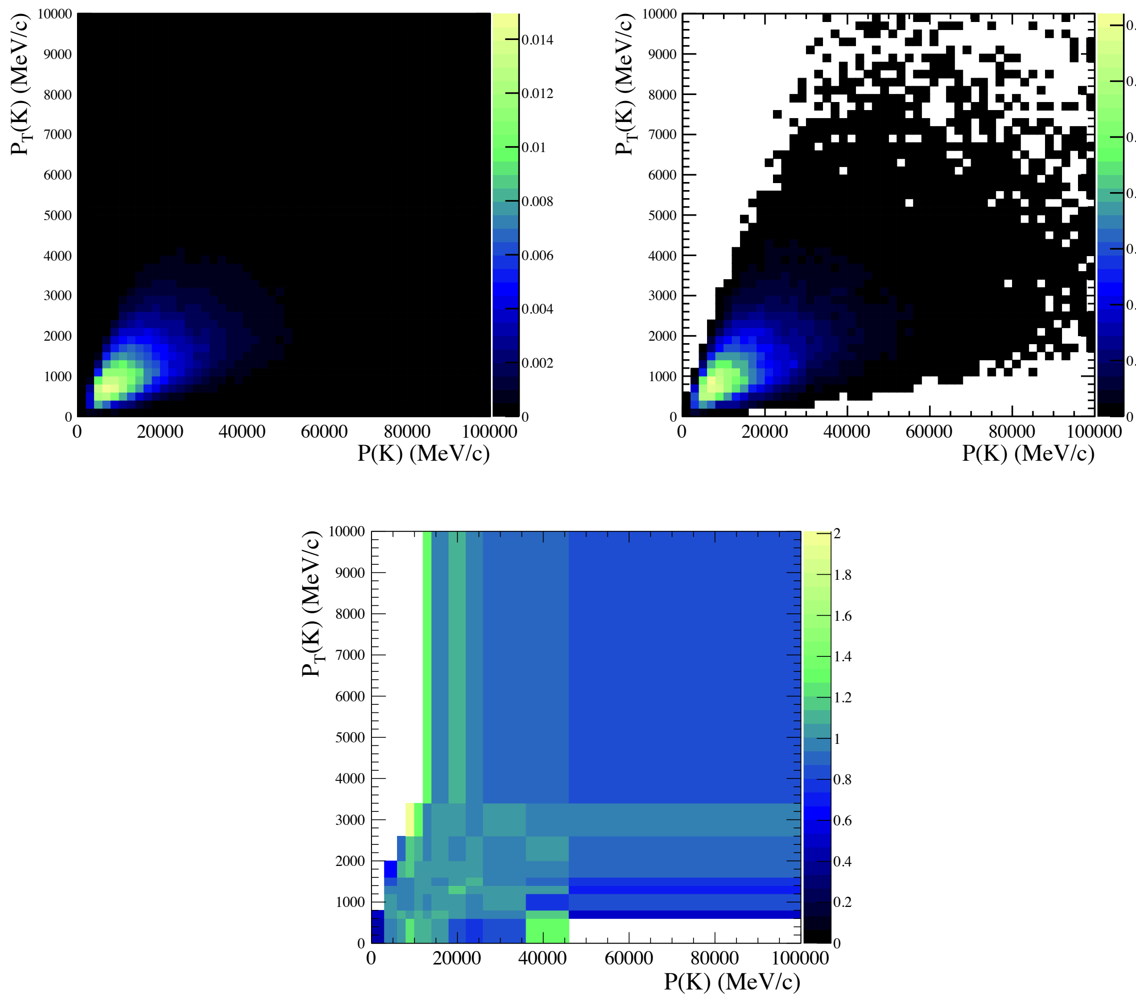
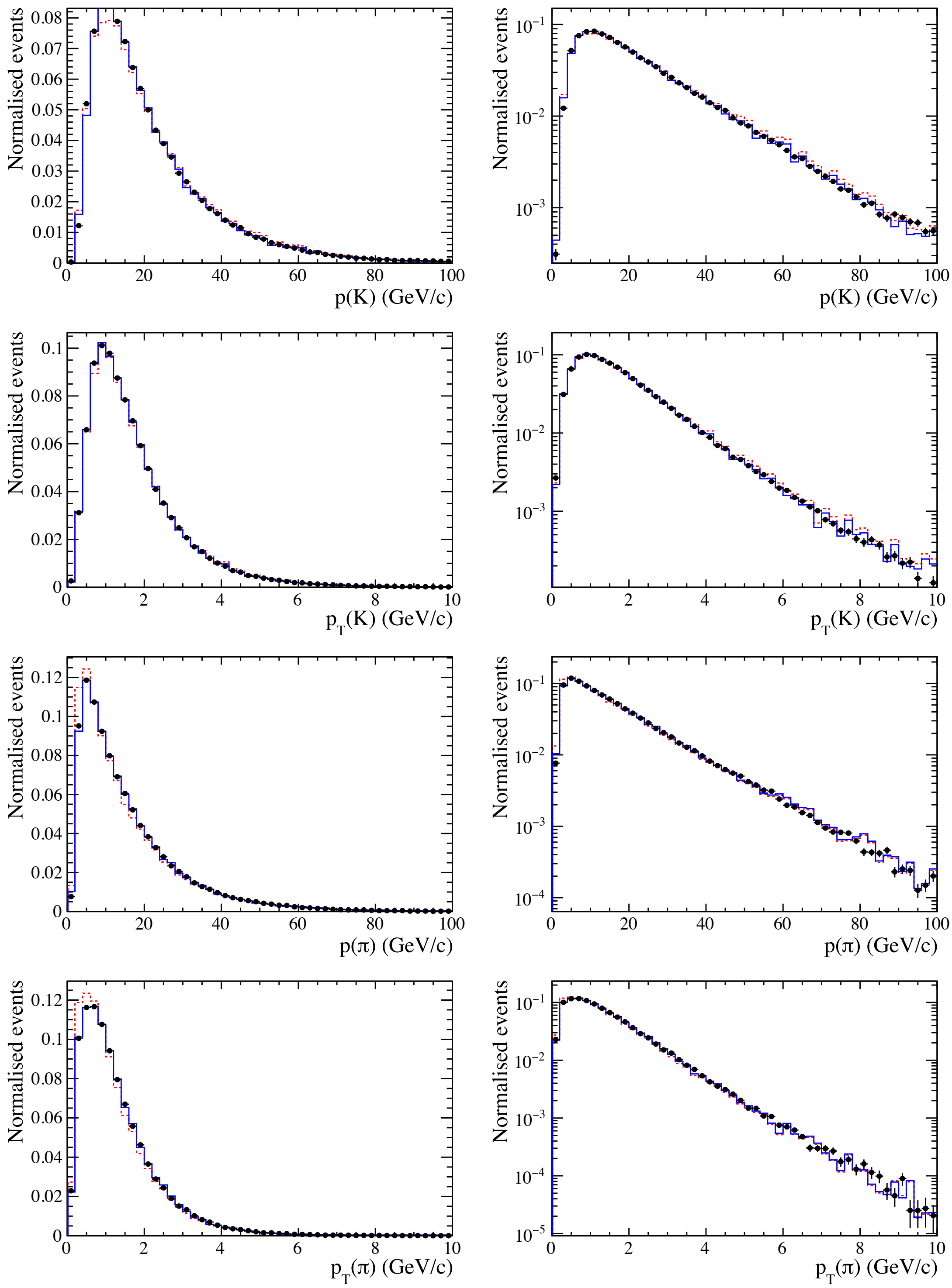


Figure C.7: Validation plots for the reweighting of hadron kinematics to account for residual data-simulation differences. The plots, in linear (left), and logarithmic (right) scales show momentum distributions for the hadrons. The simulation is corrected as described in Section 5.1 (red histogram), and compared to sFit weighted $B^0 \to J/\psi K^{*0}$ data (black points). Further weights are derived and applied resulting in the blue histogram. These weights are applied to the phase-space simulation sample and two sets of acceptance coefficients are rederived accounting for the kaon and pion reweighting.



Appendix D

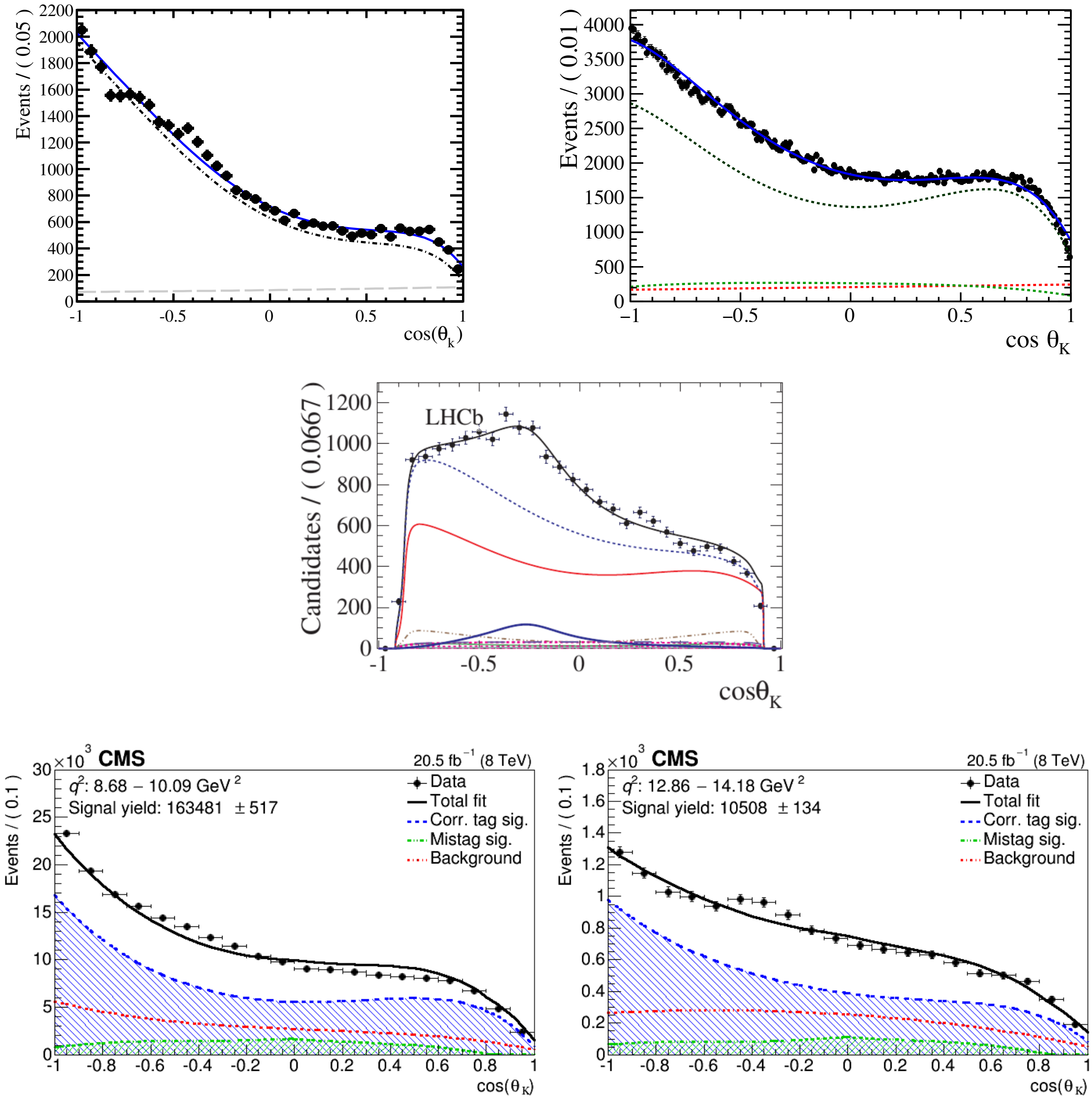
Effect of a possible new resonant state in the control channel fit

This appendix briefly describes studies performed on the control channel $B^0 \to J/\psi K^+\pi^-$ fits in order to understand the different values of F_S found by different fit methods. The appendix is supplementary to Section 6.6. Some values quoted are from the angular fit presented in Ref. [29] as well as analysis methods to be published soon, based on the methods described in Refs [86, 102].

The differences observed are consistent with the effect of a resonance, X^-, similar to the $Z(4430)^-$ state recently observed [90] in the decay process $B^0 \to \psi(2S)K^+\pi^-$. The $Z(4430)^-$ is a tetraquark candidate and contributes to these decays as $B^0 \to Z(4430)^- K^+$ where the $Z(4430)^- \to \psi(2S)\pi^-$. The contribution to the control channel decay, $B^0 \to J/\psi K^+\pi^-$, would be of the analogous form $B^0 \to X^- K^+$ with $X^- \to J/\psi\pi^-$. Such a state is claimed to be present in the Belle analysis of $B^0 \to J/\psi K^+\pi^-$ [96]. The presence of this state could explain the mismodelling in the region $\cos\theta_K \in [-0.8, -0.5]$. The effect also appears in the $\cos\theta_K$ spectrum of the CMS analysis [95]. Figure D.1 shows several $\cos\theta_K$ distributions where this mismodelling may be seen. This contribution can only affect the control channel and cannot contribute to non-resonant $B^0 \to K^+\pi^-\mu^+\mu^-$ signal decays. The mismodelling is also below the level of precision required for the analysis of $B^0 \to K^+\pi^-\mu^+\mu^-$ which suffers from low statistics. A full analysis of this structure is ongoing within the LHCb collaboration.

The value of F_S found by the fit presented in Fig 6.13 and Table 6.4 is lower than values found by other methods. In general, methods including $m_{K\pi}$ are found to give consistently lower values of F_S than methods including angular information. This is most evident in comparing the result of the nominal 3D fit model ($F_S|_{796}^{996} = 5.19 \pm 0.06\%$) with the result of a purely angular fit ($F_S|_{796}^{996} = 8.3 \pm 0.3\%$). The angular fit is also performed with the addition of the $m_{K\pi}$ dimension simultaneously, resulting in a value $F_S|_{796}^{996} = 7.9 \pm 0.3$. Whilst the angular and $m_{K\pi}$ spectra provide somewhat orthogonal information, and values found with the fit method described in this thesis are measured on a wider window of

Figure D.1: An indication of a possible X^- contribution in an excess of events around $\cos\theta_K \in [-0.8, -0.5]$ for the control channels in this analysis *cf.* figures where the $Z(4430)^-$ component was discovered. The result of an amplitude method fit [102] to a bin $m_{K\pi} \in [826, 861]\,\mathrm{MeV}/c^2$ is shown to the (upper left) and the projection of the nominal LASS model 3D fit described in this thesis (upper right). The upper right-hand figure is also shown in Fig. 6.13. These should be compared to the $\cos\theta_K$ distribution of $B^0 \to \psi(2S)K^*$ decays where a component is included for the $Z(4430)^-$ state (central figure, dark blue component). The central figure is given by private correspondence with the authors of Ref. [90]. Finally, the results of the CMS analysis control channel fits (taken from Ref. [95]) to $B^0 \to J/\psi\, K^{*0}$ (lower left) and $B^0 \to \psi(2S)K^*$ (lower right) show similar mismodelling.



$m_{K\pi}$[1], such a difference is unexpected. However, the angular and angular-with-$m_{K\pi}$ fitted values are performed on *exactly* the same dataset.

The analysis method of this thesis (both the 3D and 2D fits) agrees with the fitted F_{S} found by angular methods for pseudodata samples of a known F_{S}. This has been tested extensively with pseudodata samples generated with several different models. To test the hypothesis of a X contribution, the model developed for the analysis published in Ref. [90] is used to generate pseudodata samples. These pseudodata samples are fit with several methods, including the 3D fit documented in this thesis. Differences are observed in the fitted F_{S} values in the range $1-3\%$, dependent on the parameters of the X^{-}. A summary of several well motivated model configurations is provided in Table D.1. Also of note, the angular methods [29, 102], when performing fits to $m_{K\pi}$, suffer from poor modelling uncertainty as the fit is performed in a region $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$. The analysis method described in this thesis, is performed in the range $m_{K\pi} \in [644, 1200]\,\mathrm{MeV}/c^2$ which provides a better modelling control of the resonance parameters. Indeed, the angular analysis methods quote a systematic uncertainty on $F_{\mathrm{S}}|_{796}^{996}$ due to modelling of the $m_{K\pi}$ spectrum in the range $1-2\%$.

[1]and therefore differing datasets

Table D.1: A summary of the results of several fitters results on toy data with and without a resonant X^- contribution. All pseudodata samples are generated using the model adapted from the analysis of Ref. [90] with resonance parameters from Ref. [96] for the J/ψ resonance rather than the $\psi(2S)$. The rows marked as "Angular" fits are performed using the fit for angular observables described in Refs [29]. The '3D' fit is documented in this thesis. The rows marked "1D" are performed with a fit to only the $m_{K\pi}$ spectrum with an independent implementation from that of this thesis, but following Eqn. 6.11. The observed differences between methods are sufficient to explain the discrepancy observed in the control channel $B^0 \to J/\psi K^+\pi^-$, between angular fit methods and fit methods involving $m_{K\pi}$. Uncertainties shown are statistical only.

Model configuration	Fit method	$F_S\|_{796}^{996}$ (%)
X^- amplitude at default for the $Z(4430)^-$	Angular	0.0668 ± 0.0020
	3D	0.0545 ± 0.0013
	1D	0.0497 ± 0.0016
X^- with width $100\,\text{MeV}/c^2$ larger than $Z(4430)^-$	Angular	0.0695 ± 0.0020
	3D	0.0543 ± 0.0013
	1D	0.0484 ± 0.0016
X^- state with width $100\,\text{MeV}/c^2$ smaller than $Z(4430)^-$	Angular	0.0754 ± 0.0020
	3D	0.0557 ± 0.0013
	1D	0.0531 ± 0.0016
X^- amplitude scaled up by $\sqrt{2}$ (*i.e.* 50% larger contribution)	Angular	0.0683 ± 0.0020
	3D	0.0557 ± 0.0012
	1D	0.0510 ± 0.0016
No X^- state	Angular	0.0521 ± 0.0020
	3D	0.0502 ± 0.0015
	1D	0.0464 ± 0.0016
No X^- state, with $K_0^*(800)^0$ parameters set to averages from Section 6.2.3	Angular	0.0399 ± 0.0020
	3D	0.0339 ± 0.0007
	1D	0.0358 ± 0.0016
No X^- state including D-wave $K\pi$ states	Angular	0.0575 ± 0.0020
	3D	0.0527 ± 0.0016
	1D	0.0525 ± 0.0016

Appendix E

Supplementary fitted parameter values and fitted projections of the result for each q^2 bin

E.1 Parameters found by the fit to $K\pi\mu\mu$ invariant mass

The fitted $m_{K\pi\mu\mu}$ spectra for control channel data shown in Fig 6.1 correspond to fits to two regions of $m_{K\pi}$. The upper two figures are the result of a fit to $B^0 \to J/\psi K^{*0}$ data for the regions $q^2 \in [9.22, 9.96]\,\mathrm{GeV}^2/c^4$ and $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$ which is used to extract the yield of $B^0 \to J/\psi K^{*0}$ events. The values of the parameters found by this fit, which are not used in any other area of the analysis presented in this thesis, are given in Table E.1 for completentess.

E.2 Fitted distributions of $K\pi\mu\mu$ invariant mass spectrum for all remaining q^2 bins

This appendix contains the remaining fitted $m_{K\pi\mu\mu}$ spectra for the fitted yields given in Table 6.3. Figures E.1 and E.2 are supplementary to Fig. 6.12 in Section 6.5.1.

E.3 Fitted projections for all remaining q^2 bins

This appendix shows the remaining projections of the fit for the q^2 bins that have not been included in Chapter 7. Table E.2 provides a summary of the figures corresponding to all q^2 bins, for convenience.

Figure E.1: Continued from Fig. 6.12. Fits to the $m_{K\pi\mu\mu}$ spectrum for $B^0 \to K^+\pi^-\mu^+\mu^-$ decays shown on linear (left) and logarithmic (right) scales in bins of q^2 (rows). The set of shape parameters for the double-CB, $\vec{\alpha}$ and m_{B^0}, are fixed to those of Table 6.1. See Fig. E.2 for further bins.

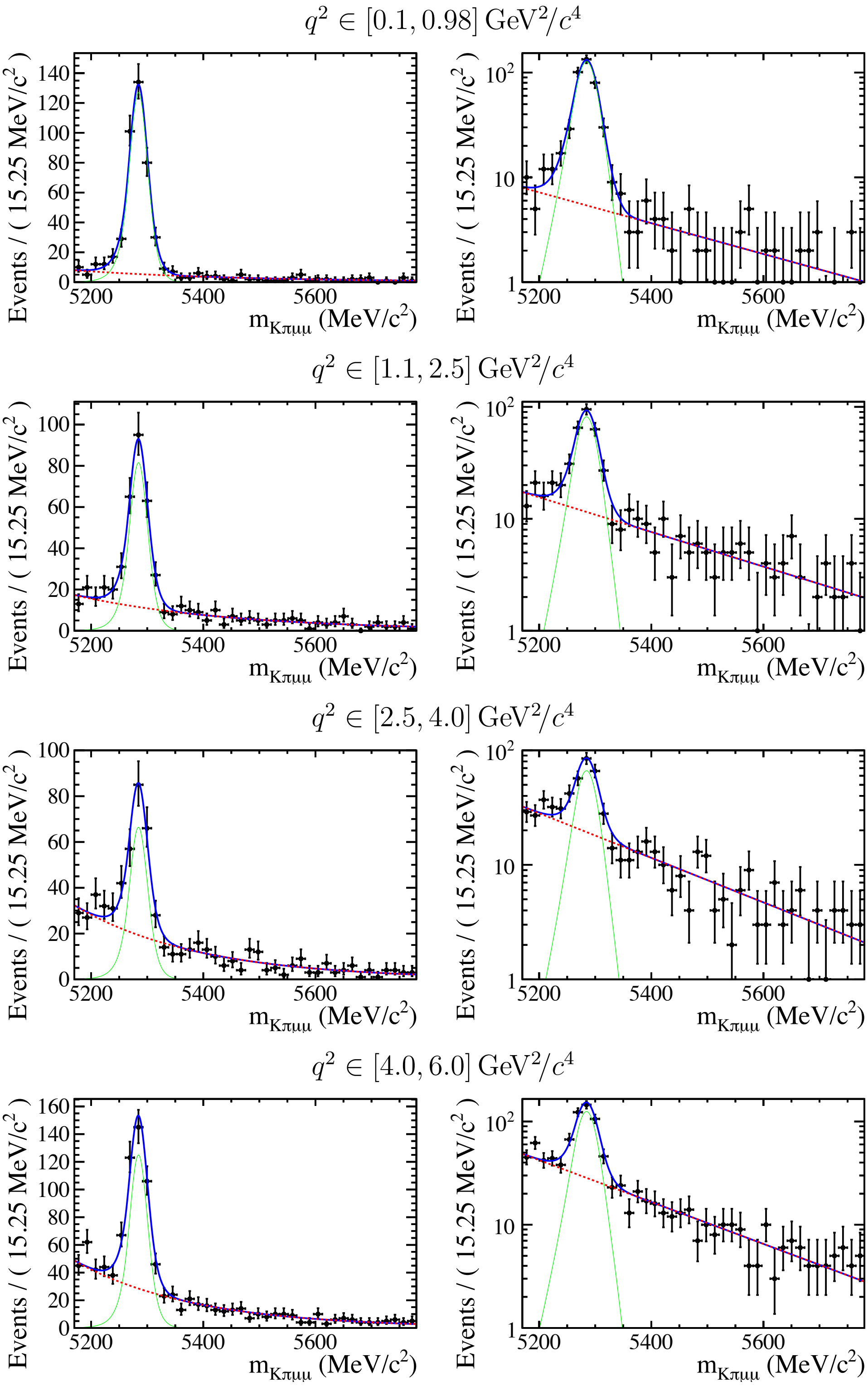


Figure E.2: Continued from Fig. E.1. Fits to the $m_{K\pi\mu\mu}$ spectrum for $B^0 \to K^+\pi^-\mu^+\mu^-$ decays shown on linear (left) and logarithmic (right) scales in bins of q^2 (rows). The set of shape parameters for the double-CB, $\vec{\alpha}$ and m_{B^0}, are fixed to those of Table 6.1.

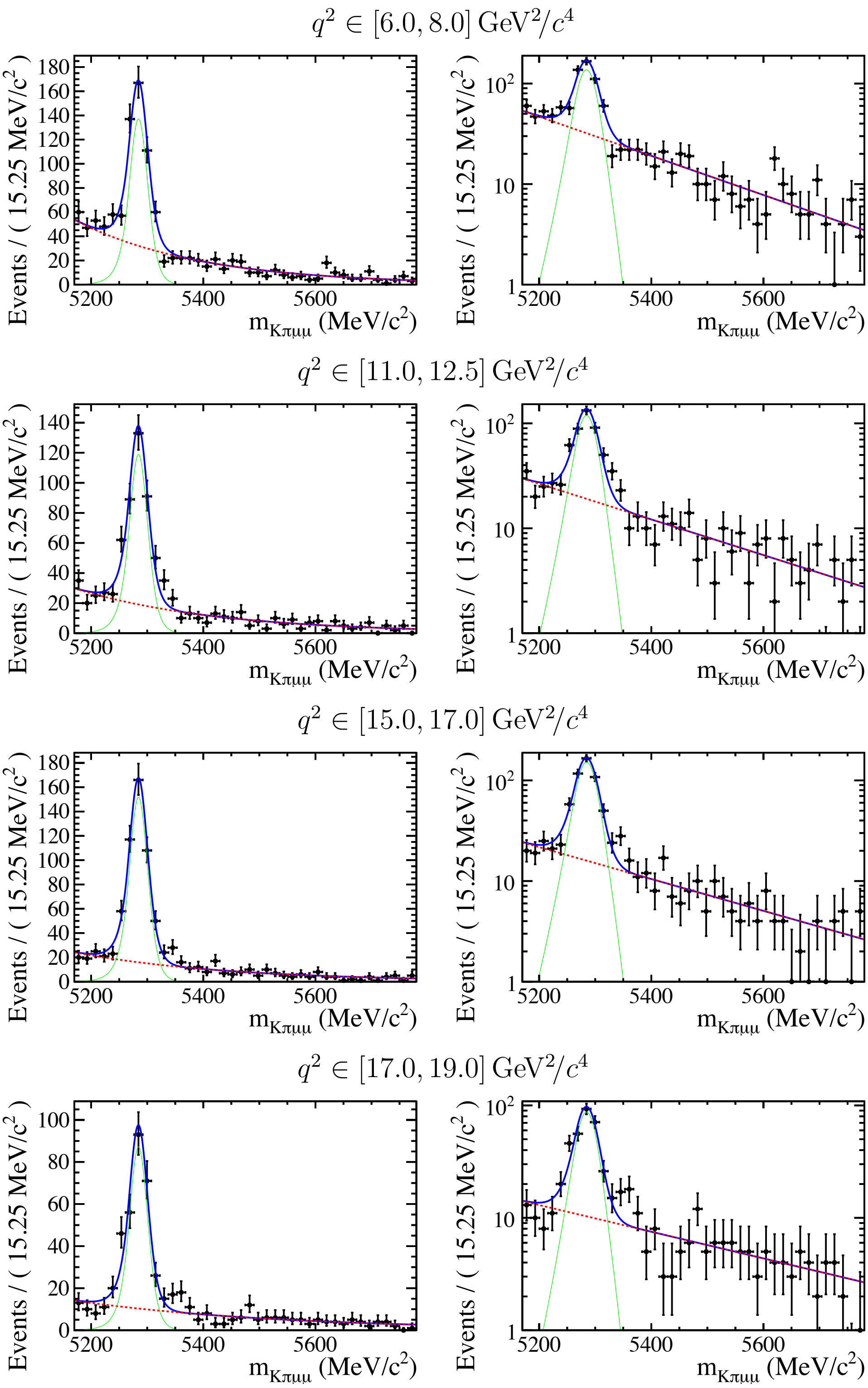


Figure E.3: The result of the full 3D fit to the data for the bin $q^2 \in [0.1, 0.98]\,\mathrm{GeV}^2/c^4$.

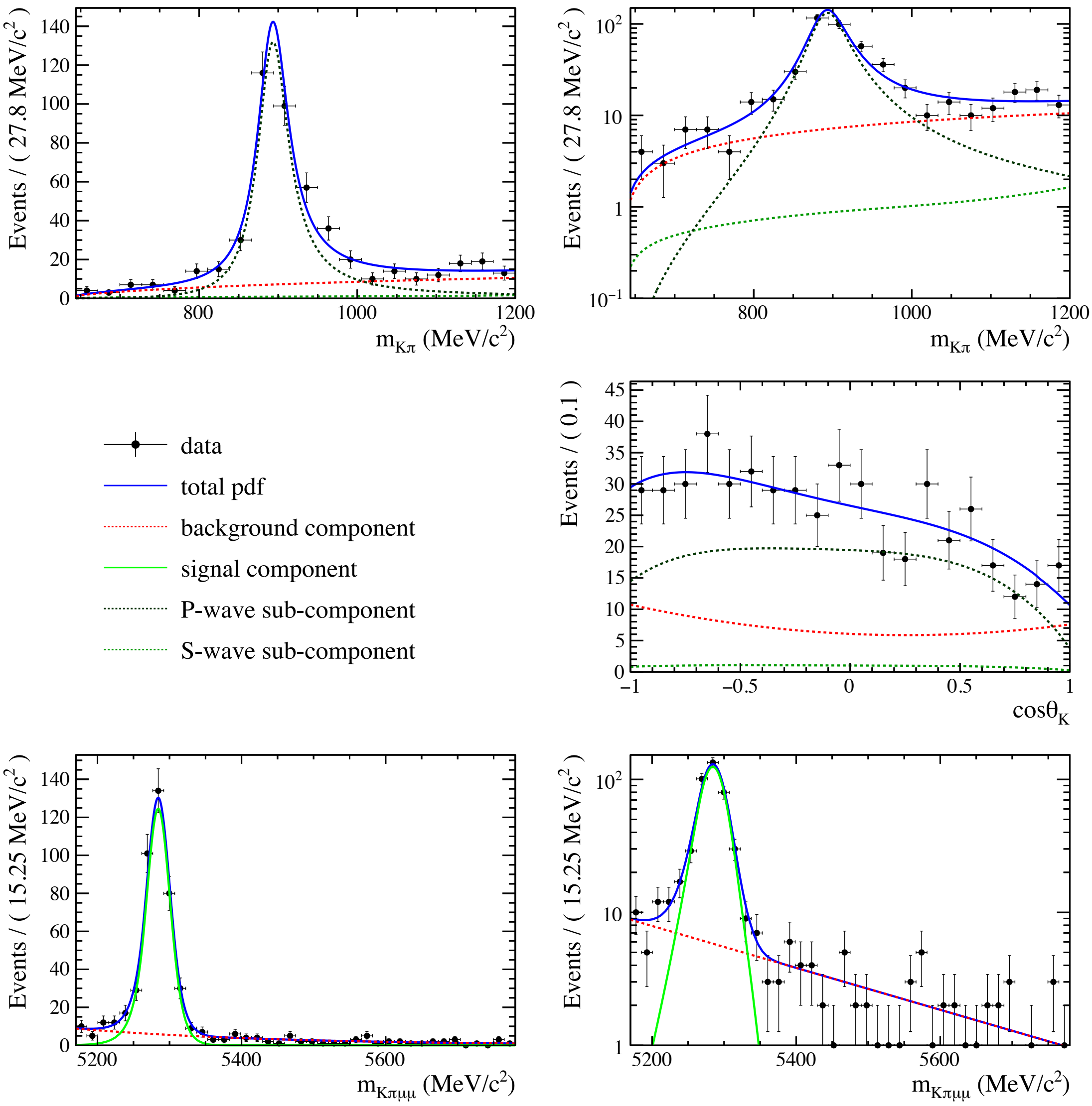


Figure E.4: The result of the full 3D fit to the data for the bin $q^2 \in [1.1, 2.5]\,\mathrm{GeV}^2/c^4$.

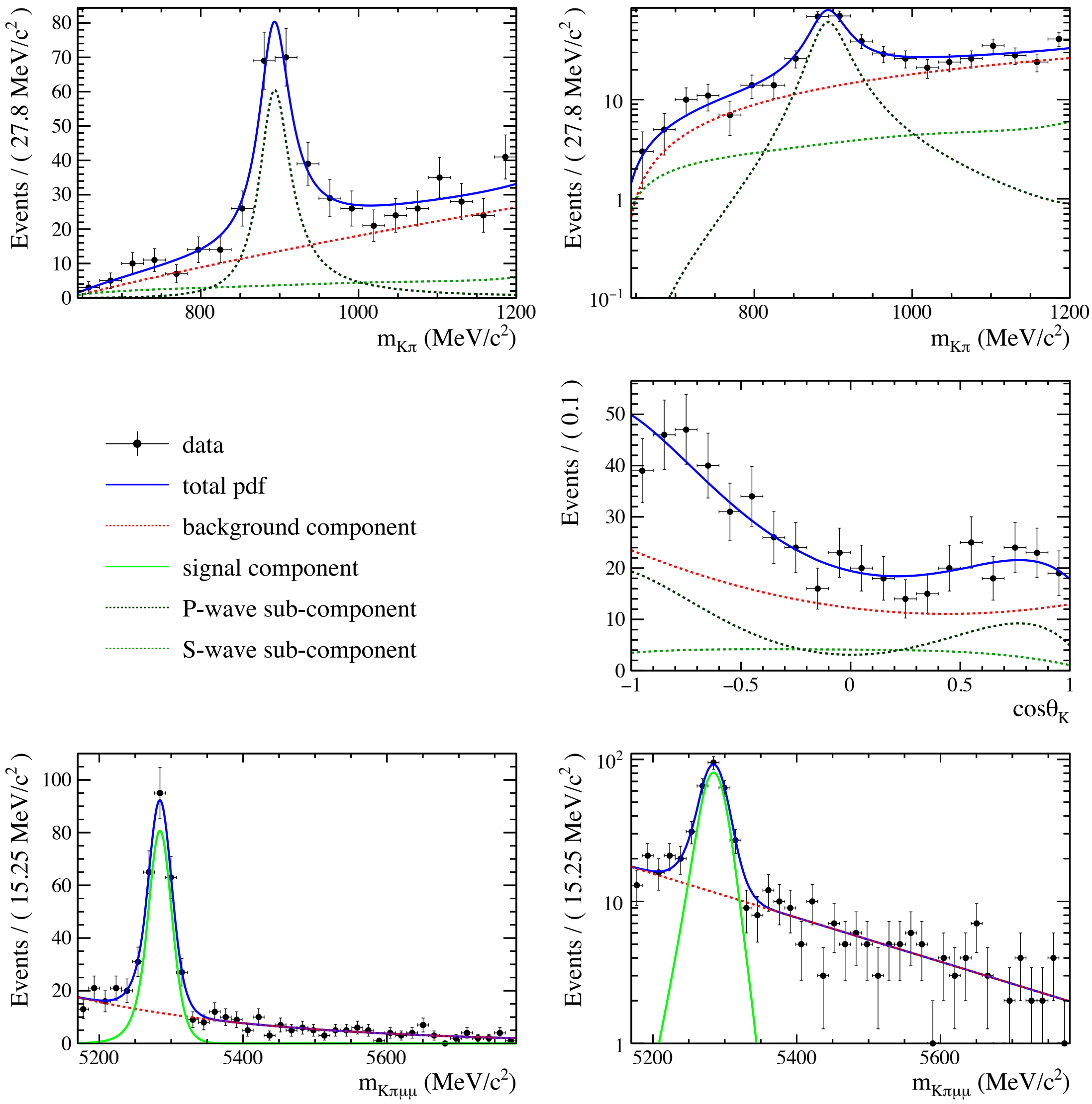


Figure E.5: The result of the full 3D fit to the data for the bin $q^2 \in [6, 8]\,\mathrm{GeV}^2/c^4$.

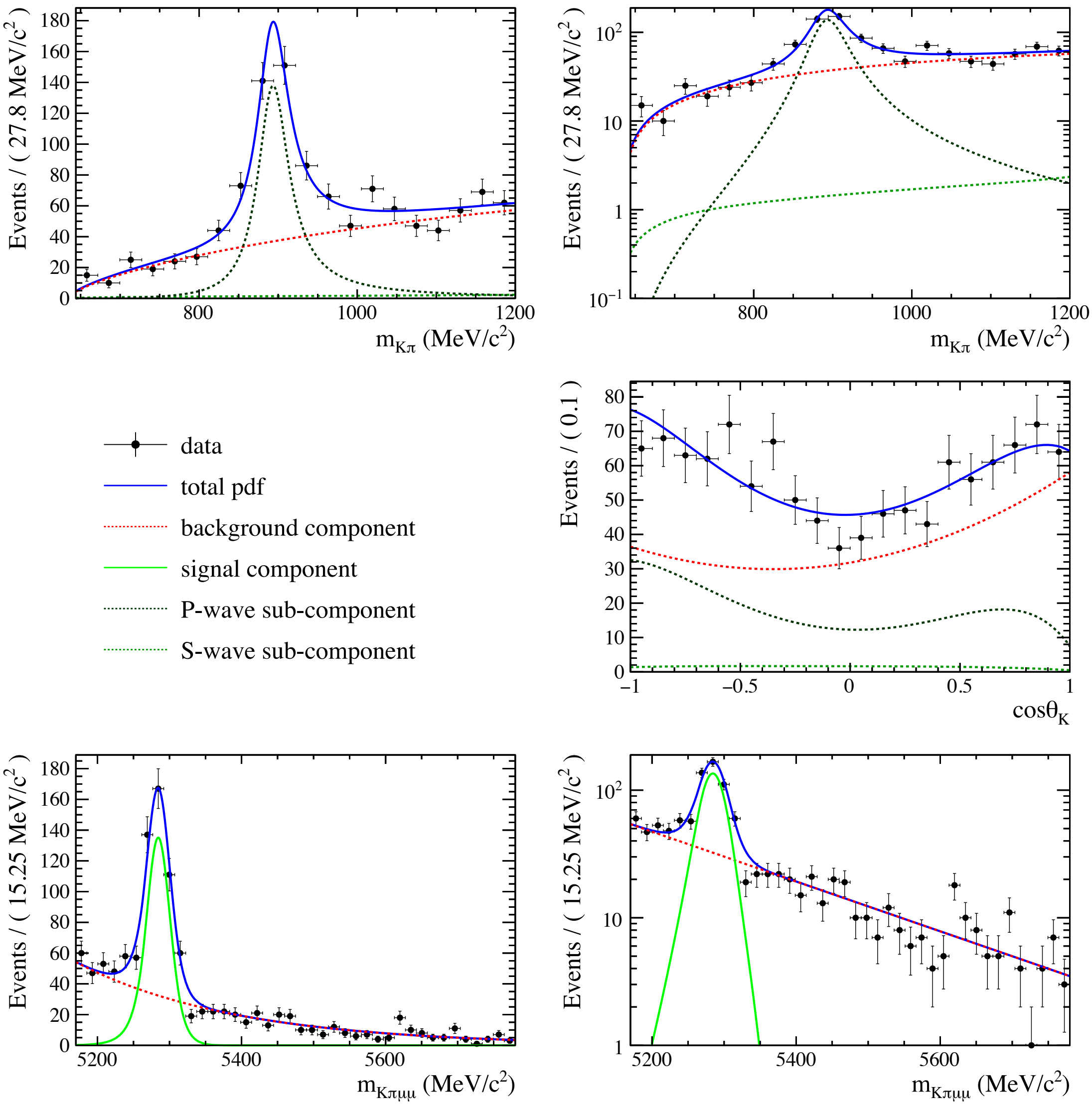


Figure E.6: The result of the full 3D fit to the data for the bin $q^2 \in [11, 12.5]\,\mathrm{GeV}^2/c^4$.

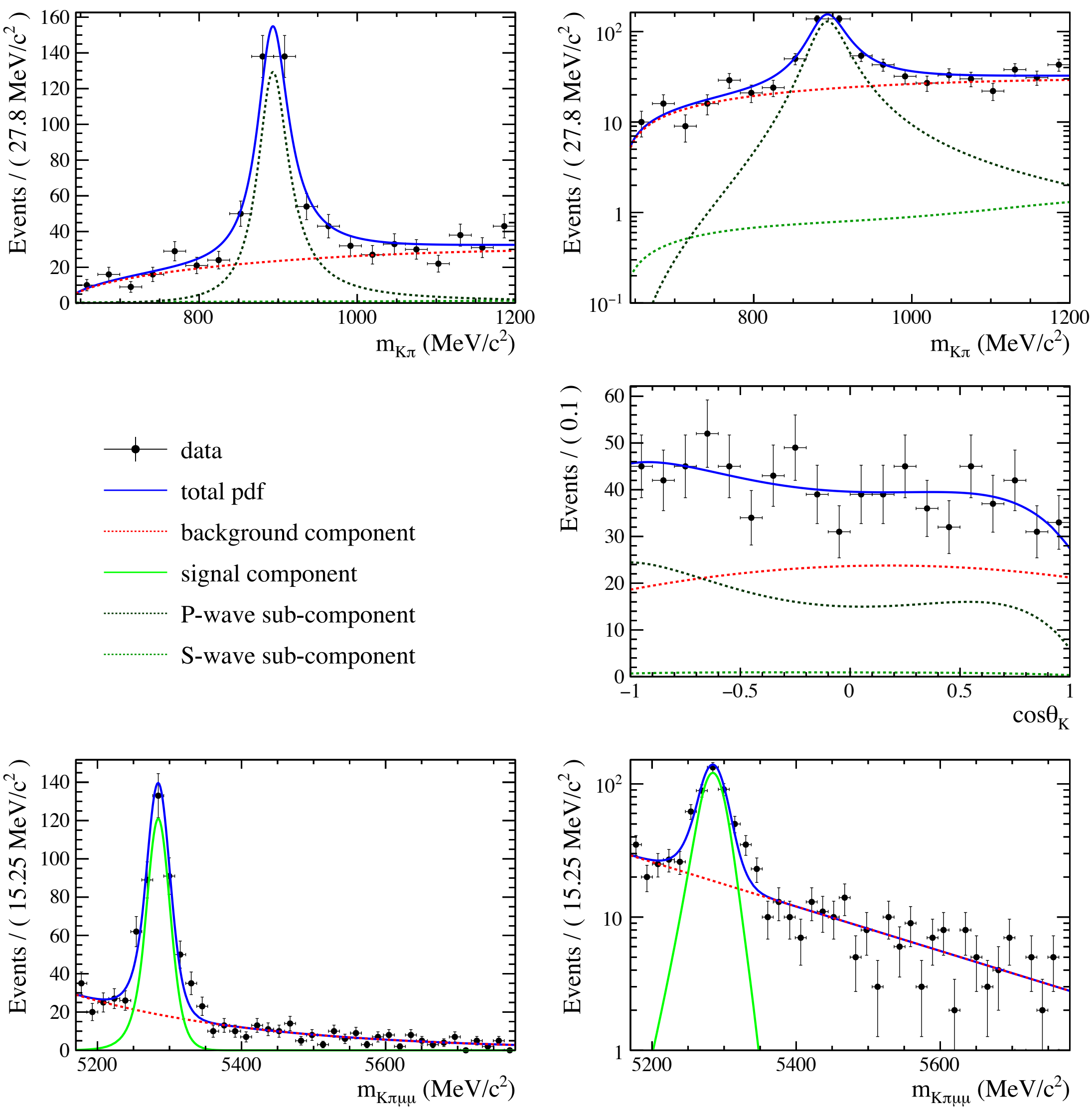


Figure E.7: The result of the full 3D fit to the data for the bin $q^2 \in [15, 17]\,\mathrm{GeV}^2/c^4$.

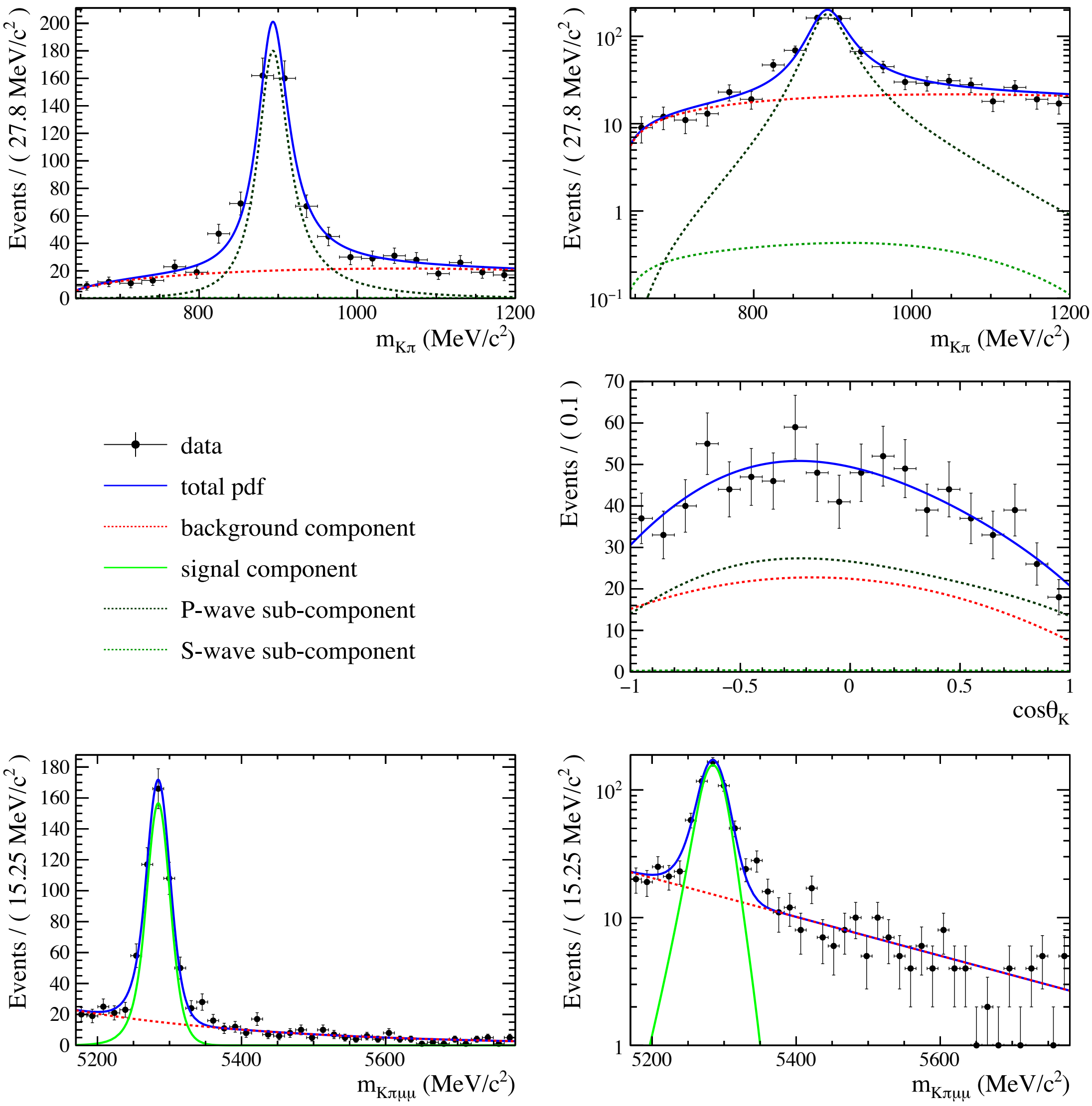


Figure E.8: The result of the full 3D fit to the data for the bin $q^2 \in [17, 19]\,\mathrm{GeV}^2/c^4$.

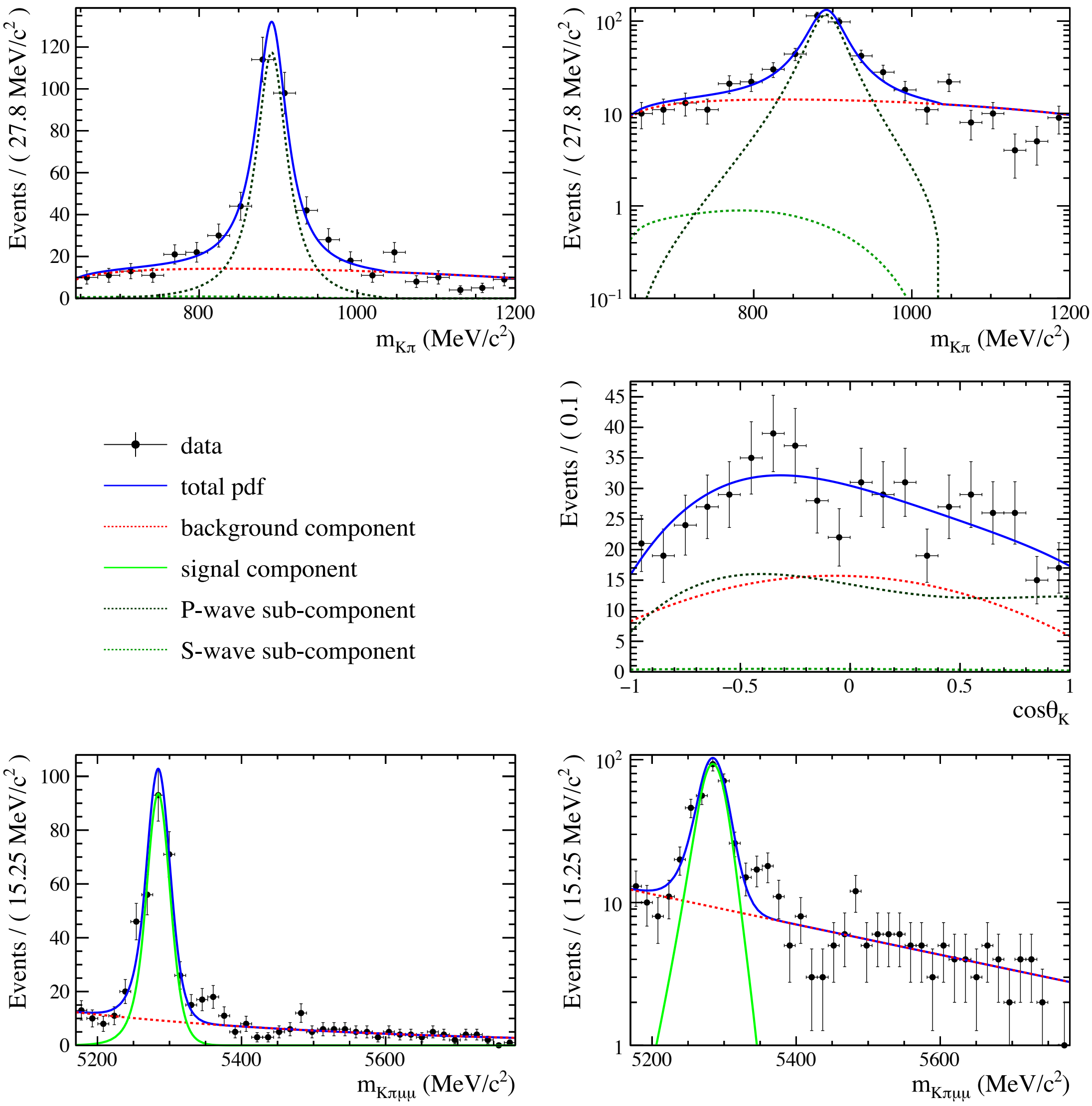


Table E.1: The fitted parameters from the fits to $B^0 \to J/\psi K^{*0}$ and $B_s^0 \to J/\psi K^*$ candidates in the window $m_{K\pi} \in [796, 996]\,\mathrm{MeV}/c^2$. These parameters correspond to the upper two plots in Fig. 6.1. Only the yield of $B^0 \to J/\psi K^{*0}$ events, n_{B^0}, is used in the analysis.

parameter	value	units
n_{bkg}	15089 ± 258	–
n_{sig}	374730 ± 616	–
m_{B^0}	5284.40 ± 0.041	MeV/c^2
$m_{B_s^0} - m_{B^0}$	87 (fixed)	MeV/c^2
$f_{B_s^0}$	1.170 ± 0.029	–
n_{B^0}	324482 ± 609	–
$n_{B_s^0}$	3844 ± 98	–
p_0	-5.511 ± 0.080	$10^{-3}[\mathrm{MeV}/c^2]^{-1}$
α	1.329 ± 0.0107	–
n	35 (fixed)	–
f_{σ_1}	0.671 ± 0.022	–
σ_1	15.12 ± 0.170	MeV/c^2
σ_2	24.78 ± 0.499	MeV/c^2

q^2 bin (GeV^2/c^4)	Fig.
$[0.1, 0.98]$	E.3
$[1.1, 2.5]$	E.4
$[2.5, 4.0]$	7.4
$[4.0, 6.0]$	7.5
$[6.0, 8.0]$	E.5
$[11.0, 12.5]$	E.6
$[15.0, 17.0]$	E.7
$[17.0, 19.0]$	E.8
$[1.1, 6.0]$	7.3

Table E.2: A summary of the figures corresponding the 3D fit results for each q^2 bin. Some are included in Chapter 7, the remainder in this Appendix.