The angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay at LHCb

Christopher Parkinson

High Energy Physics Group
Department of Physics
Imperial College London

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

June 2013

Declaration

The work presented in this thesis was carried out between October 2009 and June 2013. It is the result of my own studies, with the support of members of the Imperial College HEP group and the broader LHCb collaboration. The work of others is explicitly referenced.

The entire thesis was written by myself. Regarding the $B^0 \to K^{*0}\mu^+\mu^-$ analysis presented in this thesis, I developed the selection described in Sec. 4, performed the measurement of the differential branching fraction in Sec. 6, and performed the measurement of the angular observables in Sec. 7.

This thesis has not been submitted for any other qualification.

Christopher John Parkinson, June 2013

Acknowledgements

I would like to express my gratitude to my supervisor Mitesh Patel for his patient guidance, his valuable and constructive suggestions regarding this research work, and his numerous useful critiques of this thesis. I would also like to thank Thomas Blake and Alex Shires, with whom I had the pleasure to work closely with during my Ph.D. Furthermore I would like to thank the members of the LHCb group at Imperial College, especially Ulrik Egede, for his assistance and advice with various aspects of my Ph.D. I would also like to thank the students at Imperial College, both for useful discussions and for making the last three years much more enjoyable. Finally I would like to thank my parents for their continued support.

This thesis is dedicated to Agni: the other half of my life.

Abstract

The Standard Model (SM) of particle physics is an incomplete description of the fundamental particles and their interactions. In extensions of the SM, new physics particles are expected to appear at the $\sim$ TeV mass scale. The discovery of such new physics particles could help to answer some of the greatest questions in physics: what constitutes 'dark' matter; why is there more matter than antimatter; are there more than four dimensions?

Particles at the $\sim$ TeV scale are expected to contribute virtually to interactions between SM particles at $\sim$ GeV energy scales. In interactions that are mediated by loop processes, such as the flavour-changing neutral current decay $B^0 \to K^{*0}\mu^+\mu^-$, the contribution from new physics particles can be as large as that of SM particles. Certain experimentally accessible observables of the $B^0 \to K^{*0}\mu^+\mu^-$ decay are precisely predicted in the SM and in new physics models. Any deviation in the measured value of the observables from the SM prediction would indicate a contribution from new physics particles.

This thesis reports the analysis of the world's largest sample of $B^0 \to K^{*0}\mu^+\mu^-$ decays. The decays are isolated in data collected at the LHCb experiment in the $\sqrt{s} = 7\,\text{TeV}$ pp collisions at the Large Hadron Collider, corresponding to $1\,\text{fb}^{-1}$ of integrated luminosity.

An angular analysis of the decays is performed to extract various observables. There is no significant deviation from the SM prediction in any of the measurements. The lack of new physics contributions to the $B^0 \to K^{*0}\mu^+\mu^-$ decay indicates either that the new physics mass scale is significantly higher than $\sim 1\,\text{TeV}$, or that the flavour couplings of the new physics particles are highly suppressed. These results impose stringent, model-independent constraints on new physics particles at the $\sim$ TeV mass scale.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Standard Model (SM) of particle physics is an effective theory that describes our current understanding of fundamental particles and their interactions. The SM is experimentally validated to incredible precision. For example, the decay widths of the Z^0 and W gauge bosons are predicted to be

$$\Gamma_{Z^0}^{\mathrm{SM}} \approx 2.4960 \pm 0.0002\,\mathrm{GeV}, \qquad \Gamma_{W}^{\mathrm{SM}} \approx 2.0915 \pm 0.0005\,\mathrm{GeV}, \tag{1.0.1}$$

and are measured to be

$$\Gamma_{Z^0} \approx 2.4952 \pm 0.0023\,\mathrm{GeV}, \qquad \Gamma_{W} \approx 2.085 \pm 0.042\,\mathrm{GeV}\ [1]. \tag{1.0.2}$$

Despite the excellent consistency of experimental measurements and SM predictions across a wide variety of precision measurements, the SM has a number of problems. For example, cold dark matter (see Refs. [2, 3]) is known to make up about 27% of the energy density in the universe [4], but there is no cold dark matter candidate in the SM. Furthermore, the amount of charge-parity (CP) violation predicted by the SM falls short of that required to generate the matter-antimatter asymmetry observed in the Universe by about $\sim$10 orders of magnitude [5, 6]. There is also no description of gravity in the SM.

A more subtle problem of the SM is the low mass scale of the Higgs boson. The Higgs boson has recently been discovered by the ATLAS [7] and CMS [8] collaborations with a mass of $\sim 125\,\mathrm{GeV}/c^2$ [9,10]. The loop corrections to the Higgs mass in the SM are on the order of the Planck scale, at $\sim 10^{19}\,\mathrm{GeV}$. To obtain a Higgs mass at the level of the W and Z^0 masses therefore requires a cancellation in the theory of $\sim 10^{17}$ orders of magnitude. Such a precise cancellation is considered unnaturally 'fine tuned'. A mechanism that 'protects' the Higgs mass from the loop corrections is needed to reduce the size of the cancellation that is required.

Many theories of physics beyond the SM introduce additional loop corrections from New Physics (NP) particles that contribute with the opposite sign to those of the SM. These contributions cancel the vast majority of the loop corrections from the SM particles, leaving the mass of the Higgs at the same scale as the W and Z^0 bosons without excessive fine tuning. New particles are expected to appear at the $\sim 1\,\text{TeV}$ scale so that the Higgs mass is protected. Direct searches for these NP particles have been performed by the ATLAS and CMS collaborations at the Large Hadron Collider (LHC) [11]. These collaborations aim to produce and detect these NP particles in the high energy pp collisions at the LHC. No evidence for NP particles has yet been observed.

In the same way that NP particles contribute to the Higgs mass, they will also contribute to all SM processes through loop corrections. The NP particles can therefore be searched for 'indirectly' by observing the effect of NP particles on SM processes. Flavour-Changing Neutral Current (FCNC) processes in the SM are particularly interesting, as these decays only occur through loop diagrams. As the intermediate particles are virtual in such loops, the contribution of NP particles can be as large as that of SM particles. Flavour-changing decays are particularly interesting as the flavour changing structure in the SM is intimately related to the Higgs mechanism.

Decays of B mesons mediated by the FCNC $b \to s$ quark transition have already been observed. The B meson has the advantage of a clear experimental signature in detectors due to the relatively long lifetime of $\sim 1.5 \times 10^{-12}\,\text{s}$. A number of simplifying approximations can also be made for heavy meson decays, which aids the calculation of theoretical predictions in both the SM and in NP models. Furthermore, 'observables' can be formed specifically so that the leading theoretical uncertainties cancel, allowing precise theoretical predictions to be made. The $B^0 \to K^{*0}\mu^+\mu^-$ decay offers a large number of such observables, including the branching fraction and a variety of quantities related to the angular distribution of the decay products. The $B^0 \to K^{*0}\mu^+\mu^-$ decay is therefore a critical channel in the search for NP particles at the $\sim$ TeV scale. Any discrepancy between the measured value of the observables and the SM predictions would indicate NP contributions to the decay. If no discrepancy is observed, then the measurements impose model independent constraints on NP. Moreover, the different $B^0 \to K^{*0}\mu^+\mu^-$ observables are sensitive to different underlying processes. The measurements therefore constrain not only the size of the contribution from NP particles, but also the underlying gauge structure of their interactions.

The LHCb detector at the LHC is specialised for precise studies of B meson decays [12]. Isolating a sample of $B^0 \to K^{*0}\mu^+\mu^-$ decays from the vast quantity of pp interactions at the LHC requires a highly selective and efficient trigger. Multivariate analyses can then be used to isolate $B^0 \to K^{*0}\mu^+\mu^-$ candidates from the huge number of back-

ground processes. The LHCb detector, the trigger, and the selection of $B^0 \to K^{*0}\mu^+\mu^-$ decays each modify the angular distribution of the decay products and therefore alter the angular observables of interest. The correction of such 'acceptance effects' is required so that the angular observables can be accurately measured. The branching fraction and angular observables can then be determined, and the implications of the measurements for NP models analysed.

In Sec. 2 of this thesis, the theoretical framework of flavour changing processes in the SM is outlined. The theoretical description of the $B^0 \to K^{*0}\mu^+\mu^-$ decay is described, the $B^0 \to K^{*0}\mu^+\mu^-$ observables are defined, and the status of experimental measurements of the observables are discussed. In Sec. 3 the LHC is briefly introduced, and the LHCb detector and trigger system are described. In Sec. 4 the selection of $B^0 \to K^{*0}\mu^+\mu^-$ candidates is discussed. In Sec. 5 the effect of the LHCb detector, trigger and selection on the angular distribution of the $B^0 \to K^{*0}\mu^+\mu^-$ decay products is analysed, and a model independent correction procedure is defined. In Sec. 6 and Sec. 7 measurements of the $B^0 \to K^{*0}\mu^+\mu^-$ differential branching fraction and angular observables are described. In Sec. 8 the results of the measurements are put into a wider context, both in terms of the current status of experimental measurements and in terms of the implications of the measurements for NP models. Finally, conclusions are presented in Sec. 9.

Chapter 2

Theory of $B^0 \to K^{*0}\mu^+\mu^-$

In the SM 'flavour changing' transitions, such as the FCNC $B^0 \to K^{*0}\mu^+\mu^-$ decay, are related to the CKM matrix, which appears due to Yukawa interactions. These Yukawa interactions arise as a result of the Higgs mechanism and are responsible for generating quark masses in the SM.

The $B^0 \to K^{*0}\mu^+\mu^-$ decay is described theoretically using an effective Hamiltonian, which expresses the decay in terms of effective vertices and couplings. The couplings in the effective theory can be precisely predicted in the SM and NP models, and can be related to observables such as the differential branching fraction. There are also a variety of 'angular observables' that offer excellent sensitivity to NP contributions via the $B^0 \to K^{*0}\mu^+\mu^-$ differential decay rate.

In Sec. 2.1 the relationship between the Higgs mechanism and flavour-changing processes in the SM is outlined. In Sec. 2.2 the effective-theory framework is described, and the application of the effective theory to the $B^0 \to K^{*0}\mu^+\mu^-$ decay is discussed in Sec. 2.3. The techniques used to compute predictions for the $B^0 \to K^{*0}\mu^+\mu^-$ observables are outlined in Sec. 2.4. The (angular) differential decay rate of the $B^0 \to K^{*0}\mu^+\mu^-$ decay is expressed in terms of the $B^0 \to K^{*0}\mu^+\mu^-$ angular observables in Sec. 2.5. Finally, the status of experimental measurements of the $B^0 \to K^{*0}\mu^+\mu^-$ observables is presented in Sec. 2.6.

2.1 The Higgs mechanism and flavour violation

Flavour-changing processes arise in the SM as a direct result of the Higgs mechanism and the generation of the quark masses[1]. The Higgs boson is added to the SM to generate masses for the weak gauge bosons without violating local gauge invariance.

[1]This section was written with reference to Refs. [13–15]

The Higgs mechanism in the SM is a single complex doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{2.1.1}$$

where ϕ^+ and ϕ^0 are complex scalar fields. The Higgs Lagrangian is

$$\mathcal{L}_{\text{Higgs}} = -(D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi) + \mathcal{L}_{\text{Y}}, \tag{2.1.2}$$

where the first term is the kinetic term for a free scalar particle, $V(\phi^\dagger\phi)$ is the Higgs potential and $\mathcal{L}_{\text{Y}}$ is a Lagrangian containing Yukawa interaction terms of the Higgs field and the fermion fields. The Higgs potential is chosen to be that of a general self-interacting field

$$V(\phi^\dagger \phi) = \lambda \left(\phi^\dagger \phi\right)^2 - \mu^2 \phi^\dagger \phi + \frac{\mu^4}{4\lambda}, \tag{2.1.3}$$

which, for $\mu^2 > 0$, gives a non-zero minimum at

$$|\phi| = \sqrt{\frac{\mu^2}{2\lambda}} = \frac{1}{\sqrt{2}}\nu. \tag{2.1.4}$$

By picking a certain direction of the minima,

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix}, \tag{2.1.5}$$

the electroweak symmetry is broken. In this 'unitary gauge', fluctuations $H(\vec{x})$ around the minimum lead to

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ (\nu + H) \end{pmatrix}, \tag{2.1.6}$$

where H is a real field: the Higgs boson. The ν parameter is a real constant that minimises the scalar potential that has been measured to be $\nu = 246\,\text{GeV}$ [14]. This quantity is referred to as the Higgs 'vacuum expectation value', and defines the 'electroweak scale' to be at the $\sim$100 GeV scale. The complex conjugate of ϕ is defined as

$$\phi_C = i\tau_2 \phi^* = \frac{1}{\sqrt{2}} \begin{pmatrix} (\nu + H) \\ 0 \end{pmatrix}, \tag{2.1.7}$$

where τ_2 is the second Pauli spin matrix. This form arises due to the transformation properties of ϕ under the symmetries of the SM.

A Dirac mass term couples left- and right-handed components of the field. For a particle X, the Dirac mass term is expressed as

$$\mathcal{L}_{\text{Dirac}} = -m_X \bar{X} X = -m_X \left(\bar{X}_L X_R + \bar{X}_R X_L \right). \tag{2.1.8}$$

Such terms break the symmetries of the SM and therefore Dirac fermion masses cannot be added directly to the SM Lagrangian. Conversely, Yukawa interactions do not break the symmetries of the SM. The Yukawa term for the electron is

$$\mathcal{L}^e_{\text{Y}} = -g_e \left(\bar{e}_L \phi e_R + \bar{e}_R \phi^\dagger e_L \right), \tag{2.1.9}$$

where g_e is the Yukawa coupling strength for the electron, $e_{L,R}$ are the left- and right-handed components of the electron field and ϕ is the Higgs doublet. Using the form of the Higgs field in the unitary gauge, see Eq. (2.1.6) and Eq. (2.1.7), $\mathcal{L}^e_{\text{Y}}$ is written as

$$\mathcal{L}^e_{\text{Y}} = -\frac{1}{\sqrt{2}} g_e \nu (\bar{e}_L e_R + \bar{e}_R e_L), \tag{2.1.10}$$

$$= -\frac{1}{\sqrt{2}} g_e \nu \bar{e} e. \tag{2.1.11}$$

which is a Dirac mass term for the electron with a mass of

$$m_e = \frac{1}{\sqrt{2}} g_e \nu. \tag{2.1.12}$$

The Yukawa interaction terms allow Dirac mass terms for the fermions to be introduced to the SM Lagrangian. For a given ν, the Yukawa coupling strength g_e determines the mass of the electron m_e.

The Yukawa terms for the quarks are

$$\mathcal{L}^q_{\text{Y}} = \left(a_{ij} \bar{q}_{Li} \phi_C u_{Rj} + b_{ij} \bar{q}_{Li} \phi d_{Rj} + \text{h.c.} \right), \tag{2.1.13}$$

where the indices i and j run over the three quark generations and 'h.c.' indicates the Hermitian conjugate. The matrices a_{ij} and b_{ij} are the Yukawa coupling strengths for

each quark generation, and

$$q_{Li} = \begin{pmatrix} u_{Li} \\ d_{Li} \end{pmatrix}, \qquad u_{Ri} = \begin{pmatrix} u_{Ri} \\ 0 \end{pmatrix}, \qquad d_{Ri} = \begin{pmatrix} 0 \\ d_{Ri} \end{pmatrix}. \tag{2.1.14}$$

In the unitary gauge, Eq. (2.1.13) is

$$\mathcal{L}_{\mathrm{Y}}^{q} = -\left(1 + \frac{H}{\nu}\right)\left(\bar{u}_{Li} m_{ij}^{u} u_{Rj} + \bar{d}_{Li} m_{ij}^{d} d_{Rj} + \text{h.c.}\right), \tag{2.1.15}$$

where

$$m_{ij}^{u} = -\frac{1}{\sqrt{2}}\nu a_{ij}, \qquad m_{ij}^{d} = -\frac{1}{\sqrt{2}}\nu b_{ij}. \tag{2.1.16}$$

Terms proportional to H in Eq. (2.1.15) are Higgs interaction terms that are ignored hereafter.

The matrices $m_{ij}^{u,d}$ are not in general diagonal. Four separate rotations are required to diagonalise these matrices

$$\begin{aligned} u_{L\alpha} &= (U_L^u)_{\alpha i} u_{Li}, & u_{R\alpha} &= (U_R^u)_{\alpha i} u_{Ri}, \\ d_{L\alpha} &= (U_L^d)_{\alpha i} d_{Li}, & d_{R\alpha} &= (U_R^d)_{\alpha i} d_{Ri}, \end{aligned} \tag{2.1.17}$$

where each rotation matrix U is unitary and α runs over the mass eigenstates of the quarks

$$\begin{aligned} u_\alpha &= \{u, c, t\}, \\ d_\alpha &= \{d, s, b\}. \end{aligned} \tag{2.1.18}$$

The diagonal mass matrices are therefore defined as

$$\begin{aligned} m_\alpha^u &= (U_L^{u\dagger})_{i\alpha} m_{ij}^u (U_R^u)_{\alpha j}, \\ m_\alpha^d &= (U_L^{d\dagger})_{i\alpha} m_{ij}^d (U_R^d)_{\alpha j}, \end{aligned} \tag{2.1.19}$$

and under such rotations

$$\mathcal{L}_{\rm Y}^{q} = -\left[\bar{u}_{L\alpha}(U_L^{u\dagger})_{i\alpha}m_{ij}^{u}(U_R^{u})_{\alpha j}u_{R\alpha} + \bar{d}_{L\alpha}(U_L^{d\dagger})_{i\alpha}m_{ij}^{d}(U_R^{d})_{\alpha j}d_{R\alpha} + \text{h.c.}\right] \tag{2.1.20}$$

$$= -\left[m_{\alpha}^{u}\bar{u}_{L\alpha}u_{R\alpha} + m_{\alpha}^{d}\bar{d}_{L\alpha}d_{R\alpha} + \text{h.c.}\right] \tag{2.1.21}$$

$$= -\left[m_{\alpha}^{u}\bar{u}_{L\alpha}u_{R\alpha} + m_{\alpha}^{u}\bar{u}_{R\alpha}u_{L\alpha} + m_{\alpha}^{d}\bar{d}_{L\alpha}d_{R\alpha} + m_{\alpha}^{d}\bar{d}_{R\alpha}d_{L\alpha}\right] \tag{2.1.22}$$

$$= -\left[m_{\alpha}^{u}\bar{u}_{\alpha}u_{\alpha} + m_{\alpha}^{d}\bar{d}_{\alpha}d_{\alpha}\right], \tag{2.1.23}$$

which generates Dirac mass terms with masses

$$m_{\alpha}^{u} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix}, \qquad m_{\alpha}^{d} = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix}. \tag{2.1.24}$$

The quark masses are therefore related to the Yukawa matrices in Eq. (2.1.16) and the rotations in Eq. (2.1.17). To obtain the spectrum of quark masses that is observed in nature, the Yukawa couplings must be

$$\begin{aligned} a_u &= 2\times10^{-5}, & b_d &= 4\times10^{-5}, \\ a_c &= 9\times10^{-3}, & b_s &= 8\times10^{-4}, \\ a_t &= 1, & b_b &= 3\times10^{-2}. \end{aligned} \tag{2.1.25}$$

There is no explanation for the wide range of Yukawa coupling strengths in the SM.

The rotation matrices $U_{L,R}^{u,d}$ also appear in the weak interactions of quarks. The interaction Lagrangian for charged-current quark interactions in the generation basis is

$$\mathcal{L}_{\rm CC} = \frac{ig_2}{\sqrt{2}}\left[W_{\mu}^{+}\bar{u}_{Lj}\gamma^{\mu}d_{Lj} + W_{\mu}^{-}\bar{d}_{Lj}\gamma^{\mu}u_{Lj}\right]. \tag{2.1.26}$$

In the mass basis this is

$$\mathcal{L}_{\rm CC} = \frac{ig_2}{\sqrt{2}}\left[W_{\mu}^{+}\bar{u}_{L\alpha}\left[(U_L^{u})_{\alpha j}(U_L^{d\dagger})_{j\beta}\right]\gamma^{\mu}d_{L\beta} + W_{\mu}^{-}\bar{d}_{L\alpha}\left[(U_L^{d})_{\alpha j}(U_L^{u\dagger})_{j\beta}\right]\gamma^{\mu}u_{L\beta}\right], \tag{2.1.27}$$

where

$$V_{\alpha\beta} \equiv \left[U_L^{u}U_L^{d\dagger}\right]_{\alpha\beta}, \tag{2.1.28}$$

is the CKM matrix. This can be interpreted as the difference between the rotations

required to diagonalise the up and down quark mass matrices, or equivalently the misalignment of the up- and down-quark mass bases. The CKM matrix relates the generation and mass eigenstates as

$$\begin{pmatrix} d_{L1} \\ d_{L2} \\ d_{L3} \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d_{L\alpha=d} \\ d_{L\alpha=s} \\ d_{L\alpha=b} \end{pmatrix}. \tag{2.1.29}$$

The values of the CKM elements have a hierarchical structure, which is often expressed in terms of the expansion parameter $\lambda \approx 0.24$ as

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1 & \lambda & 0 \\ -\lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \mathcal{O}(\lambda^2). \tag{2.1.30}$$

Flavour-changing processes in the SM are therefore related to the two Yukawa matrices a_{ij} and b_{ij} and the CKM matrix. Such interactions violate the flavour symmetry of the SM, and are therefore the source of 'flavour violation' in the SM. Many NP models predict new particles that contribute to, and therefore modify, flavour-changing interactions of SM particles. Flavour observables, such as those of the $B^0 \to K^{*0}\mu^+\mu^-$ decay, are sensitive to the contributions of NP particles. By confronting SM predictions with experimental measurements the flavour structure of NP models can be constrained. Such constraints are typically discussed in the framework of the effective theories parameterised in terms of an Operator Product Expansion (OPE).

2.2 The Operator Product Expansion

The OPE defines an effective Hamiltonian in which short- and long-distance effects are separated at an energy scale μ. Short distance (high energy) effects are parameterised in terms of Wilson coefficients, $\mathcal{C}$, which are perturbatively calculable in the SM and in NP models. Long distance (low energy) effects are parameterised in terms of Wilson operators, $\mathcal{O}$, which incorporate the effects of the strong interaction and are therefore non-perturbative and difficult to calculate. Each Wilson operator $\mathcal{O}$ corresponds to a process through which a given decay is mediated. The classic example of an effective theory is the Fermi theory of weak interactions, which describes β decay in terms of a four-fermion interaction. The effect of the W boson is encoded in the value of $G_{\rm F}$, which

is equivalent to the Wilson coefficient for the process.

The OPE is a summation over all contributing operators, with the contribution of each operator given by the corresponding Wilson coefficient. Any particle that contributes to the decay and has a mass greater than the scale μ will affect the value of (at least one of) the Wilson coefficients. The size of the contribution is dependent on the mass of the particle and the strength of the coupling. The OPE is written as

$$\mathcal{H}_{\text{eff}} = -\sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i \propto -\sum_i \mathcal{C}_i\mathcal{O}_i \tag{2.2.1}$$

where i runs over the different Wilson operators. The factor c_i is the coupling strength for the operator $\mathcal{O}_i$, and Λ is the mass scale of any contributing (NP) particles. The Wilson coefficients $\mathcal{C}_i$ are related to the factor c_i/Λ^2 by a normalisation factor.

The Wilson coefficients can be calculated precisely in the SM and in NP models. Unfortunately it is difficult to measure the Wilson coefficients experimentally. The properties of a B meson decay that are relatively simple to measure, such as the branching fraction, are dependent not on just on the Wilson coefficients but also on non-perturbative hadronic processes that affect the B decay. The theoretical prediction of these properties is therefore complicated by the modelling of these processes, resulting in large theoretical uncertainties. The solution is to form ratios of decay properties that are experimentally accessible and in which the hadronic uncertainties cancel. Experimental measurements of such observables can then confront the precise theoretical predictions computed in the SM and NP models, to determine or set constraints on the size of any NP contributions.

2.2.1 The flavour problem and Minimal Flavour Violation

Typically NP models introduce additional flavour-changing processes. The most general form of NP flavour-changing processes is to have generic couplings of order unity. Experimental measurements of observables in flavour-changing interactions of SM particles are sensitive to the contribution of NP particles, and can therefore impose constraints on the flavour structure of any contributing NP particle.

Processes in which two quarks change flavour are described in terms of $\Delta_F = 2$ Wilson operators. In Ref. [16], measurements of flavour observables are used to set constraints on NP contributions to $\Delta_F = 2$ operators. The effective Hamiltonian, including only the $\Delta_F = 2$ operators responsible for meson oscillations, is

$$\mathcal{H}_{\text{eff}} = -\sum_{i\neq j} \frac{c_{ij}}{\Lambda^2}\left(\bar{\mathcal{O}}_{Li}\gamma^\mu\mathcal{O}_{Lj}\right)^2, \tag{2.2.2}$$

where c_{ij} is the coupling of the two quarks in the meson, for example, c_{bs} for B_s^0 oscillations. The analysis of Ref. [16] concludes that the magnitude of any NP contribution cannot exceed that of the SM. Given this condition, the constraints at 95% C.L. on the scale Λ are

$$\Lambda > \frac{4.4\,\text{TeV}}{|V_{ti}^* V_{tj}|/|c_{ij}|^{1/2}} \sim \begin{cases} 1.3\times10^4\,\text{TeV}\times|c_{sd}|^{1/2} \\ 5.1\times10^2\,\text{TeV}\times|c_{bd}|^{1/2} \\ 1.1\times10^2\,\text{TeV}\times|c_{bs}|^{1/2} \end{cases} \quad [16]. \tag{2.2.3}$$

For generic couplings of order unity, $c_{ij} = 1$, the analysis of Ref. [16] determines that $\Lambda \gtrsim 100\,\text{TeV}$. For NP particles with $\Lambda = 1\,\text{TeV}$, the coupling strength of the NP particles must be $c_{ij} < 10^{-5}$. These constraints can be interpreted in two ways: either the scale of NP particles contributing to $\Delta_F = 2$ operators is $\gtrsim 100\,\text{TeV}$; or the NP $\Delta_F = 2$ processes are highly suppressed [16]. This is a statement of the so-called *flavour problem*: for NP particles to exist at the $\sim$ TeV mass scale, NP must have a highly non-generic flavour structure [16].

As the presence of NP particles at the $\sim 1\,\text{TeV}$ scale is well motivated, in order to satisfy the experimental constraints, a given NP model therefore needs to explain how the suppression of flavour-changing processes arises. The largest (or least suppressed) NP contributions that are consistent with experimental measurements are obtained by postulating that NP flavour-changing processes are dependent on a hierarchical structure similar to the CKM matrix. Such a structure, with the CP-violating phase 'aligned' to that of the CKM matrix, ensures that NP processes do not cause observable deviations in measurements of the CKM matrix. The current experimental measurements of the CKM matrix are precise, measuring two of the CKM angles to $\sim 4\%$ [1], and are in good agreement with the unitarity condition(s) expected in the SM [17, 18].

This conjecture leads to the concept of Minimal Flavour Violation (MFV), which dictates that flavour- and CP-violation in the NP sector is related to the SM Yukawa matrices a_{ij} and b_{ij} and therefore to the CKM matrix [19]. The flavour-changing couplings in NP are then suppressed with the same hierarchical structure as in the SM. This allows NP contributions to be suppressed to a level that is consistent with experimental measurements even with $\Lambda \sim$ TeV.

Processes in which only one quark changes flavour, such as the $b \to s$ transition that mediate the $B^0 \to K^{*0}\mu^+\mu^-$ decay, are described in terms of $\Delta_F = 1$ operators. Ref. [16] shows that existing constraints on $\Delta_F = 1$ operators require that $\Lambda \gtrsim 1.5\,\text{TeV}$ at 95% C.L. with order unity couplings. This means that there could be $\Delta_F = 1$ flavour-changing processes mediated by NP particles at the $\sim$ TeV mass scale. It is therefore interesting to measure $\Delta_F = 1$ operators with higher precision, such that more stringent constraints,

or evidence of NP contributions, can be obtained.

2.3 Wilson operators for $B^0 \to K^{*0}\mu^+\mu^-$

There are twelve long-distance $\Delta_F = 1$ operators that contribute to the $B^0 \to K^{*0}\mu^+\mu^-$ decay, denoted $\mathcal{O}_{1\cdots10,S,P}$. Many of these operators are suppressed, or are experimentally constrained, and are not interesting when searching for NP contributions to the decay [20]. For example, $\mathcal{O}_1$ and $\mathcal{O}_2$ relate only to diagrams with internal u and c quarks, which are either doubly-Cabbibo suppressed or contribute at a level of 0.1% of those diagrams with an internal t quark [21]. The $\mathcal{O}_{S,P}$ operators correspond to diagrams with internal scalar and pseudo-scalar particles that are also extremely suppressed [20].

Ultimately the operators $\mathcal{O}_{7,9,10}$, and their right-handed partners $\mathcal{O}'_{7,9,10}$, offer the greatest sensitivity to NP in the $B^0 \to K^{*0}\mu^+\mu^-$ decay. The effective Hamiltonian is written in terms of those operators as

$$\mathcal{H}_{\text{eff}} = -\frac{4G_{\text{F}}}{\sqrt{2}}\frac{e^2}{16\pi^2}V_{tb}V_{ts}^* \sum_{i=7,9,10}(\mathcal{C}_i\mathcal{O}_i + \mathcal{C}'_i\mathcal{O}'_i) + \text{h.c.} \tag{2.3.1}$$

The operators are defined as:

$$\mathcal{O}_7 = \frac{e}{g^2}\bar{\text{m}}_b(\bar{s}\sigma_{\mu\nu}P_R b)F^{\mu\nu}, \qquad \mathcal{O}'_7 = \frac{e}{g^2}\bar{\text{m}}_b(\bar{s}\sigma_{\mu\nu}P_L b)F^{\mu\nu}, \tag{2.3.2}$$

$$\mathcal{O}_9 = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu \ell), \qquad \mathcal{O}'_9 = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_R b)(\bar{\ell}\gamma^\mu \ell), \tag{2.3.3}$$

$$\mathcal{O}_{10} = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu\gamma_5 \ell), \qquad \mathcal{O}'_{10} = \frac{e^2}{g^2}(\bar{s}\gamma_\mu P_R b)(\bar{\ell}\gamma^\mu\gamma_5 \ell), \tag{2.3.4}$$

where e is the electro-magnetic coupling strength, g is the weak coupling strength, $\sigma_{\mu\nu}$ are the Pauli spin matrices, $P_{L,R}$ are the left- and right-isospin projection operators, and $F^{\mu\nu}$ electromagnetic field tensor. Each of these operators corresponds to either a penguin- or box-type diagram in the SM. The operator $\mathcal{O}_7$ relates to penguin diagrams that are mediated via a photon, which subsequently decays into a $\mu^+\mu^-$ pair. The operators $\mathcal{O}_9$ and $\mathcal{O}_{10}$ correspond to vector and axial-vector mediated processes respectively. These operators relate to penguin diagrams with a Z^0 boson, which subsequently decays into a $\mu^+\mu^-$ pair, and W mediated box diagrams. The dominant SM Feynman diagrams for the $B^0 \to K^{*0}\mu^+\mu^-$ decay are shown in Fig. 2.1.

In general there are contributions to the Wilson coefficients from the SM and from

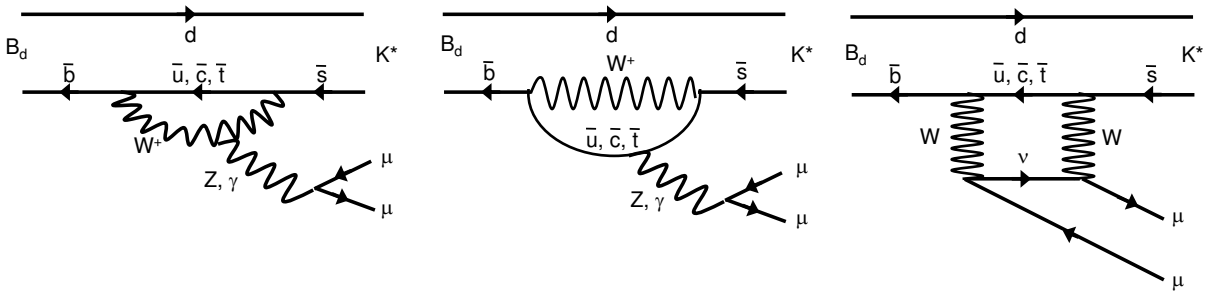


Figure 2.1: The dominant SM Feynman diagrams for the $B^0 \to K^{*0}\mu^+\mu^-$ decay. Taken from Ref. [22]

NP, so that

$$\mathcal{C}_i = \mathcal{C}_i^{\mathrm{SM}} + \mathcal{C}_i^{\mathrm{NP}}, \qquad \mathcal{C}'_i = \mathcal{C}'^{\mathrm{SM}}_i + \mathcal{C}'^{\mathrm{NP}}_i. \tag{2.3.5}$$

There are no right-handed interactions for the $i = 9, 10$ operators in the SM, so $\mathcal{C}'^{\mathrm{SM}}_{9,10} = 0$. Conversely, there are helicity-suppressed right-handed SM contributions to $\mathcal{O}'_7$ so that

$$\mathcal{C}'_7 = \frac{m_s}{\mathrm{m}_b}(\mathcal{C}_7^{\mathrm{SM}} + \mathcal{C}_7^{\mathrm{NP}}) + \mathcal{C}'^{\mathrm{NP}}_7 \text{ [23]}. \tag{2.3.6}$$

Wilson coefficients are calculated by requiring equality between each term in the effective theory and the full theoretical calculation at a 'matching scale' μ. This scale is typically the mass of the W boson. Conversely, measurements of B decays are sensitive to effects at the scale of the b quark mass m_b. The Wilson coefficients must therefore be 'evolved' down from the matching scale to the scale m_b. As the scale is reduced, effects previously encoded in the operators become contributions to the Wilson coefficients, which leads to the mixing of certain operators. The evolution therefore results in the Wilson coefficients depending on the scale μ. Effective coefficients $\mathcal{C}_i^{\mathrm{eff}}$ are defined at the scale m_b to account for the mixing of the operators during the evolution [20].

2.4 Theoretical prediction of the $B^0 \to K^{*0}\mu^+\mu^-$ decay

The prediction of $B^0 \to K^{*0}\mu^+\mu^-$ observables is complicated by hadronic contributions, particularly those involving $c\bar{c}$ loops in which the charm quarks annihilate to a di-muon final state via a virtual photon [24]. The properties of $c\bar{c}$ loops are determined by the invariant mass of the di-muon, q^2, and these properties differ markedly between three distinct regions. The 'low q^2' region is defined as $q^2 \lesssim 7\,\mathrm{GeV}^2/c^4$. The boundary arises due to the $c\bar{c}$ pair-production threshold at $q^2 \approx (2\mathrm{m}_c)^2 \approx 7\,\mathrm{GeV}^2/c^4$. In the $7 < q^2 < 15\,\mathrm{GeV}^2/c^4$

region, the J/ψ and $\psi(2S)$ di-muon resonances dominate the short-distance rate and no precise theoretical calculation is available [24]. The 'high q^2' region is defined as $q^2 \gtrsim 15\,\mathrm{GeV}^2/c^4$ and is characterised by the energy of the K^{*0} ($E_{K^{*0}}$) being less than the energy scale of QCD interactions within the meson $\Lambda_{\mathrm{QCD}} \sim 1\,\mathrm{GeV}$.

Theoretical calculations of B meson decays are simplified by working in the 'heavy quark' limit. In this limit, the b quark mass is set to infinity, and a Heavy Quark Effective Theory (HQET) can be constructed in which the heavy quark interacts only via 'soft' hadronic processes. A heavy-quark expansion in terms of $\Lambda_{\mathrm{QCD}}/\mathrm{m}_b$ then shows that the perturbatively calculable parton-level process is the leading contribution to the $B^0 \to K^{*0}\mu^+\mu^-$ decay, and the hadronic effects are relatively small. The QCD Factorisation (QCDF) technique separates the parton-level and hadronic contributions at the energy scale $\mu \sim \mathrm{m}_b$, accounting for the 'soft' non-perturbative processes using hadronic form factors [25].

At low q^2, where $E_{K^{*0}}$ is large, the energy of the light (spectator) quark in the B^0 hadron is large and so the light quark cannot be neglected. Due to the low mass of the light quark, it cannot be treated with the $\Lambda_{\mathrm{QCD}}/\mathrm{m}_b$ expansion of HQET. Furthermore, the light quark interacts not only via 'soft' hadronic processes, as in HQET, but also via so-called 'collinear' hadronic processes [1]. The hadronic interactions are handled by expanding in terms of $\Lambda_{\mathrm{QCD}}/E_{K^{*0}}$, forming the so-called Soft-Collinear Effective Theory (SCET) [26].

There are seven form factors that describe the hadronic $B \to K^*$ component of the $B^0 \to K^{*0}\mu^+\mu^-$ decay, each of which are dependent on q^2. By expanding in terms of the small ratios $\Lambda_{\mathrm{QCD}}/\mathrm{m}_b$ and $\Lambda_{\mathrm{QCD}}/E_{K^{*0}}$, the seven $B \to K^*$ form factors reduce to two calculable 'soft' form factors $\xi_\perp$ and $\xi_\parallel$ at low q^2 [27]. Corrections to these soft form factors enter at order α_s and $\Lambda_{\mathrm{QCD}}/\mathrm{m}_b$, with the corrections on the order of $\Lambda_{\mathrm{QCD}}/\mathrm{m}_b$ constituting a major source of theoretical uncertainty [23, 28].

At low q^2 hadronic contributions are systematically calculable using HQET, SCET and QCDF. There is no precise theoretical calculation for $q^2 \lesssim 1\,\mathrm{GeV}^2/c^4$, however, due to the influences of light resonances in this region and logrithmic divergeances in QCDF/SCET as $q^2 \to 0$ [29, 30]. The effects of light resonances are estimated using other techniques with larger uncertainties [30, 31]. The most precise theoretical calculations are therefore available for $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$. This q^2 range offers the most stringent comparison between SM prediction and experimental measurements.

At high q^2 the energy of the K^{*0} becomes small and the approximations in SCET become invalid. In this limit however, $c\bar{c}$ loops are short-range effects compared to the energy scale set by $E_{K^{*0}}$ and Λ_{QCD} [24] and are therefore perturbatively calculable. QCD interactions at order Λ_{QCD} are still considered as perturbations on the large scale m_b. An OPE is constructed that generates a simultaneous expansion in $\Lambda_{\mathrm{QCD}}/\mathrm{m}_b$ and $E_{K^{*0}}/\sqrt{q^2}$.

The technique used in this regime results in the calculations at low and high q^2 depending on different form factors and therefore having different theoretical uncertainties. The theoretical uncertainties at high q^2 are larger than those at low q^2.

Ref. [24] concludes that the theoretical uncertainties are under very good control in both the low q^2 ($1 \lesssim q^2 \lesssim 7\,\mathrm{GeV}^2/c^4$) and high q^2 ($q^2 \gtrsim 15\,\mathrm{GeV}^2/c^4$) regions. Moreover, it shows that the OPE used at high q^2 and QCDF used at low q^2 obtain consistent results at the cross-over of the two regimes ($q^2 \approx 15\,\mathrm{GeV}^2/c^4$).

2.5 The $B^0 \to K^{*0}\mu^+\mu^-$ differential decay rate

A $B \to VV^*$ decay such as $B^0 \to K^{*0}\mu^+\mu^-$, where $V(V^*)$ is a massive (virtual) vector particle, can be described in terms of six complex transversity amplitudes, defined as

$$\vec{A}^{\chi} = \{A_0^{\chi}, A_{\parallel}^{\chi}, A_{\perp}^{\chi}\}, \tag{2.5.1}$$

where $\chi = \{L, R\}$ and indicates the chirality to which the amplitude corresponds [20]. Each amplitude is dependent on q^2 and relates to either the longitudinal (A_0) or transverse polarisation state of the vector particles, where the transverse polarisation can either be parallel ($A_{\parallel}$) or perpendicular ($A_{\perp}$).

At low q^2 (where $E_{K^{*0}}$ is large) and to first order in $1/\mathrm{m}_b$ and α_s, the transversity amplitudes are expressed in terms of Wilson coefficients as

$$A_{\perp}^{\chi} = \quad N(1-\hat{s})\left[(\mathcal{C}_9^{\mathrm{eff}} + \mathcal{C}_9^{\prime\mathrm{eff}}) \mp (\mathcal{C}_{10} + \mathcal{C}_{10}') + \frac{2\hat{\mathrm{m}}_b}{\hat{s}}(\mathcal{C}_7^{\mathrm{eff}} + \mathcal{C}_7^{\prime\mathrm{eff}})\right]\xi_{\perp}(E_{K^{*0}}), \tag{2.5.2}$$

$$A_{\parallel}^{\chi} = -N(1-\hat{s})\left[(\mathcal{C}_9^{\mathrm{eff}} - \mathcal{C}_9^{\prime\mathrm{eff}}) \mp (\mathcal{C}_{10} - \mathcal{C}_{10}') + \frac{2\hat{\mathrm{m}}_b}{\hat{s}}(\mathcal{C}_7^{\mathrm{eff}} - \mathcal{C}_7^{\prime\mathrm{eff}})\right]\xi_{\perp}(E_{K^{*0}}), \tag{2.5.3}$$

$$A_0^{\chi} = -\frac{N(1-\hat{s})^2}{\hat{m}_{K^{*0}}\sqrt{8\hat{s}}}\left[(\mathcal{C}_9^{\mathrm{eff}} - \mathcal{C}_9^{\prime\mathrm{eff}}) \mp (\mathcal{C}_{10} - \mathcal{C}_{10}') + 2\hat{\mathrm{m}}_b(\mathcal{C}_7^{\mathrm{eff}} - \mathcal{C}_7^{\prime\mathrm{eff}})\right]\xi_{\parallel}(E_{K^{*0}}), \tag{2.5.4}$$

where $\hat{s} = q^2\mathrm{m}_B^2$, $\hat{\mathrm{m}}_i = \mathrm{m}_i/\mathrm{m}_B$ and $\xi_{\perp,\parallel}$ are the two $B \to K^*$ soft form factors [27]. The effective Wilson coefficients $\mathcal{C}_i^{\mathrm{eff}}$ are the linear combinations of $\mathcal{C}_i$ and $\mathcal{C}_{1\cdots6}$ that arise due to mixing of the Wilson operators when the scale μ is evolved. The normalisation N is given by

$$N = V_{tb}V_{ts}^*\sqrt{\frac{G_\mathrm{F}^2\alpha^2 q^2 \lambda^{1/2}}{3^{10}\pi^5 \mathrm{m}_B}}, \tag{2.5.5}$$

with $\lambda = [q^2 - (\mathrm{m}_B + \mathrm{m}_{K^{*0}})^2][q^2 - (\mathrm{m}_B - \mathrm{m}_{K^{*0}})^2]$ [32]. Similar expressions that depend on different form factors describe the transversity amplitudes at high q^2, see Ref. [28]. The

SM prediction at high q^2 is dependent on calculations of the high q^2 form factors using lattice techniques [33].

The form factors constitute the dominant source of theoretical uncertainty on the calculation of the transversity amplitudes. As the transversity amplitudes are linear in these form factors, the leading order theoretical uncertainties cancel in a ratio of transversity amplitudes. This is the guiding principle in the development of the $B^0 \to K^{*0}\mu^+\mu^-$ angular observables. Those observables in which the form factors cancel at leading order are denoted form-factor independent (FFI) observables, those in which the form factors do not cancel are denoted form-factor dependent (FFD) observables.

The $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction is one such FFD observable, which has a theoretically predicted value of $(1.19 \pm 0.39) \times 10^{-6}$ in the SM [34]. The branching fraction is defined as

$$\mathcal{B} = \int \left[1/\frac{d\Gamma}{dq^2} \right] dq^2, \tag{2.5.6}$$

where $d\Gamma/dq^2$ is the differential decay rate. The differential decay rate is expressed in terms of the transversity amplitudes as

$$\frac{d\Gamma}{dq^2} = \sum_{\chi=L,R} |A_0^\chi|^2 + |A_\parallel^\chi|^2 + |A_\perp^\chi|^2. \tag{2.5.7}$$

The branching fraction can be separated into distinct regions of q^2 so that

$$\mathcal{B} = \sum_i \frac{d\mathcal{B}(i)}{dq^2}, \tag{2.5.8}$$

where

$$\frac{d\mathcal{B}(i)}{dq^2} = \frac{1}{q^2_{\text{max};i} - q^2_{\text{min};i}} \int_{q^2_{\text{min};i}}^{q^2_{\text{max};i}} \left[1/\frac{d\Gamma(i)}{dq^2} \right] dq^2, \tag{2.5.9}$$

where $q^2_{\text{max};i}$ ($q^2_{\text{min};i}$) denotes the maximum (minimum) of the i^{th} q^2 region.

The differential branching fraction $d\mathcal{B}/dq^2$, defined in Eq. (2.5.9), is measured in Sec. 6. As the transversity amplitudes can be expressed in terms of $\mathcal{C}_{7,9,10}$ and $\mathcal{C}'_{7,9,10}$, this measurement constrains the size of NP contributions to these Wilson coefficients. Since $d\mathcal{B}/dq^2$ is dependent on the normalisation of the form factors, the precision of the theoretical prediction of $d\mathcal{B}/dq^2$ is worse than that of the angular observables discussed in the next section.

2.5.1 The angular differential decay rate

The differential decay rate of the $B^0 \to K^{*0}\mu^+\mu^-$ decay can be written in terms of the decay angles $\cos\theta_L$, $\cos\theta_K$, ϕ and q^2. In this thesis only $B^0 \to K^{*0}\mu^+\mu^-$ decays with $K^{*0} \to K^+\pi^-$ are considered. The decays $B^0 \to K^{*0}\mu^+\mu^-$ and $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ are distinguished by the charge of the kaon, with a K^+ indicating a B^0 decay and K^- indicating a $\overline{B}^0$ decay. Since the flavour of the decaying meson is identified, the angular basis can be defined such that the angular definition for the $\overline{B}^0$ decay is a *CP* transformation of that for the B^0 decay.

The angle θ_L is defined as the angle between the direction of the μ^+ (μ^-) in the dimuon rest frame and the direction of the dimuon in the B^0 ($\overline{B}^0$) rest frame. The angle θ_K is defined as the angle between the direction of the charged kaon in the K^{*0} ($\overline{K}^{*0}$) rest frame and the direction of the K^{*0} ($\overline{K}^{*0}$) in the B^0 ($\overline{B}^0$) rest frame. The angle ϕ is the angle between the plane containing the μ^+ and μ^- and the plane containing the kaon and pion from the K^{*0}. These definitions are shown in Fig. 2.2. Explicitly, $\cos\theta_L$ and $\cos\theta_K$ are defined for the B^0 decay as

$$\cos\theta_L = \left(\hat{p}^{(\mu^+\mu^-)}_{\mu^+} \cdot \hat{p}^{(B^0)}_{\mu^+\mu^-}\right) = \left(\hat{p}^{(\mu^+\mu^-)}_{\mu^+} \cdot \left(-\hat{p}^{(\mu^+\mu^-)}_{B^0}\right)\right) \ , \tag{2.5.10}$$

$$\cos\theta_K = \left(\hat{p}^{(K^{*0})}_{K^+} \cdot \hat{p}^{(B^0)}_{K^{*0}}\right) = \left(\hat{p}^{(K^{*0})}_{K^+} \cdot \left(-\hat{p}^{(K^{*0})}_{B^0}\right)\right) \ , \tag{2.5.11}$$

and for the $\overline{B}^0$ decay as

$$\cos\theta_L = \left(\hat{p}^{(\mu^+\mu^-)}_{\mu^-} \cdot \hat{p}^{(\overline{B}^0)}_{\mu^+\mu^-}\right) = \left(\hat{p}^{(\mu^+\mu^-)}_{\mu^-} \cdot \left(-\hat{p}^{(\mu^+\mu^-)}_{\overline{B}^0}\right)\right) \ , \tag{2.5.12}$$

$$\cos\theta_K = \left(\hat{p}^{(K^{*0})}_{K^-} \cdot \hat{p}^{(\overline{B}^0)}_{K^{*0}}\right) = \left(\hat{p}^{(K^{*0})}_{K^-} \cdot \left(-\hat{p}^{(K^{*0})}_{\overline{B}^0}\right)\right) . \tag{2.5.13}$$

The definition of the angle ϕ for the B^0 decay is given by

$$\cos\phi = \left(\left(\hat{p}^{(B^0)}_{\mu^+} \times \hat{p}^{(B^0)}_{\mu^-}\right) \cdot \left(\hat{p}^{(B^0)}_{K^+} \times \hat{p}^{(B^0)}_{\pi^-}\right)\right) \ , \tag{2.5.14}$$

$$\sin\phi = \left(\left(\hat{p}^{(B^0)}_{\mu^+} \times \hat{p}^{(B^0)}_{\mu^-}\right) \times \left(\hat{p}^{(B^0)}_{K^+} \times \hat{p}^{(B^0)}_{\pi^-}\right)\right) \cdot \hat{p}^{(B^0)}_{K^{*0}} \ , \tag{2.5.15}$$

and for the $\overline{B}^0$ decay

$$\cos\phi = \left(\left(\hat{p}^{(\overline{B}^0)}_{\mu^-} \times \hat{p}^{(\overline{B}^0)}_{\mu^+}\right) \cdot \left(\hat{p}^{(\overline{B}^0)}_{K^-} \times \hat{p}^{(\overline{B}^0)}_{\pi^+}\right)\right) \ , \tag{2.5.16}$$

$$\sin\phi = -\left(\left(\hat{p}_{\mu^-}^{(\overline{B}^0)} \times \hat{p}_{\mu^+}^{(\overline{B}^0)}\right) \times \left(\hat{p}_{K^-}^{(\overline{B}^0)} \times \hat{p}_{\pi^+}^{(\overline{B}^0)}\right)\right) \cdot \hat{p}_{\overline{K}^{*0}}^{(\overline{B}^0)}. \tag{2.5.17}$$

The $\hat{p}_X^{(Y)}$ are unit vectors describing the direction of a particle X in the rest frame of the system Y. In every case the particle momenta are first boosted to the B^0 (or $\overline{B}^0$) rest frame.

The differential decay rate for $B^0 \to K^{*0}\mu^+\mu^-$ is

$$\begin{aligned}
\frac{d^4\Gamma'}{dq^2 d\cos\theta_L d\cos\theta_K d\phi} = \frac{9}{32\pi}\,[\quad & J_{1s}\sin^2\theta_K + J_{1c}\cos^2\theta_K \\
+ \quad & (J_{2s}\sin^2\theta_K + J_{2c}\cos^2\theta_K)\cos 2\theta_L \\
+ \quad & J_3 \sin^2\theta_K \sin^2\theta_L \cos 2\phi \\
+ \quad & J_4 \sin 2\theta_K \sin 2\theta_L \cos\phi \\
+ \quad & J_5 \sin 2\theta_K \sin\theta_L \cos\phi \\
+ \quad & J_{6s}\sin^2\theta_K \cos\theta_L \\
+ \quad & J_7 \sin 2\theta_K \sin\theta_L \sin\phi \\
+ \quad & J_8 \sin 2\theta_K \sin 2\theta_L \sin\phi \\
+ \quad & J_9 \sin^2\theta_K \sin^2\theta_L \sin 2\phi\,]\,,
\end{aligned} \tag{2.5.18}$$

where Γ' is the rate of the $B^0 \to K^{*0}\mu^+\mu^-$ decay and the J_i are decay amplitudes.

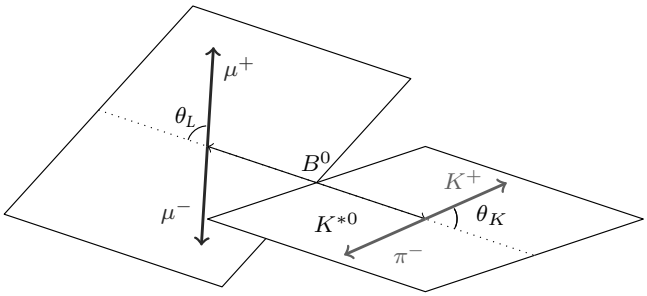
The same form is obtained for $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays with $\Gamma' \to \bar{\Gamma}'$ and $J_i \to \bar{J}_i$, where $\bar{\Gamma}'$ is the rate of the $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ decay. This assumes that there is no CP asymmetry in the $B^0 \to K^{*0}\mu^+\mu^-$ decay. The decay rate relevant to the present analysis is the CP-averaged rate

$$\Gamma = \frac{1}{2}(\Gamma' + \bar{\Gamma}'). \tag{2.5.19}$$

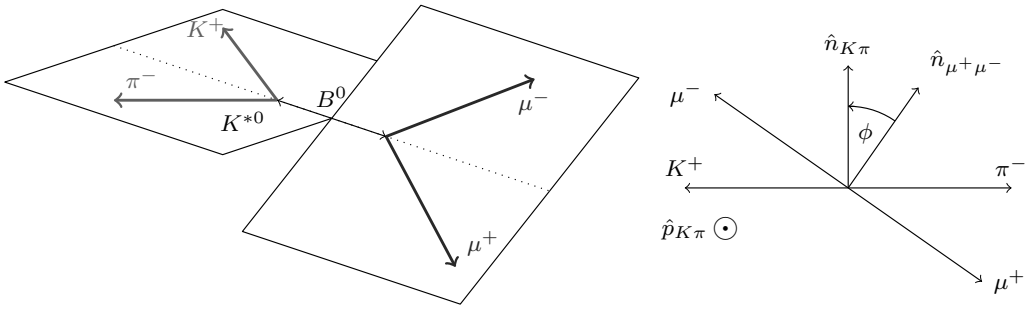
Since the flavour of the b quark is identified by the charge of the kaon, produced in the $K^{*0} \to K^+\pi^-$ decay, the angular definition which is relevant for any given $B^0 \to K^{*0}\mu^+\mu^-$ (or $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$) decay can be used. If a misidentification occurs so that the meson flavour is assigned incorrectly, then

$$\begin{aligned}
\theta_L &\to \pi - \theta_L, \\
\theta_K &\to \pi - \theta_K, \\
\phi &\to -\phi,
\end{aligned} \tag{2.5.20}$$

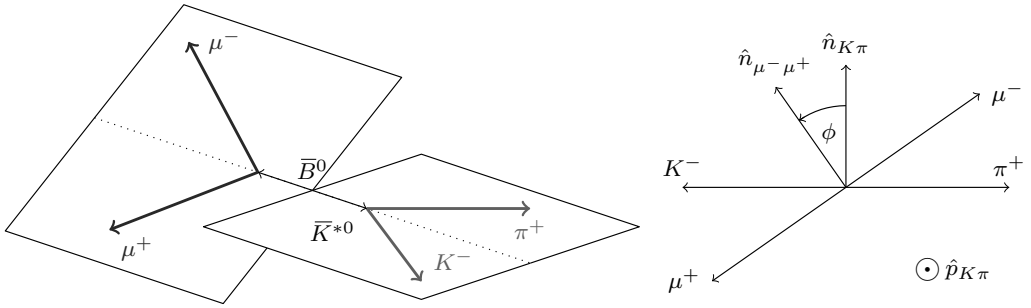
which results in a sign-flip in the J_6 and J_9 terms.



(a) θ_K and θ_L definitions for the B^0 decay



(b) ϕ definition for the B^0 decay



(c) ϕ definition for the $\overline{B}^0$ decay

Figure 2.2: Definition of the angular basis used in the present analysis. The notation $\hat{n}_{ab}$ is used to represent the normal to the plane containing particles a and b in the B^0 (or $\overline{B}^0$) rest frame. An explicit description of the angular basis is given in the text.

Due to the limited number of candidates available in the present analysis, the ϕ angle is transformed according to

$$\phi = \begin{cases} \phi + \pi & \text{if } \phi < 0 \\ \phi & \text{otherwise.} \end{cases} \tag{2.5.21}$$

This transformation results in the cancellation of the angular terms that correspond to $J_{4,5,7,8}$, which simplifies the analysis at the expense of sensitivity to these four terms. The transformation also renders $0 < \phi < \pi$ and therefore changes the normalisation of the differential decay rate from $9/32\pi$ to $9/16\pi$.

The remaining J_i terms are defined in terms of the transversity amplitudes as

$$\begin{aligned}
J_{1s} &= \frac{3}{4}\left[|A_\perp^L|^2 + |A_\parallel^L|^2 + |A_\perp^R|^2 + |A_\parallel^R|^2\right] \\
J_{1c} &= |A_0^L|^2 + |A_0^R|^2 \\
J_{2s} &= \frac{1}{4}\left[|A_\perp^L|^2 + |A_\parallel^L|^2 + |A_\perp^R|^2 + |A_\parallel^R|^2\right] \\
J_{2c} &= -\left[|A_0^L|^2 + |A_0^R|^2\right] \\
J_3 &= \frac{1}{2}\left[|A_\perp^L|^2 - |A_\parallel^L|^2 + |A_\perp^R|^2 - |A_\parallel^R|^2\right] \\
J_{6s} &= 2\left[\mathrm{Re}(A_\parallel^L A_\perp^{L*} - A_\parallel^R A_\perp^{R*})\right] \\
J_9 &= \mathrm{Im}(A_\parallel^{L*} A_\perp^L + A_\parallel^{R*} A_\perp^R).
\end{aligned} \tag{2.5.22}$$

The differential decay rate $d\Gamma/dq^2$ is related to the J_i terms by

$$\frac{d\Gamma}{dq^2} = \frac{1}{4}\left(3J_{1c} + 6J_{1s} - J_{2c} - 2J_{2s}\right). \tag{2.5.23}$$

2.5.2 The angular observables

The differential decay rate of Eq. (2.5.18) can be expressed in terms of four angular observables, at least one of which must be FFD. The angular observables are formed either as the CP-average or the CP-asymmetry of specific J_i terms. The CP-average is defined as

$$S_i = \frac{1}{2}\left[J_i + \bar{J}_i\right] = \frac{1}{2}\Sigma^{\mathrm{CP}}\left[J_i\right], \tag{2.5.24}$$

and the CP-asymmetry is defined as

$$A_i = \frac{1}{2}\left[J_i - \bar{J}_i\right] = \frac{1}{2}\Delta\left[J_i\right]. \tag{2.5.25}$$

In the present analysis five FFD angular observables

$$F_{\rm L},\ A_{\rm FB},\ S_3,\ S_9,\ A_9, \tag{2.5.26}$$

and three FFI angular observables

$$A_{\rm T}^{\rm R},\ A_{\rm T}^{2},\ A_{\rm T}^{\rm I}, \tag{2.5.27}$$

are measured. The FFD observable $F_{\rm L}$ is the CP-average of the fraction of $B^0 \to K^{*0}\mu^+\mu^-$ decays with a longitudinally polarised K^{*0}, and is expressed as

$$\begin{aligned} F_{\rm L} &= \frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[J_{1c}\right] = -\frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[J_{2c}\right], \\ &= \frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[|A_0^L|^2 + |A_0^R|^2\right]. \end{aligned} \tag{2.5.28}$$

The fraction of $B^0 \to K^{*0}\mu^+\mu^-$ decays with a transversely polarised K^{*0} is then

$$\begin{aligned} (1 - F_{\rm L}) &= \frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[\frac{4}{3}J_{1s}\right] = \frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[4J_{2s}\right], \\ &= \frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[|A_\parallel^L|^2 + |A_\perp^L|^2 + |A_\parallel^R|^2 + |A_\perp^R|^2\right]. \end{aligned} \tag{2.5.29}$$

The observables $A_{\rm FB}$, S_3 and S_9 are FFD CP-averages, while A_9 is a FFD CP-asymmetry. The observable $A_{\rm FB}$ has been measured in previous experiments and is well known in the literature, since it is highly sensitive to NP contributions [35]. In the previous measurements of $A_{\rm FB}$ by the Belle and BaBar experiments there are hints that the sign of $\mathcal{C}_7^{\rm eff}$ could be flipped ($\mathcal{C}_7^{\rm eff} \to -\mathcal{C}_7^{\rm eff}$), which, if verified, would be a clear signal for right-handed NP contributions to the $B^0 \to K^{*0}\mu^+\mu^-$ decay [36,37]. This is discussed further in Sec. 2.6. The observable $A_{\rm FB}$ is the CP-average of the forward-backward

asymmetry in $B^0 \to K^{*0}\mu^+\mu^-$ decays. This observable is expressed as

$$\begin{aligned} A_{\rm FB} &= \frac{3}{2}\frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[J_{6s}\right] \\ &= \frac{3}{2}\frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[{\rm Re}(A_\parallel^L A_\perp^{L*} - A_\parallel^R A_\perp^{R*})\right]. \end{aligned} \tag{2.5.30}$$

The S_3 and A_9 observables are highly sensitive to NP processes [38]. The observable S_3 is the CP-average of the asymmetry in the two transverse polarisation states of the K^{*0} in $B^0 \to K^{*0}\mu^+\mu^-$ decays and is particularly interesting for constraining NP models as it is sensitive only to right-handed currents [38]. It is expressed as

$$\begin{aligned} S_3 &= \frac{1}{2}\frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[J_3\right] \\ &= \frac{1}{2}\frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[|A_\perp^L|^2 - |A_\parallel^L|^2 + |A_\perp^R|^2 - |A_\parallel^R|^2\right]. \end{aligned} \tag{2.5.31}$$

The observable S_9 is the CP-average of J_9, and is expressed as

$$\begin{aligned} S_9 &= \frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[J_9\right] \\ &= \Sigma^{\rm CP}\left[{\rm Im}(A_\parallel^{L*} A_\perp^L + A_\parallel^{R*} A_\perp^R)\right]. \end{aligned} \tag{2.5.32}$$

By re-definition of the ϕ angle for $\overline{B}^0$ meson decays such that $\phi \to -\phi$, the T-odd CP-asymmetry A_9 is formed [38]. Similar to S_3, the A_9 observable is interesting for constraining NP models as it is sensitive only to right-handed currents [38]. It is expressed as

$$\begin{aligned} A_9 &= \frac{1}{d\Gamma/dq^2}\Delta^{\rm CP}\left[J_9\right] \\ &= \Delta^{\rm CP}\left[{\rm Im}(A_\parallel^{L*} A_\perp^L + A_\parallel^{R*} A_\perp^R)\right]. \end{aligned} \tag{2.5.33}$$

Assuming equal numbers of B^0 and $\overline{B}^0$ mesons, the (normalised) differential decay

rate is

$$\begin{aligned}\frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{dq^2 d\cos\theta_L d\cos\theta_K d\phi} = \frac{9}{16\pi}\Bigg[&\frac{3}{4}(1-F_{\rm L})\sin^2\theta_K + F_{\rm L}\cos^2\theta_K \\ &+ \left(\frac{1}{4}(1-F_{\rm L})\sin^2\theta_K - F_{\rm L}\cos^2\theta_K\right)\cos 2\theta_L \\ &+ S_3\sin^2\theta_K\sin^2\theta_L\cos 2\phi \\ &+ \frac{4}{3}A_{\rm FB}\sin^2\theta_K\cos\theta_L \\ &+ (S_9, A_9)\sin^2\theta_K\sin^2\theta_L\sin 2\phi\Bigg].\end{aligned} \tag{2.5.34}$$

As the observables S_9 and A_9 both depend on the J_9 term, and differ only in the definition of the ϕ angle for $\overline{B}^0$ decays, they are expressed here as (S_9, A_9).

The observables $A_{\rm FB}$, S_3 and (S_9, A_9) are 'transverse observables' as these measure the properties of decays with transversely polarised K^{*0}. As not all $B^0 \to K^{*0}\mu^+\mu^-$ decays have transversely polarised K^{*0}, these observables are implicitly dependent on $(1 - F_{\rm L})$ and are therefore FFD. This dependence is made explicit by factoring out the $(1 - F_{\rm L})$ dependence by forming a combination of an FFI transverse observable with the factor $(1 - F_{\rm L})$ in the differential decay distribution (Eq. (2.5.34)).

The FFI transverse observables are $A_{\rm T}^{\rm R}$, $A_{\rm T}^2$, and $A_{\rm T}^{\rm I}$ [27, 39]. Of these only $A_{\rm T}^2$ has been measured experimentally, and with poor precision [40]. The observable $A_{\rm T}^{\rm R}$ is the CP-average of the forward-backward asymmetry, measured only in $B^0 \to K^{*0}\mu^+\mu^-$ decays with a transversely polarised K^{*0}, and is expressed as

$$\begin{aligned} A_{\rm T}^{\rm R} &= \frac{4}{3}\frac{A_{\rm FB}}{(1-F_{\rm L})} \\ &= 2\frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[\frac{{\rm Re}(A_\parallel^L A_\perp^{L*} - A_\parallel^R A_\perp^{R*})}{|A_\parallel^L|^2 + |A_\perp^L|^2 + |A_\parallel^R|^2 + |A_\perp^R|^2}\right]. \end{aligned} \tag{2.5.35}$$

The observable $A_{\rm T}^2$ is the CP-average of the asymmetry in the two transverse polarisation states of the K^{*0}, measured only in $B^0 \to K^{*0}\mu^+\mu^-$ decays with a transversely polarised

K^{*0}, and is expressed as

$$\begin{aligned} A_{\rm T}^2 &= 2\frac{S_3}{(1-F_{\rm L})} \\ &= \frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[\frac{|A_\perp^L|^2 - |A_\parallel^L|^2 + |A_\perp^R|^2 - |A_\parallel^R|^2}{|A_\parallel^L|^2 + |A_\perp^L|^2 + |A_\parallel^R|^2 + |A_\perp^R|^2}\right]. \end{aligned} \tag{2.5.36}$$

The observable $A_{\rm T}^{\rm I}$ is the CP-average of J_9, measured only in $B^0 \to K^{*0}\mu^+\mu^-$ decays with a transversely polarised K^{*0}, and is expressed as

$$\begin{aligned} A_{\rm T}^{\rm I} &= 2\frac{S_9}{(1-F_{\rm L})} \\ &= 2\frac{1}{d\Gamma/dq^2}\Sigma^{\rm CP}\left[\frac{{\rm Im}(A_\parallel^{L*}A_\perp^L + A_\parallel^{R*}A_\perp^R)}{|A_\parallel^L|^2 + |A_\perp^L|^2 + |A_\parallel^R|^2 + |A_\perp^R|^2}\right]. \end{aligned} \tag{2.5.37}$$

These observables can be inserted into the (normalised) differential decay rate assuming equal numbers of B^0 and $\overline{B}^0$ mesons such that

$$\begin{aligned} \frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{dq^2 d\cos\theta_L d\cos\theta_K d\phi} = \frac{9}{16\pi}\Bigg[&\frac{3}{4}(1-F_{\rm L})\sin^2\theta_K + F_{\rm L}\cos^2\theta_K \\ &+ \left(\frac{1}{4}(1-F_{\rm L})\sin^2\theta_K - F_{\rm L}\cos^2\theta_K\right)\cos 2\theta_L \\ &+ \frac{1}{2}A_{\rm T}^2(1-F_{\rm L})\sin^2\theta_K\sin^2\theta_L\cos 2\phi \\ &+ A_{\rm T}^{\rm R}(1-F_{\rm L})\sin^2\theta_K\cos\theta_L \\ &+ \frac{1}{2}A_{\rm T}^{\rm I}(1-F_{\rm L})\sin^2\theta_K\sin^2\theta_L\sin 2\phi\Bigg]. \end{aligned} \tag{2.5.38}$$

As each of the angular observables are normalised by a factor $1/d\Gamma/dq^2$ the integrated differential decay rate is equal to one, and so Eq. (2.5.34) and Eq. (2.5.38) can be used as pdfs, see Sec. 7.

2.5.3 Limiting values of the angular observables

Integrating the differential decay rate, Eq. (2.5.34), gives the differential decay rate as a function of only one decay angle and q^2. For $\cos\theta_L$, this expression is

$$\frac{1}{d\Gamma/dq^2}\frac{d^2\Gamma}{d\cos\theta_L dq^2} = \frac{3}{4}F_{\rm L}\sin^2\theta_L + \frac{3}{8}(1-F_{\rm L})(1+\cos^2\theta_L) + A_{\rm FB}\cos\theta_L. \qquad (2.5.39)$$

By requiring Eq. (2.5.39) to be a valid pdf (greater than zero for all values of $\cos\theta_L$, $\cos\theta_K$ and ϕ) the expression

$$|A_{\rm FB}| \leq \frac{3}{4}(1-F_{\rm L}), \qquad (2.5.40)$$

is obtained. This indicates an $F_{\rm L}$-dependent constraint on $A_{\rm FB}$. The definition of $A_{\rm T}^{\rm R}$ is such that this constraint is always satisfied with $-1 < A_{\rm T}^{\rm R} < 1$.

Requiring that

$$\frac{1}{d\Gamma/dq^2}\frac{d^2\Gamma}{d\phi dq^2} = \frac{1}{\pi}\left[1 + F_{\rm L} + S_3\cos 2\phi + (S_9, A_9)\sin 2\phi\right], \qquad (2.5.41)$$

is a valid pdf, the expression

$$S_3\cos 2\phi + (S_9, A_9)\sin 2\phi > -(1+F_{\rm L}), \qquad (2.5.42)$$

is obtained. This expression imposes a constraint that S_3 and (S_9,A_9) cannot be large simultaneously, which becomes looser as $F_{\rm L}$ increases. These relationships between the angular observables indicate that there are regions of the observable space (*e.g.* simultaneously large $A_{\rm FB}$ and $F_{\rm L}$) in which the pdf for the differential decay rate will become negative given certain values of $\cos\theta_L$, $\cos\theta_K$ and ϕ. The impact of this on the analysis is discussed in Sec. 7.

2.5.4 Incorporating an S-wave component

About 7% of the $B^0 \to J/\psi K^{*0}$ events isolated by the BaBar experiment are produced with the K^{*0} in an S-wave state [41]. The S-wave contribution to the $B^0 \to J/\psi K^{*0}$ decay, and the potential S-wave contribution to the $B^0 \to K^{*0}\mu^+\mu^-$ decay, are therefore considered in the analysis presented in this thesis. With two contributing amplitudes, the decay rate becomes

$$\Gamma = \Gamma^0 + \Gamma^1, \qquad (2.5.43)$$

where Γ^0 is the decay rate for the S-wave component and Γ^1 is the decay rate for the P-wave component. This is expressed in terms of the relevant transversity amplitudes as

$$\frac{d\Gamma}{dq^2} = \sum_{\chi=L,R} |A^\chi_{1\,0}|^2 + |A^\chi_{1\,\parallel}|^2 + |A^\chi_{1\,\perp}|^2 + |A^\chi_{0\,0}|^2, \tag{2.5.44}$$

where $A^\chi = \{A^\chi_{1\,0}, A^\chi_{1\,\parallel}, A^\chi_{1\,\perp}\}$ are the P-wave transversity amplitudes, $A^\chi_{0\,0}$ is the S-wave transversity amplitude and $\chi = \{L, R\}$ indicates the left- or right-handed chiral state.

The S-wave state cannot be transversely polarised. However, the P- and S-wave longitudinal components $A^\chi_{1\,0}$ and $A^\chi_{0\,0}$ interfere. The differential decay rate for the combined S- and P-wave decays is modified by the introduction of two additional observables. These are the fraction of S-wave decays, expressed as

$$F_{\rm S} = \frac{1}{d\Gamma/dq^2} \sum_{\chi=L,R} |A^\chi_{0\,0}|^2, \tag{2.5.45}$$

and the interference between the longitudinally polarised P- and S-wave states, expressed as

$$A_{\rm S} = \frac{1}{d\Gamma/dq^2} \sum_{\chi=L,R} |A^\chi_{0\,0}||A^\chi_{1\,0}| \cos(\delta_{\rm S}), \tag{2.5.46}$$

where $\delta_{\rm S}$ is a strong phase that expresses the difference between the P- and S-wave propagators. The interference term $A_{\rm S}$ results in an asymmetry in the $\cos\theta_K$ distribution.

The normalisation of the existing observables also changes to account for the modified expression of Γ

$$\frac{\text{Obs}}{d\Gamma^1/dq^2} \to \frac{\text{Obs}}{d\Gamma/dq^2(1 - F_{\rm S})} \tag{2.5.47}$$

such that each angular observable acquires a factor $(1 - F_{\rm S})$. The expression for the

differential decay rate in the presence of P- and S-wave components is then

$$\begin{aligned}\frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{dq^2 d\cos\theta_L d\cos\theta_K d\phi} = \frac{9}{16\pi}\Bigg[&\left(\frac{2}{3}F_{\rm S}+\frac{4}{3}A_{\rm S}\cos\theta_K\right)\sin^2\theta_L \\ &+ (1-F_{\rm S})\Bigg(2F_{\rm L}\cos^2\theta_K\sin^2\theta_L \\ &+ \frac{1}{2}(1-F_{\rm L})\sin^2\theta_K(1+\cos^2\theta_L) \\ &+ S_3\sin^2\theta_K\sin^2\theta_L\cos 2\phi \\ &+ \frac{4}{3}A_{\rm FB}\sin^2\theta_K\cos\theta_L \\ &+ (S_9, A_9)\sin^2\theta_K\sin^2\theta_L\sin 2\phi\Bigg)\Bigg].\end{aligned} \tag{2.5.48}$$

2.6 Experimental status

Prior to the present analysis, a subset of the $B^0\to K^{*0}\mu^+\mu^-$ observables discussed above had been measured by the CDF, LHCb, Belle and BaBar collaborations. The measurements made by the different collaborations are summarised in Tab. 2.1. The first measurements were performed by the Belle collaboration. Due to the limited number of $B^0\to K^{*0}\mu^+\mu^-$ events available, the observables were extracted in six bins of q^2 covering the full kinematically-allowed $0.00<q^2<(m_B-m_{K^{*0}})^2\,{\rm GeV}^2/c^4$ range. A further q^2 bin was defined as $1.00<q^2<6.00\,{\rm GeV}^2/c^4$, which corresponds to the q^2 range in which the theoretical prediction of the observables is the most precise. For the present analysis, following the binning scheme of the LHCb collaboration in Ref. [42], fiducial requirements are imposed on q^2 such that the six q^2 bins cover the slightly reduced $0.10<q^2<19.00\,{\rm GeV}^2/c^4$ range. The seven q^2 bins used for this analysis are detailed in Tab. 2.2.

The measurements of the BaBar collaboration (in the $0.10<q^2<6.25\,{\rm GeV}^2/c^4$ range) and those of the Belle and CDF collaborations (in the $2.00<q^2<4.30\,{\rm GeV}^2/c^4$ range) favour a positive value of $A_{\rm FB}$. These measurements are consistent with a sign-flip in $\mathcal{C}_7$ due to large contributions from right-handed currents. Although this is not supported by the measurements of the LHCb collaboration (which are more precise) large right-handed currents are not excluded. There is also tension with the SM prediction in $F_{\rm L}$ at low q^2. The measurement of $F_{\rm L}$ by the BaBar collaboration (in the $0.10<q^2<6.25\,{\rm GeV}^2/c^4$ range) and those of the CDF and LHCb collaborations (in the $0.10<q^2<2.00\,{\rm GeV}^2/c^4$

Collaboration	Observables					q^2 range (GeV^2/c^4)	Reference
	$d\mathcal{B}/dq^2$	A_{FB}	F_{L}	A_{T}^2	A_9		
LHCb	✓	✓	✓			$0.10 < q^2 < 19.00$	[42]
CDF	✓	✓	✓	✓	✓	$0.00 < q^2 < 19.30$	[40, 43]
Belle	✓	✓	✓			$0.00 < q^2 < (m_B - m_{K^{*0}})^2$	[36]
BaBar	✓	✓	✓			$0.10 < q^2 < (m_B - m_{K^{*0}})^2$	[44]

Table 2.1: Experimentally measured $B^0 \to K^{*0}\mu^+\mu^-$ observables prior to the present analysis, and the corresponding collaborations. The previous LHCb measurement mentioned here used data corresponding to $\sim 0.37\,\mathrm{fb}^{-1}$ of integrated luminosity.

q^2 bin name	q^2 range (GeV^2/c^4)
$q^2_{\mathrm{ana}\,1}$	$0.10 < q^2 < 2.00$
$q^2_{\mathrm{ana}\,2}$	$2.00 < q^2 < 4.30$
$q^2_{\mathrm{ana}\,3}$	$4.30 < q^2 < 8.68$
$q^2_{\mathrm{ana}\,4}$	$10.09 < q^2 < 12.86$
$q^2_{\mathrm{ana}\,5}$	$14.18 < q^2 < 16.00$
$q^2_{\mathrm{ana}\,6}$	$16.00 < q^2 < 19.00$
$q^2_{\mathrm{ana}\,T}$	$1.00 < q^2 < 6.00$

Table 2.2: The bins of q^2 in which the $B^0 \to K^{*0}\mu^+\mu^-$ observables are extracted.

range) are lower than the values predicted in the SM. The observables S_3 and A_9, which are both highly sensitive to right-handed NP contributions, are measured only by the CDF collaboration and with relatively poor precision.

Since the $0.37\,\mathrm{fb}^{-1}$ analysis of the LHCb collaboration published in Ref. [42], the LHCb experiment has collected three times more data. This thesis presents a measurement using $1\,\mathrm{fb}^{-1}$ of integrated luminosity, which gives the world's largest sample of $B^0 \to K^{*0}\mu^+\mu^-$ decays.

Chapter 3

The LHCb detector at the Large Hadron Collider

The LHCb detector is situated at one of the four interaction points of the LHC and is specifically designed for the study of flavour physics via the decays of B hadrons. The LHC produces a large number of $b\bar{b}$ pairs and has sufficient energy to produce the full spectrum of B hadrons, making the LHC the perfect place for the study of B hadron decays. Due to the large number of pp interactions at the LHC, a highly selective trigger is required.

Flavour physics at the LHC is outlined in Sec. 3.1. The LHCb detector and trigger are detailed in Secs. 3.2 and 3.3 respectively. As simulated events are used extensively in the $B^0 \to K^{*0}\mu^+\mu^-$ analysis, the production of simulated events is outlined in Sec. 3.4.

3.1 Flavour physics at the Large Hadron Collider

The LHC is a superconducting particle accelerator that is designed to collide two beams of protons at four interaction points. The protons in the beams are accelerated and shaped into bunches using the accelerator chain at CERN. The beams are 'injected' into the LHC at 450 GeV and are then accelerated to an energy of 3.5 TeV. The bunches are brought to collision at each interaction point every 50 ns, resulting in pp interactions at a $\sqrt{s} = 7$ TeV centre of mass energy.

At LHCb there is an average of 1.4 visible interactions per bunch crossing [45], where a visible interaction is defined as one from which at least two charged particles pass through the LHCb detector. As there is a bunch crossing every 50 ns, a huge number of pp interactions occur at the LHCb interaction point every second. Each pp interaction will typically produce $\sim$80 charged particles that form a so-called primary vertex (PV) [46]. Any vertex with five or more tracks that is within 300 μm of the mean position of the

pp interaction region in the transverse plane is considered to be a PV [47]. The quality of the vertex is encapsulated in the vertex quality χ^2/dof, which is defined as the χ^2 of the hypothesis of the tracks originating from the same vertex, divided by the number of degrees of freedom of the vertex.

The cross section to produce a $b\bar{b}$ pair is around three hundred times smaller than the total pp interaction cross section. The events containing $b\bar{b}$ pairs therefore need to be isolated from a vast background of uninteresting processes in the trigger. As b quarks form B hadrons at the interaction point, the unique properties of B hadron decays can be exploited to identify events in which a $b\bar{b}$ pair was produced. Decays of B hadrons occur with a lifetime of order $\sim 10^{-12}\,\text{s}$, resulting in a displaced secondary vertex (SV) from the B hadron decay. Given the typical B hadron momentum of $\sim 100\,\text{GeV}/c$, the separation of the SV from the PV is $\sim 5\,\text{mm}$. For the purpose of identifying displaced vertices, the separation is expressed in terms of the vertex separation χ^2, which is defined as the increase in χ^2 of the PV vertex fit when the charged particles from the SV are added into the PV. The long lifetime of B hadrons also results in the decay products having a large impact parameter (IP). The IP is defined as the length of the vector from a given PV to the nearest point on the particle trajectory. The IP χ^2 is defined as the increase in χ^2 of the PV vertex fit when the track is added into the vertex. The IP quantities can also be calculated for n-track combinations, such as the K^{*0} or the B^0 meson, by using the momentum vector of the n-track combination. In events that contain multiple PVs, the IP quantities are defined with respect to the PV to which the track (or n-track combination) has the smallest IP. Due to the relatively large B hadron mass, the products of B hadron decays typically have higher transverse momentum (p_T), which is defined with respect to the beam axis, than other particles from displaced vertices.

As detailed in Sec. 3.3, the quantities defined above are exploited in the LHCb trigger to isolate events that contain B hadrons. Subsequently, these quantities are used in Sec. 4 to isolate the B hadron decay products from the other ~ 150 particles in the same event.

3.2 The LHCb detector

The LHCb detector comprises a dipole magnet, tracking stations, calorimeters and systems for particle identification (PID). The layout of the LHCb detector is shown in Fig. 3.1. The ideal orbit of the protons within the LHC defines the 'beam axis', which is taken as the z axis for the LHCb detector coordinate system. The horizontal axis is then defined as x, and the vertical axis as y. The trajectory of charged particles bends in the horizontal (x) plane due to the magnetic field, allowing the determination of the particle charge and momentum. The tracking system is spread over a 10 m baseline, and comprises

the Vertex Locator (VELO), the Tracker Turicensis (TT) and the tracking stations T1-T3. The VELO is situated around the interaction point to enhance the measurement of the IP and the vertex separation. The Ring Imaging Cherenkov (RICH) system provides separation of kaons, pions and protons. The Calorimeters help to identify, and measure the energy of, electrons, photons and hadrons. The Muon system identifies muons.

Typically the quarks in a $b\bar{b}$ pair produced at the LHC both travel in either the positive or negative z direction and approximately along the beam axis. The LHCb detector is a single-arm spectrometer with a geometric acceptance of $10 - 250\,\text{mrad}$ ($300\,\text{mrad}$) in the vertical (horizontal) plane, defined with respect to the beam axis in the positive z direction. This corresponds to a pseudo-rapidity region $2 < \eta < 5$, where η is the pseudo-rapidity, which is defined as

$$\eta = -\log\left[\tan\frac{\theta}{2}\right], \tag{3.2.1}$$

where θ is the angle between the particle momentum and the beam axis. The acceptance is larger in the horizontal plane to account for the bending of charged particle trajectories due to the dipole magnet. All four final state particles traverse the detector in approximately $\sim 16\%$ of $B^0 \to K^{*0}\mu^+\mu^-$ decays.

Each particle traversing the detector will deposit energy in the detector material. Active detector materials, such as silicon microstrips or scintillating material, are used to identify these 'hits'. The particle occupancy in the detector varies by more than two orders of magnitude between the inner- and outer-most parts of the detector. Many of the LHCb detector components are therefore more finely segmented close to the beam axis to account for the larger number of particles in this region.

The trajectory of a particle can be reconstructed by associating hits to a candidate particle, forming a 'track' through the detector. Candidate B hadrons are produced by forming n-track combinations in a given event. A candidate $B^0 \to K^{*0}\mu^+\mu^-$ decay with $K^{*0} \to K^+\pi^-$ is a 4-track combination with a K^+, π^- and two oppositely charged muons. The present analysis of $B^0 \to K^{*0}\mu^+\mu^-$ is limited to those decays with $K^{*0} \to K^+\pi^-$, as at LHCb the determination of particle trajectories, and the PID performance, is best for charged particles.

The precision with which the $B^0 \to K^{*0}\mu^+\mu^-$ observables are measured is dependent on a number of properties. The trajectory of each decay product must be precisely measured, which provides the precise determination of: the particle momenta; the B hadron properties exploited in the trigger; and the $B^0 \to K^{*0}\mu^+\mu^-$ decay angles and q^2. To correctly reconstruct a B hadron, and therefore isolate $B^0 \to K^{*0}\mu^+\mu^-$ decays from other B hadron decays, the charge and mass of each decay product must be correctly assigned. This requires robust identification of kaons, pions and muons. It is important

to suppress any spurious sources of CP asymmetry as the analysis presented in this thesis assumes zero CP asymmetry in the $B^0 \to K^{*0}\mu^+\mu^-$ decay.

The tracking and magnet systems are described in Sec. 3.2.1. The particle identification systems, including the Calorimeters, are described in Sec. 3.2.2. The overall performance of the detector is discussed in Sec. 3.2.3.

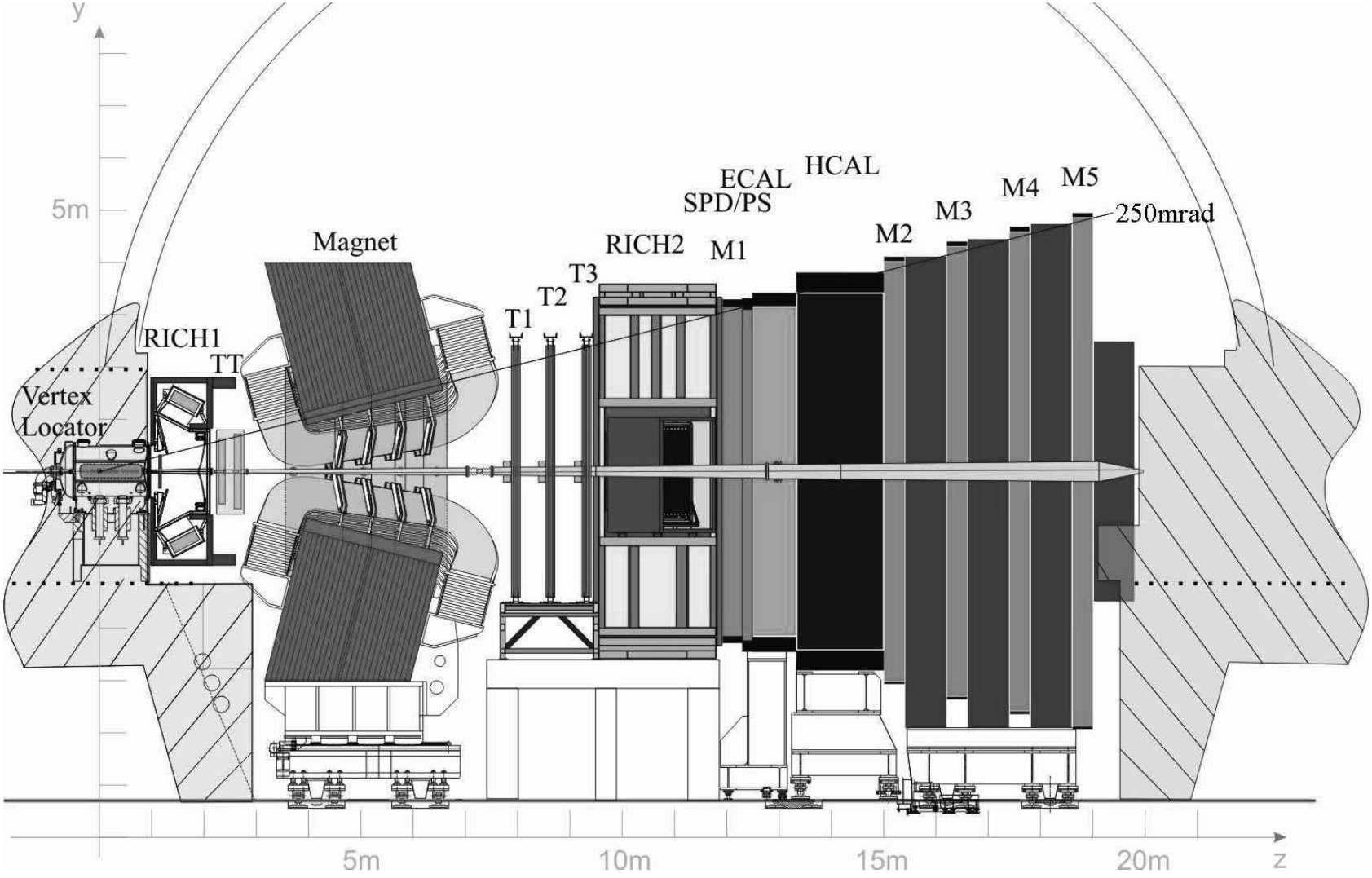


Figure 3.1: A side view of the LHCb detector. Taken from Ref. [12].

3.2.1 Tracking and magnet

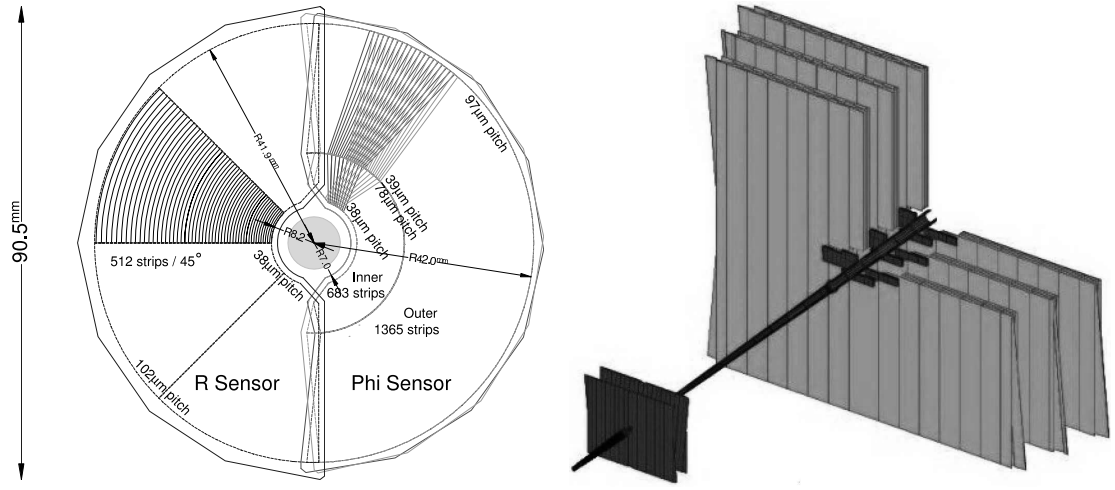
The combination of the bending of charged particle trajectories in the magnetic field with the precise measurement of the particle positions in the transverse plane provides the identification of the charge and the precise measurement of the momentum of particles traversing the detector. The LHCb dipole magnet has a strength of 4 Tm integrated over the 10 m baseline of the tracking system. The system is comprised of the VELO, located close to the interaction point, the TT, located upstream of the magnet, and the tracking stations T1-T3, located downstream of the magnet. The tracking stations T1-T3 are separated into the Inner Tracker (IT) and the Outer Tracker (OT) due to the varying particle occupancy, with the combination of the IT and OT covering the full geometric acceptance. The TT and IT are both based on silicon technologies and form the Silicon Tracker (ST).

The VELO detector is formed of two halves, each of which hold 21 modules of silicon microstrip detectors. To reduce the extrapolation from the first measurement of the particle trajectory to the vertex, and achieve the optimum measurement of particle properties, the modules of the VELO are situated only 8 mm away from the beam axis. The strip pitch varies from 40 µm to 100 µm across each module, providing a single hit resolution on the order of $\sim$10 µm. The two halves of the VELO detector overlap to ensure full azimuthal coverage, so the modules on one side of the VELO are displaced in z by 1.5 cm.

This displacement introduces a left-right asymmetry in the detector. As the dipole magnet bends particles of opposite charge in opposite directions in the horizontal plane, this asymmetry introduces a CP asymmetry in the LHCb data. To reduce the impact of this asymmetry, in approximately half of the data the polarity of the LHCb magnet is reversed.

The VELO system is based on silicon microstrip detectors. The strips are either aligned such that they emanate radially from the beam axis, in the case of a ϕ-sensor, or such that they form concentric circles around the beam axis, in the case of an r-sensor. This allows precise measurement of the particle position in both the r and ϕ coordinates. The ϕ-sensors are separated into inner- and outer-sections to reduce the particle occupancy in each strip. The ϕ-sensor strips in adjacent modules are rotated by a few degrees to obtain a stereo view of the particle trajectory. An image of a VELO sensor is shown in Fig. 3.2(a).

The ST is based on silicon microstrip detectors that have a pitch of approximately 200 µm. The TT, and each station of the IT, are formed of four detector layers in a $xuvx$ arrangement. In the u and v layers the strips are rotated by $\pm 5^\circ$ relative to the x layers, to obtain a stereo view of the particle trajectory. This affords a single-hit position



(a) VELO sensor

(b) ST and OT layout

Figure 3.2: A VELO sensor is shown with the r- and ϕ-sensor components (a). The overlap of the right and left halves can be seen. Two layers of ϕ-sensors are shown, of which the inner and outer sections, and the stereo angle between the layers, can be seen. Taken from Ref. [46]. The Silicon Tracker (dark grey) and Outer Tracker (light grey) are shown relative to the beam pipe (b). The TT can be seen in the foreground. Taken from Ref. [12].

resolution of 50 μm. In the TT, the vx layers are displaced along z from the xu layers.

The OT is a drift-time detector based on straw tube modules. The gas mixture used in the tubes gives a drift-coordinate resolution of 200 μm with a drift time of less than 50 ns. Each station contains four layers of modules, with each module containing two staggered layers of drift tubes. The four layers are situated in the same $xuvx$ arrangement as the ST. The layout of the ST and OT is shown in Fig. 3.2(b).

3.2.2 Particle identification systems

The LHCb detector contains systems that specialise in the identification of charged hadrons, muons, electrons and photons. The identification of kaons, pions and muons is critical for the study of the $B^0 \to K^{*0}\mu^+\mu^-$ decay.

3.2.2.1 The RICH system

The RICH subdetectors collect Cherenkov light, which is produced by charged particles that travel faster than light through a 'radiator' medium. Cherenkov light is produced in a cone around the direction of travel of the particle. The angle between the cone and the direction of travel (θ_C) is dependent on the speed at which the particle travels. By measuring the momentum of the particle and θ_C, the mass of the particle can be

computed.

The RICH system uses three types of radiator in two separate subdetectors, RICH1 and RICH2, each of which give optimal PID over a specific momentum range. The RICH1 subdetector is located between the VELO and TT and provides hadron identification for particles with $2 < p < 60\,\text{GeV}/c$. This is achieved by using an aerogel radiator, which gives $K - \pi$ separation in the $2 < p < 10\,\text{GeV}/c$ region, and C_4F_{10} gas, which gives $K - \pi$ separation in the remaining $10 < p < 60\,\text{GeV}/c$ range. The RICH2 subdetector uses C_4F_{10} gas to provide hadron identification for particles with $15 < p < 100\,\text{GeV}/c$. Such particles are typically found close to the beam pipe and the RICH2 subdetector is therefore situated downstream of tracking stations T1-T3 and has a reduced geometric acceptance of 15 mrad to 100 mrad (120 mrad) in the vertical (horizontal) plane.

The cones of Cherenkov light are focused using mirrors to form rings on planes of Hybrid Photon Detectors (HPDs). The HPDs convert each photon to a pulse of electrons, which is amplified and directed onto a silicon pixel detector using an electric field. The silicon pixel detector has 1024 pixels arranged in a 500 μm square, affording a granularity of 2.5 μm per pixel with a readout time of $\sim$25 ns.

The trajectory and momentum of particles are obtained by the tracking subdetectors. The position of rings of Cherenkov photons on the HPD plane can therefore be predicted given a mass hypothesis for each particle. Using a fit to the photons on the HPD plane, the difference in log-likelihood (DLL) between two mass hypotheses for a given particle is calculated. For example, the DLL between the kaon and pion hypothesis $\text{DLL}_{K\pi}$ is defined as

$$\text{DLL}_{K\pi} = \log(\mathcal{L}_K) - \log(\mathcal{L}_\pi), \tag{3.2.2}$$

such that it is positive for a kaon-like particle and negative for a pion-like particle. The DLL quantities therefore indicate the preferred mass hypothesis for each particle in the event.

3.2.2.2 The Calorimeters

The calorimetry is separated into the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL), which measure the energy and the position of any electrons, photons, and hadrons in the event. Photons and electrons cause large energy deposits in the ECAL, while hadrons cause large energy deposits in the HCAL. The ECAL is situated downstream of the RICH2 detector and the HCAL is situated immediately downstream of the ECAL.

The ECAL includes a Scintillator Pad Detector (SPD), which detects hits from charged particles to distinguish electrons from photons and π^0 decays. The ECAL also

includes a Pre-Shower (PS) detector to obtain longitudinal segmentation of the ECAL energy deposits, reducing the otherwise high background from charged pions.

The SPD and PS are formed from two layers of scintillating pads which sandwich a lead plate that is 2.5 radiation lengths (X_0) thick. The main section of the ECAL is formed from alternating layers of lead plate and scintillating tiles. It is $25X_0$ thick and has a Moliere radius of 3.5 cm. Similarly, the HCAL is formed of alternating layers of iron and scintillating tiles, but is only 5.9 interaction lengths thick. The ECAL (HCAL) subdetectors are separated into three (two) regions of differing segmentation to account for the varying particle occupancy within the detector.

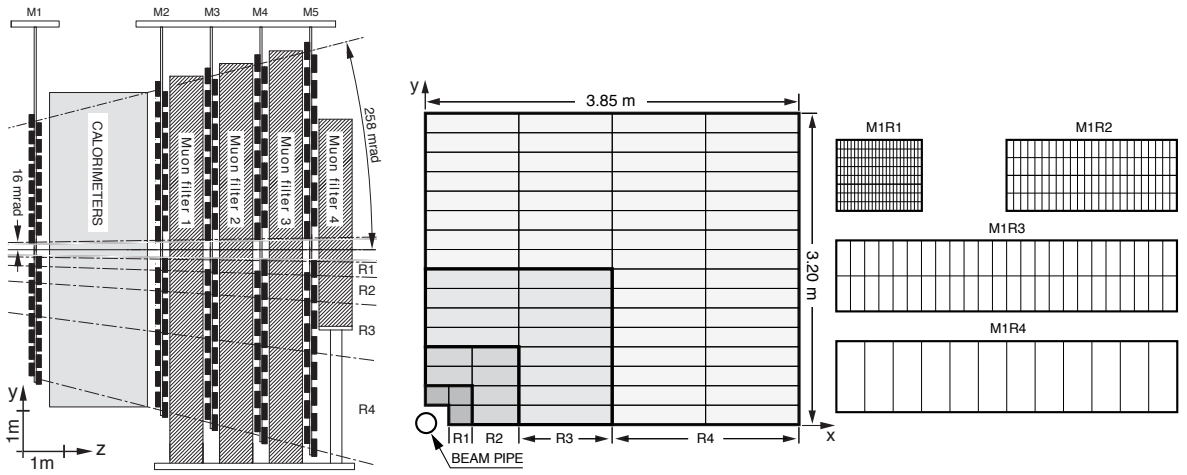
3.2.2.3 The muon system

Muons are highly penetrating particles and leave a distinctive signature in the detector. The Muon system is specially designed for triggering, as the production of muons in pp interactions is indicative of an interesting process having occurred. The muons produced in a $B^0 \to K^{*0}\mu^+\mu^-$ decay are critical to the selection of these events in the trigger.

The Muon system comprises five detector stations M1-M5. Station M1 is situated between RICH2 and the ECAL, while M2-M5 are situated downstream of the HCAL. The stations M2 to M5 are interleaved with 80 cm thick iron plates, as shown in Fig. 3.3(a). The location of M1 affords a longer baseline for the muon tracking and therefore improves the muon momentum resolution in the trigger. The Muon system uses Multi-Wire Proportion Chambers (MWPCs) for most of the active area, but triple-GEM detectors are used in the region of M1 closest to the beam axis where the particle flux is highest. The MWPCs (triple-GEMs) have a 5 ns (3 ns) time resolution, making them ideal for use in the trigger. In stations M2-M5, there are four layers of MWPCs. The information from the two most downstream (most upstream) layers are combined to form two 'logical' layers, which reduces the amount of information sent to the trigger. In M1 there are only two layers of MWPCs and triple-GEMs to reduce the amount of material in front of the ECAL.

Each region of the Muon stations is separated into logical pads, as shown in Fig. 3.3(b), to reduce the information sent to the trigger. The number of logical pads defines the spatial resolution of the Muon system. The stations M1-M3 are highly segmented in the horizontal plane to obtain a fast measurement of particle p_{T} for the trigger with a resolution of $\sim$25% [48]. The stations M4 and M5 are more coarsely segmented.

A muon must have $p > 3\,\mathrm{GeV}/c$ to traverse the magnetic field and reach the Muon system, and then only muons with $p > 6\,\mathrm{GeV}/c$ are expected to penetrate through the iron filters and produce hits in all of the Muon stations [48]. This information is exploited by two muon identification criteria: a stringent muon identification criteria (IsMuon);



(a) Side view

(b) Logical pad arrangement

Figure 3.3: A side view of the Muon system with the location of the stations and iron filters indicated (a). The arrangement of logical pads in the M1 station (b). The four regions of M1 and the individual chambers therein are indicated. On the right the separation of the chambers into logical pads is indicated. Both (a) and (b) are taken from Ref. [48].

and a less stringent muon identification criteria (IsMuonLoose). Both criteria require that a track has at least one hit in a subset of the Muon stations. The required subsets for the two criteria are momentum-dependent and are defined in Tab. 3.1 and Tab. 3.2 respectively. The hits are required to be in a field of interest defined by the extrapolation of the particle trajectory to the Muon system [49].

A further muon identification quantity, $\mathrm{DLL}_{\mu\pi}$, is formed by combining PID information from the RICH detector and Calorimeters with a 'squared distance' variable from the Muon system. This variable is computed from the x and y displacement of the closest hit to the track in a given logical pad, and is defined as

$$D = \frac{1}{N}\sum_{i=0}^{N}\left\{\left(\frac{x_{\mathrm{close,i}} - x_{\mathrm{track}}}{\mathrm{pad}_x}\right)^2 + \left(\frac{y_{\mathrm{close,i}} - y_{\mathrm{track}}}{\mathrm{pad}_y}\right)^2\right\}, \tag{3.2.3}$$

where $x_{\mathrm{close,i}}$ ($y_{\mathrm{close,i}}$) are the x (y) location of the closest hit to the track in a given pad x_{track} (y_{track}) is the x (y) location of the track, and pad_x (pad_y) is the x (y) size of the relevant logical pad. The index i runs over each Muon station for which the track satisfies the IsMuonLoose criteria [49].

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

Table 3.1: IsMuon identification requirements.

Momentum range	Hits required in muon stations
$3\,\text{GeV}/c < p < 6\,\text{GeV}/c$	at least two of M2, M3 and M4
$p > 6\,\text{GeV}/c$	at least three of M2, M3, M4 and M5

Table 3.2: IsMuonLoose identification requirements.

3.2.3 Performance of the LHCb detector

The performance of the LHCb magnet and tracking systems are encapsulated in a number of critical quantities, such as the tracking efficiency; the IP resolution; the PV resolution; and the angular resolution.

The momentum of charged particles with momenta 3 (100) GeV/c is determined with a resolution of $\sigma_p/p \sim 0.4$ (0.6)% [50]. The tracking efficiency is measured in data and in simulated events as function of p and η. The results are shown in Fig. 3.4. The average tracking efficiency is 98%, which varies as a function of the particle momentum, but not as a function of η [51]. The ratio of the tracking efficiency in data and in simulated events is shown in Fig. 3.4. The uncertainty on this ratio is considered as a source of systematic uncertainty in Sec. 6.3.2. Conversely, charge misidentification in LHCb is a negligible source of systematic uncertainty.

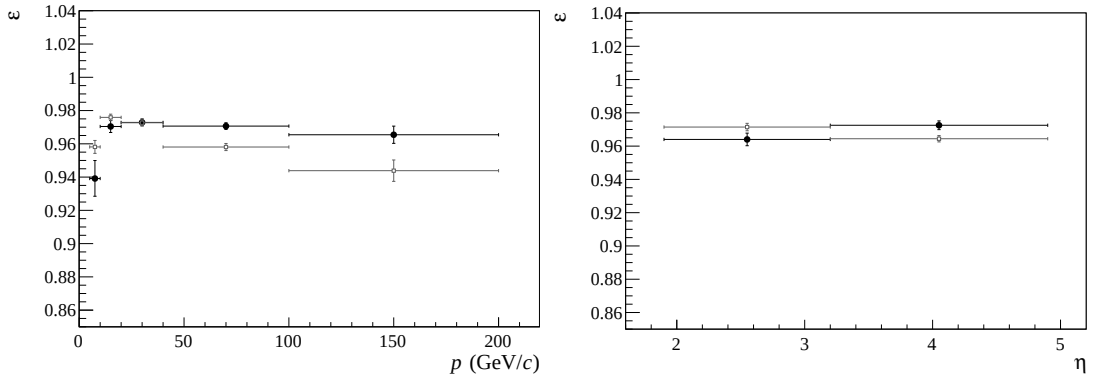
The PV position resolution in the transverse and z directions is measured in data and in simulated events. The results are shown in Figs. 3.5(a) and 3.5(b). The PV position resolution is $\sim 20\,\mu\text{m}$ in the transverse direction and $\sim 100\,\mu\text{m}$ in the z direction. The IP resolution in the transverse direction is also measured in data and in simulated events. The results are shown in Fig. 3.5(c). The IP resolution (σ_{IP}) in the transverse direction is

$$\sigma_{\text{IP}} = 13.7 \pm 24.6/p_{\text{T}}\,\mu\text{m} \tag{3.2.4}$$

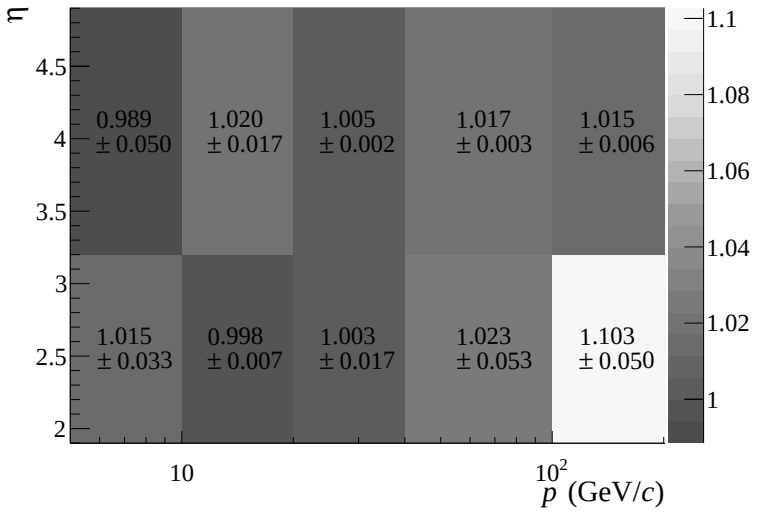
in data and

$$\sigma_{\text{IP}} = 11.3 \pm 20.8/p_{\text{T}}\,\mu\text{m} \tag{3.2.5}$$

in simulated events [46]. The difference in the PV and IP resolutions in data and in simulated events is highly correlated, and is considered as a source of systematic uncertainty

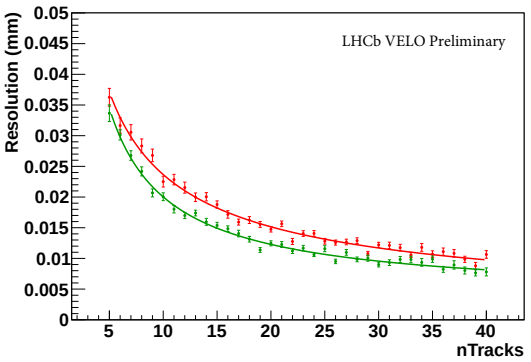


(a) p (b) η

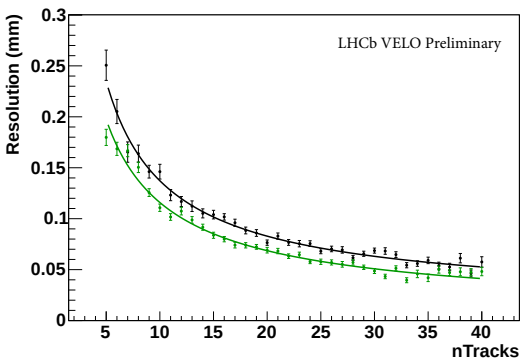


(c) Ratio in data and in simulated events

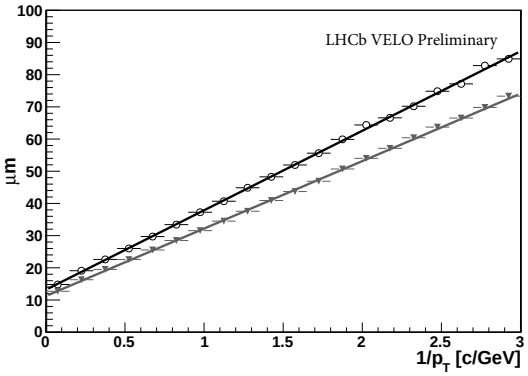
Figure 3.4: The tracking efficiency as a function of momentum (a) and η (b) in data (black points) and simulated events (open points); and the ratio of tracking efficiency in data and in simulated events as a function of p and η (c). Taken from Ref. [51].



(a) PV x resolution



(b) PV z resolution



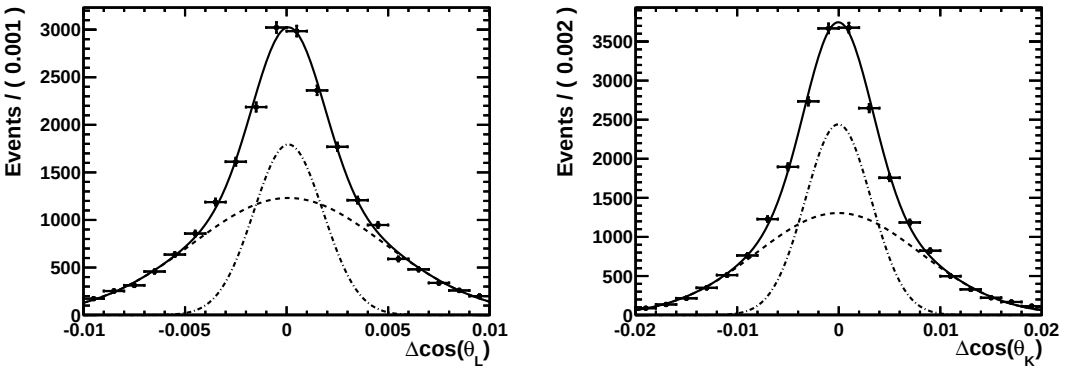
(c) IP x resolution

Figure 3.5: Comparison of the primary vertex resolution in the horizontal (a) and z (b) directions in data (black points) and simulated events (red or green points), and a comparison of the impact parameter resolution in the horizontal direction (c) in data (black open points) and simulated events (black triangles). Adapted from Ref. [46].

in Sec. 6.3.2.

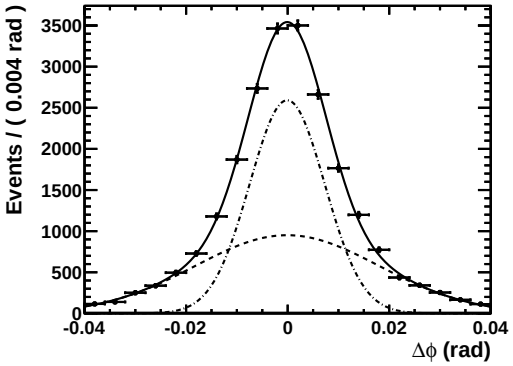
Of particular relevance to the measurement of the $B^0 \to K^{*0}\mu^+\mu^-$ angular observables is the angular resolution of the decay angles $\cos\theta_L$, $\cos\theta_K$ and ϕ. The angular resolution is extracted using a fit to simulated events, based on two Gaussian probability density functions (pdfs) that share the same mean value but have different widths, defined as 'narrow' and 'wide'. The distribution of events is shown in Fig. 3.6 with the fit overlaid. For each angle, the 'narrow' component of the resolution is about 0.1%. The 'wide' component is about three times larger in each angle (0.3%). The angular resolution is considered as a negligible source of systematic uncertainty.

The performance of the RICH detector is encapsulated in the ability to distinguish between particles of different masses using the DLL information. The efficiency with which kaons are separated from pions as a function of momentum is measured in data. The results are shown in Fig. 3.7(a). Requiring that the likelihood for a track is greater



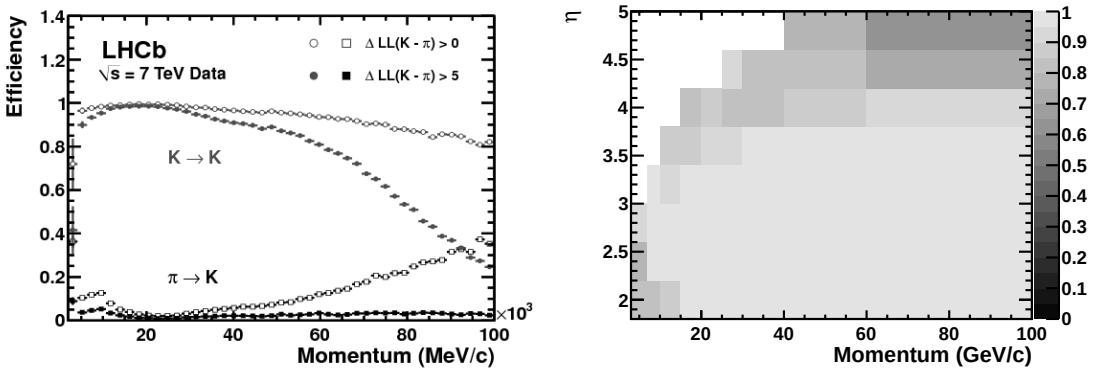
(a) $\cos\theta_L$ resolution

(b) $\cos\theta_K$ resolution



(c) ϕ resolution

Figure 3.6: Comparison of the angular resolution in $\cos\theta_L$ (a) and $\cos\theta_K$ (b) and ϕ (c) using simulated events.



(a) $K-\pi$ separation

(b) IsMuon efficiency ratio

Figure 3.7: The PID efficiency for kaons and the pion misidentification efficiency, measured in data as a function of momentum, for two different $\mathrm{DLL}_{K\pi}$ requirements (a). Taken from Ref. [52]. The ratio of the IsMuon efficiency for muons in data and in simulated events (b).

for the kaon hypothesis than the pion hypothesis, $\mathrm{DLL}_{K\pi} > 0$, and averaging over the momentum range $2 < p < 100\,\mathrm{GeV}/c$, the kaon efficiency is $\sim 95\%$ for a pion misidentification efficiency of $\sim 10\%$ [52].

The muon system has a track finding efficiency above 98% in the most central region of M1, and above 99% elsewhere [48]. The efficiency of the IsMuon criteria is measured using $B^0 \to J/\psi\, K^{*0}$ decays in the data and in simulated $B^0 \to J/\psi\, K^{*0}$ events. The efficiency is p_{T} dependent and varies between $95 < \epsilon_{\mathrm{IsMuon}} < 98\%$ for a hadron misidentification rate of $\sim 1\%$. The ratio of the efficiency of the IsMuon criteria is measured using $B^0 \to J/\psi\, K^{*0}$ decays in the data and simulated $B^0 \to J/\psi\, K^{*0}$ events as a function of p and η. The results are shown in Fig. 3.7(b). The uncertainty on the measurement of the IsMuon efficiency ratio is considered as a source of systematic uncertainty in Sec. 6.3.2.

3.3 Trigger

The LHCb trigger is designed to identify the $\sim 0.3\%$ of pp interactions that produce a $b\bar{b}$ pair. The trigger is divided into two parts: the First Level Trigger (L0); and the High Level Trigger (HLT). The L0 trigger is based on bespoke hardware and runs synchronously with the 40 MHz LHC clock. It reduces the 11 MHz event rate to about 870 kHz [47].

The HLT is a software application that is run on a large processor farm. It is separated into two stages: the first stage of the HLT (HLT1) and the second stage of the HLT (HLT2). The HLT1 algorithms use a partial event reconstruction to reduce the L0 output rate to 43 kHz, and must process each event within 30 ms. The HLT2 algorithms have

access to the fully reconstructed event and reduce the event rate down to 3 kHz, which is written to storage. Both HLT1 and HLT2 take an inclusive approach to selecting events that contain B hadrons. These inclusive algorithms impose requirements on general properties of B hadron decays. The HLT inclusive algorithms have a rejection factor of $\sim$1000 with respect to the events that passed the L0 trigger, and a $b\bar{b}$ purity of close to $\sim$100% [53].

The trigger uses the kinematic properties of B hadron decays to identify events in which $b\bar{b}$ pairs are produced. The rate of muon production in pp interactions is considerably smaller than that of hadrons. Muons also leave an easily identifiable signature in the detector, meaning they can be selected with higher efficiency than hadrons. The trigger strategy for the $B^0 \to K^{*0}\mu^+\mu^-$ decay is therefore based around the muons, and the following discussion therefore focuses on those parts of the trigger which use muons.

3.3.1 First level trigger

The L0 trigger is divided into three independent triggers: the L0Calorimeter trigger; the L0PileUp trigger; and the `L0Muon` trigger. Each of these uses information from a subset of the LHCb detector systems, and must process each event within a total of 4 µs, with $\sim$2 µs available for the computation of the trigger decision.

Only events that satisfy the `L0Muon` trigger are used in the $B^0 \to K^{*0}\mu^+\mu^-$ analysis. This trigger uses information from the five Muon stations M1-M5 to reconstruct and measure the properties of candidate muons. The Muon stations are separated into four quadrants. In each quadrant, a search is performed for hits that define a straight line through the five muon stations that points towards to the interaction region in the vertical plane. In the horizontal plane the search is limited to muons with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}/c$. The p_{T} is measured with a precision of $\sim$25% using the information from the M1 and M2 stations only.

The two tracks with the largest p_{T} in each quadrant are used to form the trigger decision. There are two `L0Muon` algorithms: the `L0Muon` algorithm imposes requirements on the track with the largest p_{T}; the `L0DiMuon` algorithm imposes requirements on the product of the p_{T} of the two tracks with the largest p_{T}. The efficiency of the `L0Muon` trigger algorithm is evaluated as a function of the average p_{T} of the two muons using $B^0 \to J/\psi\, K^{*0}$ decays in the data and simulated $B^0 \to J/\psi\, K^{*0}$ events. The results are shown in Fig. 3.8. The efficiency of the `L0Muon` trigger is typically $\sim$90% and varies as a function of the average p_{T} of the two muons. The difference in the efficiency measured in data and in simulated events is considered as a source of systematic uncertainty in Sec. 6.3.2. If the criteria of any of the L0 trigger algorithms are satisfied, the event is sent to the HLT for further processing.

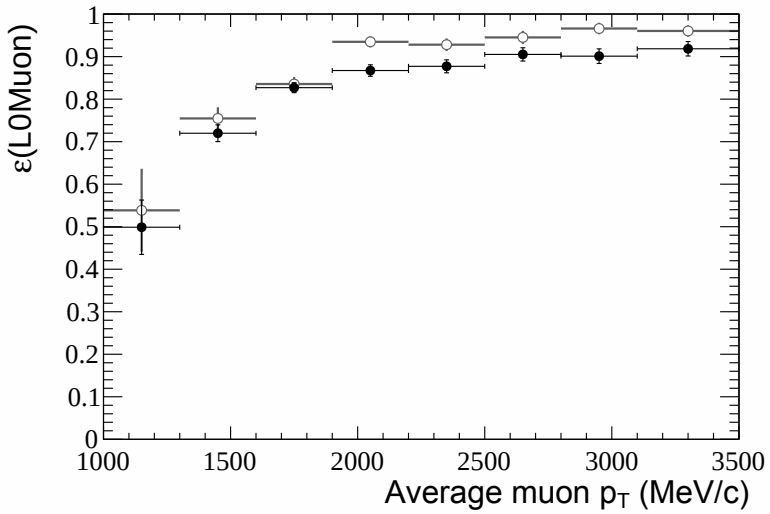


Figure 3.8: Efficiency of the `L0Muon` trigger as a function of the average p_{T} of the two muons in $B^0 \to J/\psi K^{*0}$ decays in the data (black solid points) and in simulated events (grey open points).

3.3.2 First stage of the High Level Trigger

In HLT1, VELO tracks are selected either by being tagged as a muon candidate or by satisfying a number of other criteria. Candidate muon tracks are created by opening a search window in the M3 station for each VELO track. The search window is defined by extrapolating the VELO track assuming a 6 GeV/c muon. If any hits are found in the M3 search window, the hits are combined with the VELO track to define a search windows in the M2, M4 and M5 stations. A candidate track is provisionally accepted if another hit is found within any of the additional search windows. A linear χ^2 fit of the track is then performed in the horizontal plane, with the quality of the track fit expressed in terms of the 'track quality' χ^2/dof. The requirement that track quality $\chi^2/\mathrm{dof} < 25$ is imposed on each track, and as soon as the first candidate track is accepted, the VELO track is tagged as a muon candidate and the algorithm stops [54]. To select a VELO track without tagging it as a muon, the track must satisfy criteria imposed on the IP, the number of hits on the track, and the expected number of hits.

Those VELO tracks that satisfy either of the two selections described above are used to open search windows in the OT and IT sections of the tracking stations in the 'forward tracking' stage. Requirements on the p and p_{T} of the tracks are imposed to control the size and number of search windows, which reduces the processing time to an acceptable level. Track segments reconstructed in the T1-T3 stations are combined with a given VELO track and the momentum of the track is determined. Each of the tracks are then fitted with a Kalman filter based track fit [55, 56] to obtain the χ^2 and full covariance matrix at the start of the track.

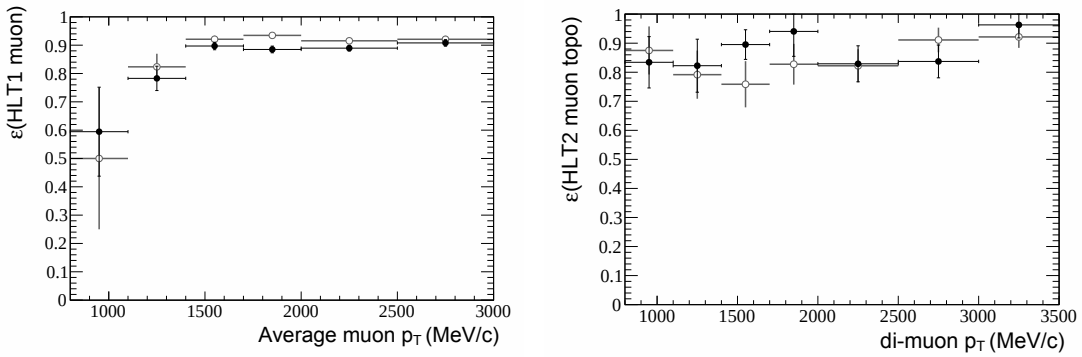
There are two types of misreconstructed tracks, which are removed as part of the forward tracking stage defined above. The first type of misreconstructed tracks are 'ghost' tracks, which are formed from a combination of hits produced by two different particles. This occurs when track segments in the VELO and tracking stations are reconstructed, but are matched together incorrectly. Such tracks typically have a low track quality χ^2/dof and are removed during the track reconstruction by imposing requirements on this quantity. The second type are 'clone' tracks, which are formed when the track from a genuine particle is reconstructed as multiple tracks. Clone tracks typically share a large number of hits, and will therefore have very similar momentum and trajectory. Clone tracks are identified by computing the Kullback-Leibler distance (KL) between pairs of tracks to quantify the amount of information shared between the two tracks [57]. In LHCb a KL < 5000 indicates a large sharing of information and flags the two tracks as clone tracks. One track from the pair is then selected by picking the track with the most hits [58].

The two HLT1 algorithms used in the present analysis are the `Hlt1TrackAllL0` and `Hlt1TrackMuon` algorithms [59]. The `Hlt1TrackAllL0` algorithm imposes requirements on the IP and momentum of any track from the forward tracking. Since muon tracks form only a small percentage of the total number of tracks, by requiring a track to satisfy the IsMuon criteria the `Hlt1TrackMuon` algorithm can impose less stringent requirements than the `Hlt1TrackAllL0` algorithm and remain within the time limitations. The efficiency of the `Hlt1TrackMuon` algorithm is evaluated as a function of the average muon p_{T} using $B^0 \to J/\psi\, K^{*0}$ decays in the data and simulated $B^0 \to J/\psi\, K^{*0}$ events. The results are shown in Fig. 3.9(a). The efficiency vary as a function of the average muon p_{T}. There is no significant difference between the efficiency in data and in simulated events.

3.3.3 Second stage of the High Level Trigger

The HLT2 stage performs an improved forward tracking step (including the removal of any clone tracks) and then selects events using both inclusive and exclusive criteria. The exclusive HLT2 algorithms select events by imposing stringent requirements on the invariant mass and kinematic properties of B hadron decays. The inclusive HLT2 algorithms select certain decay topologies, and are therefore collectively called the 'topological' trigger. In the analysis presented in this thesis only $B^0 \to K^{*0}\mu^+\mu^-$ candidates that satisfy one of two of the HLT2 exclusive algorithms, or one of six of the topological trigger algorithms, are used.

In the HLT2 stage, the event rate is sufficiently low to allow the forward tracking of all the VELO tracks that satisfy minimum requirements on the p and p_{T}. The topological trigger forms so-called 'topo-tracks' by imposing further requirements on the track quality



(a) HLT1 muon

(b) HLT2 muon topo

Figure 3.9: Efficiency of the HLT1 muon algorithms as a function of the average muon p_{T} (a) and efficiency of the HLT2 muon topological algorithms as a function of the di-muon p_{T} (b) in data (black solid circles) and simulated events (black open circles). Taken from Ref. [47].

χ^2/dof and IP. There are three distinct topological algorithms that correspond to the selection of 2-, 3- and 4-body topo objects. The topological trigger combines two topo-tracks, which are required to have a distance of closest approach (DOCA) of less than 0.2 mm, to form a 2-body object. At least one of the two topo-tracks must have satisfied the `Hlt1TrackAllL0` or `Hlt1TrackMuon` algorithm in HLT1. A 3- or 4-body object is produced by combining a 2- or 3-body object with a further topo-track, requiring the DOCA between the 2- or 3-body object and the topo-track to be less than 0.2 mm.

If an n-body candidate does not contain all of the particles from a given B hadron decay, the invariant mass will not equal that of the decaying hadron. In lieu of imposing requirements on the invariant mass, the corrected mass

$$\mathrm{m}_{corr} = \sqrt{m^2 + |p'_{T\,miss}|^2} + |p'_{T\,miss}|^2 \tag{3.3.1}$$

is used, where m^2 is the invariant mass of the n-body object and $|p'_{T\,miss}|^2$ is the missing momentum transverse to the direction of travel. The direction of travel is defined by the PV and the n-body vertex. The quantity m_{corr} is the minimal mass of the decaying hadron if a massless particle was omitted from the n-body object. To prevent D hadron decays that are combined with a random track (from elsewhere in the event) from forming valid n-body candidates, all $(n-1)$-body objects are required to have either an invariant mass greater than 2.5 GeV/c^2 or a large IP χ^2.

For an n-body candidate to pass the corresponding n-body topological trigger algorithm, it must satisfy a requirement imposed on the response of a bespoke multivariate classifier [53]. This classifier combines various kinematic properties of the n-body object, such as the vertex separation χ^2, and kinematic properties of the n tracks which form the

object, such as their DOCA and IP χ^2.

Muon tracks are obtained by imposing the IsMuon criteria on each track from the HLT2 forward tracking step. For these muon tracks, distinct 'muon topological' trigger algorithms are used. These algorithms form n-body candidates using the same procedure as described above, however the lower rate from requiring a muon to be present allows a less stringent requirement on the classifier response to be imposed. This brings the number of topological trigger algorithms used in this analysis up to six. The efficiency of the HLT2 muon topological algorithms are evaluated as a function of the di-muon p_{T} using $B^0 \to J/\psi K^{*0}$ decays in the data and simulated $B^0 \to J/\psi K^{*0}$ events. The results are shown in Fig. 3.9(b). The efficiency is independent of the di-muon p_{T}. There is no significant difference between the efficiency in data and in simulated events.

3.4 Production of simulated events

Simulated events are used extensively in the analysis presented in this thesis. The events are produced by the physics simulation application Gauss [60]. This simulates a pp interaction using Pythia 6.4 [61] with a specific LHCb configuration [62]. The signal decay is simulated using EvtGen [63], in which final state radiation is generated with Photos [64]. The Geant4 toolkit [65] is used to simulate the interactions of each generated particle with the LHCb detector, as described in Ref. [66]. The response of the detector to the particles is subsequently simulated, digitised, and stored in the same fashion as the data.

Chapter 4

Selection

In each event, four-track combinations are formed into $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ (hereafter denoted $B^0 \to K^{*0}\mu^+\mu^-$) candidates. With the order of $\sim$150 tracks per event and the 3 kHz output rate from the LHCb trigger, too many candidates are produced to use in an angular fit. The selection defined in this section removes the least signal-like candidates to reduce the data sample that is fitted.

A signal candidate is formed by correctly combining the four final state particles of a $B^0 \to K^{*0}\mu^+\mu^-$ decay, with each particle assigned the correct mass and charge hypothesis. Any other candidate is background, of which the following classes are considered in the analysis:

- **Class I** background, which is formed from candidates with the wrong mass hypothesis assigned to one or more tracks of a genuine B hadron decay;
- **Class II** background, which is formed from candidates with one or more tracks added to a genuine B hadron decay to form a four-track final state;
- **Class III** background, which is formed from random combinations of tracks, typically from other B hadron decays, in the event.

Class III background has a smoothly varying shape in $K\pi\mu\mu$ invariant mass ($\mathrm{m}_{K\pi\mu\mu}$). The amount and angular shape of these candidates is extracted from a fit to candidates in the data in the upper $\mathrm{m}_{K\pi\mu\mu}$ mass sideband. This sideband is defined as the mass range $5350 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. Class I and Class II backgrounds have a peaking structure in $\mathrm{m}_{K\pi\mu\mu}$ and do not contribute above $\mathrm{m}_{K\pi\mu\mu} \gtrsim 5450\,\mathrm{MeV}/c^2$. Therefore neither the amount nor the angular distribution can be taken from a fit to the upper mass sideband. Furthermore the angular distribution is not in general known, and so cannot be included in the angular fit using simulated events. Veto criteria are therefore used to

reduce Class I and Class II backgrounds to a level at which the angular fit is not sensitive to their contribution.

The selection introduces acceptance effects due to correlations between particle kinematics and the decay angles. For example, candidates with large momentum asymmetry in the lepton system will accumulate at $\cos\theta_L \sim \pm 1$ and will typically contain one low momentum muon. Requirements on quantities such as muon p_{T} or IP χ^2, which are correlated with the particle momentum, therefore remove candidates in these $\cos\theta_L$ regions. Similarly, requirements on hadron IP χ^2 or $\mathrm{DLL}_{K\pi}$ introduce acceptance effects at $\cos\theta_K \sim \pm 1$. The selection is designed to minimise the use of quantities that introduce such acceptance effects. In this way the acceptance effects are moderated, reducing the systematic error introduced to account for any uncertainty in the acceptance correction.

The pre-selection criteria detailed in Sec. 4.1 retain the most signal-like candidates for further analysis. Each candidate is required to satisfy the subset of the trigger algorithms detailed in Sec. 4.2, which select the majority of $B^0 \to K^{*0}\mu^+\mu^-$ events. Class I and Class II backgrounds are rejected with the specific vetoes detailed in Sec. 4.3. Class III background is reduced using the multivariate analysis (MVA) detailed in Sec. 4.4.

4.1 Pre-selection

Pre-selection requirements are imposed to reduce the number of candidates input in to the MVA (detailed in Sec. 4.4) and to the subsequent angular fit. The requirements are summarised in Tab. 4.1. The quantities used in the selection which were not defined in Sec. 3 are defined below. The pointing angle (θ_{p}) is defined as the angular separation of the momentum and displacement vectors of the candidate. The displacement vector is the displacement of the SV from the associated PV. The PV associated with a given candidate is that with respect to which the candidate has the smallest IP. The quantity θ_{sep} is defined as the angular separation of the momentum vectors of two particles, and is calculated for each pair of particles in a given candidate. These quantities are shown schematically in Fig. 4.1.

In order to reject candidates containing tracks from beam-gas interactions in the LHC beam pipe, requirements are placed on the separation of the PV position $(x,\ y,\ z)_{\mathrm{PV}}$ from the mean PV position $(\bar{x},\ \bar{y},\ \bar{z})_{\mathrm{PV}}$. Tracks originating from a PV are removed by a minimum requirement on the IP χ^2 of the tracks. To remove ghost- and clone-tracks that satisfy the criteria applied in the forward tracking, minimum requirements are placed on the track quality χ^2/dof and θ_{sep}. Misidentified particles are removed by requiring the kaon and pion to have PID information from the RICH detector system and to satisfy minimum requirements imposed on the kaon and pion $\mathrm{DLL}_{K\pi}$. Each muon is

Vertex/Track/Particle	Selection Requirement
PV	$\lvert x_{\rm PV} - \bar{x_{\rm PV}}\rvert < 5\,{\rm mm}$
PV	$\lvert y_{\rm PV} - \bar{y_{\rm PV}}\rvert < 5\,{\rm mm}$
PV	$\lvert z_{\rm PV} - \bar{z_{\rm PV}}\rvert < 200\,{\rm mm}$
Track	IP $\chi^2 > 9$
Track	quality $\chi^2/{\rm dof} < 5$
Track	$\theta_{\rm sep} > 1\,{\rm mrad}$
Track	KL > 5000
Kaon	hasRICH criteria
Kaon	${\rm DLL}_{K\pi} > -5$
Pion	hasRICH criteria
Pion	${\rm DLL}_{K\pi} < 25$
Muons	IsMuon criteria
Muons	${\rm DLL}_{\mu\pi} > -3$
K^{*0}	vertex quality $\chi^2/{\rm dof} < 12$
K^{*0}	vertex separation $\chi^2 > 9$
K^{*0}	$792 < {\rm m}_{K\pi} < 992\,{\rm MeV}/c^2$
$\mu^+\mu^-$	vertex quality $\chi^2/{\rm dof} < 12$
$\mu^+\mu^-$	vertex separation $\chi^2 > 9$
B^0	$4850 < {\rm m}_{K\pi\mu\mu} < 5780\,{\rm MeV}/c^2$
B^0	$\cos(\theta_{\rm p}) > 0.9999$
B^0	IP $\chi^2 < 16$
B^0	vertex quality $\chi^2/{\rm dof} < 6$
B^0	vertex separation $\chi^2 > 121$

Table 4.1: Pre-selection requirements. Here 'Track' refers to each of the four final state particles and $\mu^+\mu^-$ refers to the di-muon system.

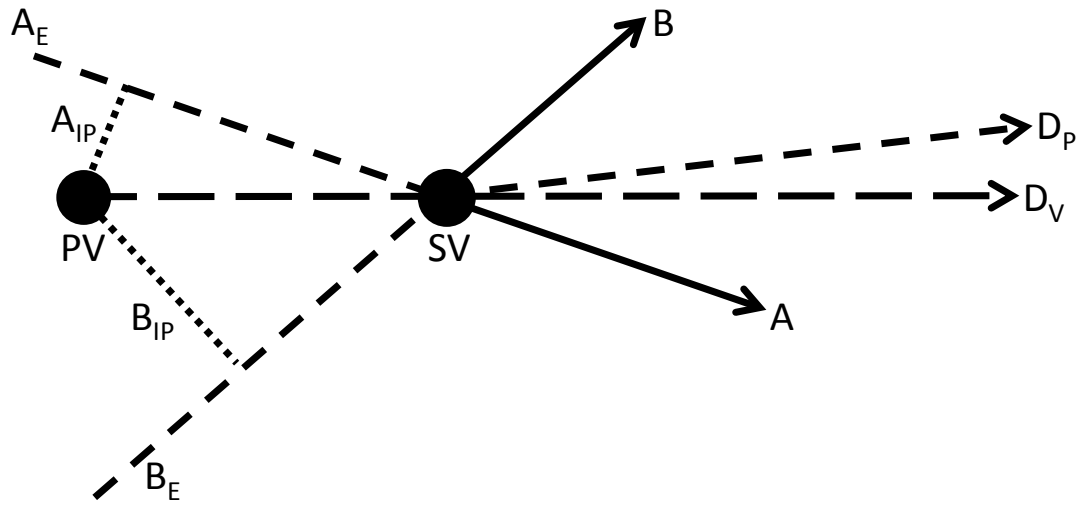


Figure 4.1: Schematic of a B meson, which is produced at the PV, decaying to two particles labelled A and B, which form a SV. The extrapolation of the particle tracks, A_E and B_E, are used to calculate the IP of the particles, A_{IP} and B_{IP}. The displacement vector of the B meson is labelled D_V and the momentum vector of the B meson is labelled D_P. The quantity $\theta_{\rm p}$ is the angle between D_V and D_P. The quantity $\theta_{\rm sep}$ is the angle between particle A and particle B.

required to satisfy the IsMuon criteria and minimum requirements on $\mathrm{DLL}_{\mu\pi}$. To remove candidates that do not contain a K^{*0} resonance, the $K\pi$ invariant mass ($\mathrm{m}_{K\pi}$) is required to be consistent with that of the K^{*0}, and requirements are placed on the K^{*0} vertex quality χ^2/dof and vertex separation χ^2. To ensure the candidate contains two muons that are consistent with originating from a single vertex, minimum requirements are placed on the di-muon vertex quality χ^2/dof and the vertex separation χ^2. Finally, to ensure the candidate is consistent with a B^0 decay, the B^0 candidate must satisfy minimum requirements on $\cos(\theta_{\rm p})$, IP χ^2, vertex quality χ^2/dof and vertex separation χ^2.

4.2 Trigger

The tracks that form each $B^0 \to K^{*0}\mu^+\mu^-$ candidate are required to pass the subset of the trigger algorithms that accept the majority of the signal events. Events that were triggered on tracks other than those that form the signal candidate are rejected, ensuring that the analysis is sensitive only to the accurate modelling of the signal process in simulated events.

At the L0 trigger stage, the signal candidate is required to contain a muon track with $p_{\rm T} > 1.48\,\mathrm{GeV}/c$ and so pass the `L0Muon` trigger algorithm [54]. At the HLT1 stage, the signal candidate is required to pass either the `Hlt1TrackAllL0` or `Hlt1TrackMuon`

Requirement	`Hlt1TrackAllL0`	`Hlt1TrackMuon`
N_{miss}	< 3	Zero
quality χ^2/dof	< 2.5	< 2
p_{T}	$> 1.7\,\text{GeV}/c$	$> 1.0\,\text{GeV}/c$
p	$> 10\,\text{GeV}/c$	$> 8\,\text{GeV}/c$
IP	$> 0.1\,\text{mm}$	$> 0.1\,\text{mm}$
IP χ^2	> 16	> 16
IsMuon	Any	Yes

Table 4.2: Requirements for a track to satisfy the HLT1 `Hlt1TrackAllL0` and `Hlt1TrackMuon` trigger algorithms.

algorithm. The requirements imposed by each algorithm are summarised in Tab. 4.2.

Signal candidates are also required to pass any of the six inclusive HLT2 'topological' trigger algorithms described in Sec. 3.3.3 or either of two exclusive HLT2 algorithms. The `Hlt2SingleMuon` algorithm selects a single displaced muon by requiring a track to pass `Hlt1TrackMuon`, and then satisfy $p_{\text{T}} > 1.3\,\text{GeV}/c$, track quality $\chi^2/\text{dof} < 2$, IP $> 0.5\,\text{mm}$ and IP $\chi^2 > 200$. The `Hlt2DiMuonDetached` algorithm selects displaced di-muon vertices with $\mu^+\mu^-$ invariant mass $\text{m}_{\mu\mu} > 1.0\,\text{GeV}/c^2$, where each of the muon tracks is required to satisfy the IsMuon criteria and IP $\chi^2 > 9$, track quality $\chi^2/\text{dof} < 5$ and $p_{\text{T}} > 500\,\text{MeV}/c$. Furthermore, the di-muon system is required to satisfy $p_{\text{T}} > 1.5\,\text{GeV}/c$, vertex quality $\chi^2/\text{dof} < 25$ and vertex separation $\chi^2 > 49$ [54].

4.3 Specific vetoes for exclusive backgrounds

Processes that form Class I and Class II background, as well as other B^0 meson decays with the same final state as the $B^0 \to K^{*0}\mu^+\mu^-$ decay, are not smoothly distributed in $\text{m}_{K\pi\mu\mu}$. These contributions are reduced to a negligible level to reduce the sensitivity of the angular fit to their unknown angular distribution, thereby reducing the associated systematic uncertainty, and to make the $\text{m}_{K\pi\mu\mu}$ distribution easier to fit.

The largest contributions to Class I and Class II backgrounds are expected to be from a small number of exclusive decays. Simulated events are used to design the veto criteria and to evaluate the residual background contribution.

Irreducible contributions from the decays $B^0 \to J/\psi\,(\to \mu^+\mu^-)K^{*0}(\to K^+\pi^-)$ (hereafter denoted $B^0 \to J/\psi\,K^{*0}$) and $B^0 \to \psi(2S)(\to \mu^+\mu^-)K^{*0}(\to K^+\pi^-)$ (hereafter denoted $B^0 \to \psi(2S)K^{*0}$) are considered in Sec. 4.3.1. Irreducible contributions to the upper $\text{m}_{K\pi\mu\mu}$ mass sideband from $\overline{B}{}^0_s \to J/\psi\,K^{*0}$ and $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ decays are considered in

Sec. 4.3.2. Further contributions to the Class I and Class II backgrounds are considered in Sec. 4.3.4 to Sec. 4.3.9. The residual contribution from each of the Class I and Class II backgrounds is summarised in Sec. 4.3.10.

4.3.1 Irreducible contributions from $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to \psi(2S) K^{*0}$ decays

The decays $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to \psi(2S) K^{*0}$ constitute an irreducible background to $B^0 \to K^{*0}\mu^+\mu^-$. Contributions from these decays can only be removed by discarding regions of $\mathrm{m}_{\mu\mu}$. Therefore, $B^0 \to K^{*0}\mu^+\mu^-$ candidates that satisfy $2946 < \mathrm{m}_{\mu\mu} < 3176\,\mathrm{MeV}/c^2$ and $3586 < \mathrm{m}_{\mu\mu} < 3766\,\mathrm{MeV}/c^2$ are rejected.

Mis-reconstructed $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to \psi(2S) K^{*0}$ decays contribute to the upper mass sideband, which is defined as $5330 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. To remove this contribution the vetoes are extended to $2946 < \mathrm{m}_{\mu\mu} < 3216\,\mathrm{MeV}/c^2$ and $3586 < \mathrm{m}_{\mu\mu} < 3808\,\mathrm{MeV}/c^2$ in the $5330 < \mathrm{m}_{K\pi\mu\mu} < 5450\,\mathrm{MeV}/c^2$ range.

The radiative decays $B^0 \to J/\psi\,(\to \mu^+\mu^-\gamma) K^{*0}$ and $B^0 \to \psi(2S)(\to \mu^+\mu^-\gamma) K^{*0}$ have reduced values of $\mathrm{m}_{\mu\mu}$ and $\mathrm{m}_{K\pi\mu\mu}$ as the photon is not reconstructed. These radiative decays therefore contribute to the lower $\mathrm{m}_{K\pi\mu\mu}$ mass sideband $5150 < \mathrm{m}_{K\pi\mu\mu} < 5230\,\mathrm{MeV}/c^2$. To remove this contribution, the vetoes are extended to $2796 < \mathrm{m}_{\mu\mu} < 3176\,\mathrm{MeV}/c^2$ and $3436 < \mathrm{m}_{\mu\mu} < 3766\,\mathrm{MeV}/c^2$ in the lower $\mathrm{m}_{K\pi\mu\mu}$ mass sideband. Candidate $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to \psi(2S) K^{*0}$ decays in the data are shown in Fig. 4.2 with the veto regions indicated.

These vetoes also remove other background decays that have a J/ψ or $\psi(2S)$ resonance. The contribution from decays such as $B^0_s \to J/\psi\,\phi$ or $B^+ \to J/\psi\, K^+$ is therefore reduced to a negligible level.

Extending the vetoes in the upper and lower $\mathrm{m}_{K\pi\mu\mu}$ mass sidebands results in the modification of the $\mathrm{m}_{K\pi\mu\mu}$ distribution of the Class III background in the three q^2 bins adjacent to the J/ψ and $\psi(2S)$ veto regions. This is taken into account by increasing the weight of each candidate in the appropriate $\mathrm{m}_{K\pi\mu\mu}$ and q^2 ranges to correct for the fraction of the q^2 bin that is removed.

The weighting assumes a uniform q^2 distribution of Class III background within each q^2 bin. From first principles, the distribution of Class III background is expected to be uniform across q^2. However any acceptance effect will modify this distribution. From the total q^2 acceptance effect shown in Fig. 5.5, the efficiency variation within the relevant analysis q^2 bins is $\approx 10\%$. When the weights are varied by $\pm 5\%$, the signal and background yields extracted in each analysis q^2 bin are consistent. Assuming a uniform q^2 distribution of Class III background within each q^2 bin does not affect the results of the analysis.

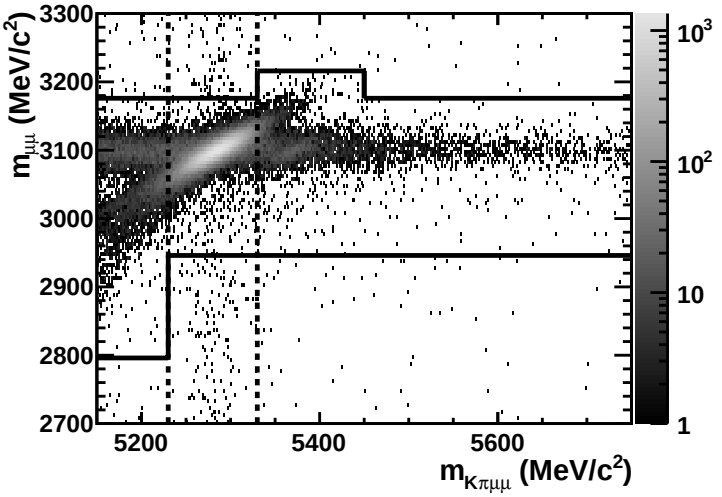


Figure 4.2: The ($m_{K\pi\mu\mu}$, $m_{\mu\mu}$) plane after the selection (including the multivariate selection described in Sec. 4.4) is applied. The decays $B^0 \to J/\psi\, K^{*0}$ and $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ peak at the nominal (B^0,J/ψ) and (B^0_s,J/ψ) masses respectively. The solid lines indicate the $B^0 \to J/\psi\, K^{*0}$ veto, as described in Sec. 4.3.1. The dashed lines indicate the signal region in $m_{K\pi\mu\mu}$.

4.3.2 Irreducible contributions from $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ and $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ decays

The decay $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ contributes to the upper $m_{K\pi\mu\mu}$ mass sideband, peaking at $m_{K\pi\mu\mu} \sim 5366\,\text{MeV}/c^2$ and $m_{\mu\mu} \sim 3096\,\text{MeV}/c^2$, and causes a spurious enhancement of the combinatoric background in the $m_{K\pi\mu\mu}$ fit to $B^0 \to J/\psi\, K^{*0}$ candidates.

The measured ratio of branching fractions

$$\frac{\mathcal{B}(\overline{B}{}^0_s \to J/\psi\, K^{*0})}{\mathcal{B}(B^0 \to J/\psi\, K^{*0})} = (3.43^{+0.36}_{-0.34} \pm 0.5) \times 10^{-2}\ [67], \tag{4.3.1}$$

and the measured ratio of hadronisation fractions of B^0_s and B^0 mesons

$$f_s/f_d = 0.256 \pm 0.020\ [68], \tag{4.3.2}$$

are used to calculate the contribution of $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ to be $(0.88 \pm 0.17)\%$ of the $B^0 \to J/\psi\, K^{*0}$ yield. Differences in the efficiency of the two channels are ignored. The ratio of the efficiency of the two decays is measured to be 0.97 ± 0.01 in Ref. [67]. The difference is small compared to the large uncertainty on the $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ contribution, and is therefore considered as a negligible source of systematic uncertainty. The contribution of $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ decays are therefore included, and constrained, in the fit of the $m_{K\pi\mu\mu}$ distribution of $B^0 \to J/\psi\, K^{*0}$ candidates in the data.

The decay $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ is similarly expected to contribute to the upper $\mathrm{m}_{K\pi\mu\mu}$ mass sideband, however the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ branching fraction has not been measured. By analogy to the decays $B^0 \to J/\psi\, K^{*0}$ and $\overline{B}{}^0_s \to J/\psi\, K^{*0}$, *i.e.*

$$\frac{f_s}{f_d}\frac{\mathcal{B}(\overline{B}{}^0_s \to J/\psi\, K^{*0})}{\mathcal{B}(B^0 \to J/\psi\, K^{*0})} \sim \frac{f_s}{f_d}\frac{\mathcal{B}(\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)} \sim 1\%, \tag{4.3.3}$$

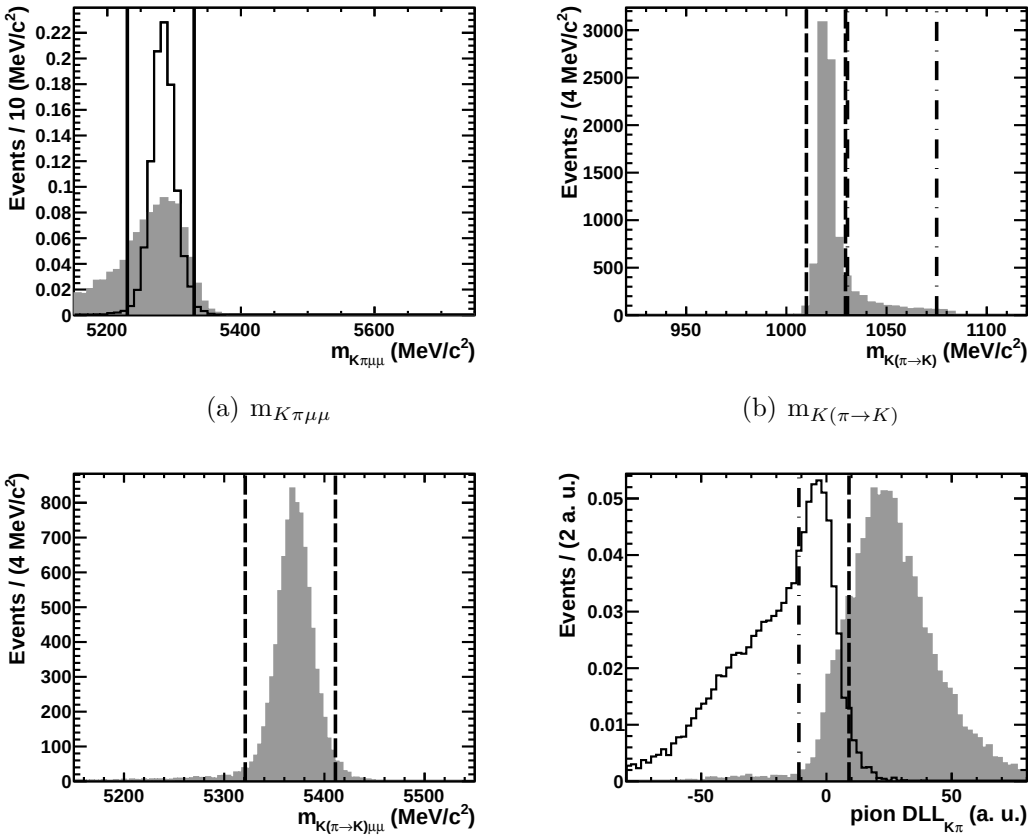
the contribution of $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ is estimated to be $\approx (1 \pm 1)\%$ of the $B^0 \to K^{*0}\mu^+\mu^-$ yield. A large uncertainty is assumed to account for potential NP contributions to the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ decay. This contribution is included in the branching fraction measurement, however owing to the unknown angular distribution, it is ignored in the angular fit. As only $\sim$10 $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ events are expected over the full q^2 range in $1\,\mathrm{fb}^{-1}$ of integrated luminosity, the systematic uncertainty associated with ignoring the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ contribution in the angular fit is neglected in this analysis.

4.3.3 Irreducible contributions from $B^+ \to K^{*+}\mu^+\mu^-$ decays

The decay $B^+ \to K^{*+}\mu^+\mu^-$ contributes to the background when a π^- from elsewhere in the event is added to the $K^{*+} \to K^+\pi^0$ decay (where the π^0 is not reconstructed) to form a four-track final state. These candidates cannot be isolated from $B^0 \to K^{*0}\mu^+\mu^-$ decays by imposing requirements on invariant mass distributions or DLL quantities. Therefore, no veto criteria are imposed to remove $B^+ \to K^{*+}\mu^+\mu^-$ backgrounds. The branching fraction of this decay is well known, therefore the level of this background can be estimated using simulated events.

4.3.4 Contributions from $B^0_s \to \phi\mu^+\mu^-$ decays

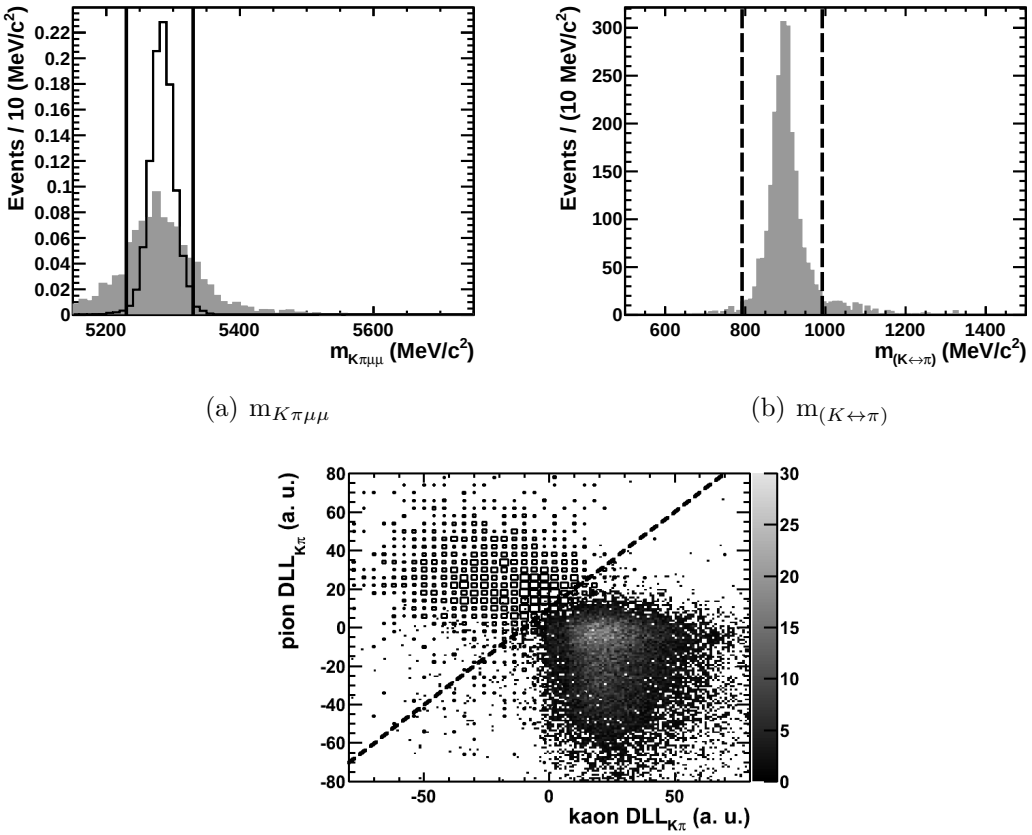
The decay $B^0_s \to \phi(\to K^+K^-)\mu^+\mu^-$ (hereafter denoted $B^0_s \to \phi\mu^+\mu^-$) contributes to the Class I background if the K^- is misidentified as a π^-. These candidates can be separated from signal as, if the π^- is assigned the nominal mass of a K^-, the $K\pi$ invariant mass ($\mathrm{m}_{K(\pi\to K)}$) should then be consistent with that of a ϕ. Similarly, the $K\pi\mu\mu$ invariant mass ($\mathrm{m}_{K(\pi\to K)\mu\mu}$) should then be consistent with that of a B^0_s. Candidates are rejected if $5321 < \mathrm{m}_{K(\pi\to K)\mu\mu} < 5411\,\mathrm{MeV}/c^2$ and either : $1010 < \mathrm{m}_{K(\pi\to K)} < 1030\,\mathrm{MeV}/c^2$ and the pion satisfies $\mathrm{DLL}_{K\pi} > -10$; or $1030 < \mathrm{m}_{K(\pi\to K)} < 1075\,\mathrm{MeV}/c^2$ and the pion satisfies $\mathrm{DLL}_{K\pi} > 10$. These two vetoes are defined in separate regions of $\mathrm{m}_{K(\pi\to K)}$ to reduce the number of $B^0 \to K^{*0}\mu^+\mu^-$ decays that are removed. The distributions of invariant masses and DLL values used to veto this background are shown for $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0_s \to \phi\mu^+\mu^-$ simulated events in Fig. 4.3.



(a) $\mathrm{m}_{K\pi\mu\mu}$ (b) $\mathrm{m}_{K(\pi\to K)}$

(c) $\mathrm{m}_{K(\pi\to K)\mu\mu}$ (d) pion $\mathrm{DLL}_{K\pi}$

Figure 4.3: The distributions of invariant masses and DLL values used to veto $B_s^0\to\phi\mu^+\mu^-$ decays for $B^0\to K^{*0}\mu^+\mu^-$ (black solid histogram) and misidentified $B_s^0\to\phi\mu^+\mu^-$ (grey shaded histogram) simulated events. Solid lines indicate the selected regions; the dashed and dot-dashed lines indicate the veto requirements described in Sec. 4.3.4. The $\mathrm{m}_{K\pi\mu\mu}$ (a), $\mathrm{m}_{K(\pi\to K)}$ (b), $\mathrm{m}_{K(\pi\to K)\mu\mu}$ (c) and pion $\mathrm{DLL}_{K\pi}$ (d) distributions are shown.



(a) $\mathrm{m}_{K\pi\mu\mu}$ (b) $\mathrm{m}_{(K\leftrightarrow\pi)}$

(c) (kaon $\mathrm{DLL}_{K\pi}$, pion $\mathrm{DLL}_{K\pi}$)

Figure 4.4: The distributions of invariant masses and DLL values used to veto misidentified $B^0 \to K^{*0}\mu^+\mu^-$ decays, plotted for $B^0 \to K^{*0}\mu^+\mu^-$ events that are correctly identified (black solid histogram, or grey colour scale) and that are misidentified (grey shaded histogram, or black boxes). Solid lines indicate selected regions; dashed lines indicate veto requirements. The $\mathrm{m}_{K\pi\mu\mu}$ (a) and $\mathrm{m}_{K\pi}$ (b) distributions, and the (kaon $\mathrm{DLL}_{K\pi}$, pion $\mathrm{DLL}_{K\pi}$) plane (c), are shown.

4.3.5 Contributions from $B^0 \to K^{*0}\mu^+\mu^-$ decays

The decay $B^0 \to K^{*0}\mu^+\mu^-$ contributes to the Class I background when the K^+ is misidentified as a π^+ and the π^- is misidentified as a K^-. These misidentified candidates can be separated from correctly identified signal candidates as, if the π^- and K^+ mass assignments are swapped, the $K\pi$ invariant mass ($\mathrm{m}_{(K\leftrightarrow\pi)}$) should then be consistent with that of a K^{*0}. Candidates are rejected if the kaon and pion satisfy kaon $\mathrm{DLL}_{K\pi}$ + 10 < pion $\mathrm{DLL}_{K\pi}$ and $792 < \mathrm{m}_{(K\leftrightarrow\pi)} < 992\,\mathrm{MeV}/c^2$. The distributions of invariant masses and DLL values used to veto $B^0 \to K^{*0}\mu^+\mu^-$ backgrounds are shown for simulated $B^0 \to K^{*0}\mu^+\mu^-$ events in Fig. 4.4.

4.3.6 Contributions from $B^0 \to J/\psi K^{*0}$ decays

The $B^0 \to J/\psi K^{*0}$ decay contributes to the Class I background if the π^- (K^+) is misidentified as a μ^- (μ^+) and the μ^- (μ^+) is misidentified as a π^- (K^+). Candidates where the π^- and μ^- are misidentified can be separated from signal candidates as, if the π^- and μ^- mass assignments are swapped, the $\pi\mu$ invariant mass $\mathrm{m}_{(\pi\to\mu)\mu}$ should be consistent with that of a J/ψ. The equivalent quantity for candidates where the K^+ and μ^+ are misidentified, the $K\mu$ invariant mass $\mathrm{m}_{(K\to\mu)\mu}$, is calculated by swapping the K^+ and μ^+ mass assignments. The candidates are rejected if $3036 < \mathrm{m}_{(\pi\to\mu)\mu} < 3156\,\mathrm{MeV}/c^2$ and the pion satisfies either the IsMuon criteria or $\mathrm{DLL}_{\mu\pi} > 5.0$; or if $3036 < \mathrm{m}_{(K\to\mu)\mu} < 3156\,\mathrm{MeV}/c^2$ and the kaon satisfies either the IsMuon criteria or $\mathrm{DLL}_{\mu\pi} > 5.0$. The distributions of invariant masses and DLL values used to veto $B^0 \to J/\psi K^{*0}$ backgrounds are shown for $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to J/\psi K^{*0}$ simulated events in App. A. The contribution from $\overline{B}{}^0_s \to J/\psi K^{*0}$ decays is negligible.

4.3.7 Contributions from $\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$ decays

The decay $\Lambda_b^0 \to \Lambda^{*0}(1520)(\to p^+K^-)\mu^+\mu^-$ (hereafter denoted as $\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$) contributes to the Class I background if either: the p^+ is misidentified as a π^+; or the p^+ is misidentified as a K^+ and the K^- is misidentified as a π^-. Candidates where the p^+ is misidentified as a π^+ can be separated from signal candidates as, if the π^+ is assigned the nominal mass of a p^+, the $K\pi$ invariant mass $\mathrm{m}_{K(\pi\to p)}$ should then be consistent with that of a Λ^{*0}. Similarly, the $K\pi\mu\mu$ invariant mass $\mathrm{m}_{K(\pi\to p)\mu\mu}$ should then be consistent with that of a Λ_b^0. Candidates are rejected if $5575 < \mathrm{m}_{K(\pi\to p)\mu\mu} < 5665\,\mathrm{MeV}/c^2$, $1490 < \mathrm{m}_{K(\pi\to p)} < 1550\,\mathrm{MeV}/c^2$ and the pion satisfies $\mathrm{DLL}_{p\pi} > 15$.

The equivalent quantities for candidates where the p^+ is misidentified as a K^+ and the K^- is misidentified as a π^-, defined as $\mathrm{m}_{(K\to p)(\pi\to K)}$ and $\mathrm{m}_{(K\to p)(\pi\to K)\mu\mu}$, are calculated by assigning the π^- and K^+ the nominal mass of a K^- and p^+ respectively. Candidates are rejected if $5575 < \mathrm{m}_{(K\to p)(\pi\to K)\mu\mu} < 5665\,\mathrm{MeV}/c^2$, $1500 < \mathrm{m}_{(K\to p)(\pi\to K)} < 1550\,\mathrm{MeV}/c^2$ and the pion satisfies $\mathrm{DLL}_{K\pi} > 10$. The distributions of invariant masses and DLL values used to veto $\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$ backgrounds are shown for $B^0 \to K^{*0}\mu^+\mu^-$ and $\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$ simulated events in App. A.

4.3.8 Contributions from $B^0 \to \rho^0\mu^+\mu^-$ decays

The $B^0 \to \rho^0(770)(\to \pi^+\pi^-)\mu^+\mu^-$ (hereafter denoted $B^0 \to \rho^0\mu^+\mu^-$) decay contributes to the Class I background if the π^+ is misidentified as a K^+. The decay proceeds by the same process as the $B^0 \to K^{*0}\mu^+\mu^-$ decay, however, is further suppressed by the ratio of

CKM factors

$$\left(\frac{V_{cd}}{V_{cs}}\right) \approx 0.05 \text{ or } \left(\frac{V_{td}}{V_{ts}}\right) \approx 0.04. \tag{4.3.4}$$

Taking the rate of $\pi \to K$ mis-identification as 10% [52], the contribution of $B^0 \to \rho^0 \mu^+ \mu^-$ decays is calculated to be $\ll 1\%$ of the signal yield, and is therefore neglected.

4.3.9 Contributions from $B^+ \to K^+ \mu^+ \mu^-$ decays

A Class II background can be formed if a π^- from elsewhere in the event is added to a genuine $B^+ \to K^+ \mu^+ \mu^-$ decay to form a four-track final state. Given that $B^+ \to K^+ \mu^+ \mu^-$ decays accumulate at the nominal B^0 mass, this background can only contribute to the upper $\mathrm{m}_{K\pi\mu\mu}$ mass sideband. These candidates are typically asymmetric in $\cos\theta_K$, due to a momentum imbalance between the kaon and pion. This modifies the $\cos\theta_K$ distribution of candidates in the upper mass sideband, causing a mis-measurement of the angular observables. The candidates are removed by requiring $\mathrm{m}_{K\pi\mu\mu} > 5380\,\mathrm{MeV}/c^2$ and $5220 < \mathrm{m}_{K\mu\mu} < 5340\,\mathrm{MeV}/c^2$, where $\mathrm{m}_{K\mu\mu}$ is the $K\mu\mu$ invariant mass. The distributions of invariant masses used to veto $B^+ \to K^+ \mu^+ \mu^-$ backgrounds are shown for $B^0 \to K^{*0} \mu^+ \mu^-$ and $B^+ \to K^+ \mu^+ \mu^-$ simulated events in App. A.

4.3.10 Summary of the specific vetoes

The residual contributions from exclusive B hadron decays to the Class I and Class II backgrounds are evaluated using simulated events, and are summarised in Tab. 4.3. The level of residual background is given in the 'signal' $\mathrm{m}_{K\pi\mu\mu}$ mass region $5230 < \mathrm{m}_{K\pi\mu\mu} < 5330\,\mathrm{MeV}/c^2$ and in the full $\mathrm{m}_{K\pi\mu\mu}$ mass region $5150 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. Any contribution from Class I and Class II backgrounds results in a different angular distribution of selected candidates in the signal and sideband regions, resulting in the mis-measurement of the angular observables. In the signal region, the residual background is at the level of $\sim 4\%$ of the signal yield, with the total signal loss from the veto criteria $\lesssim 0.3\%$. These expectations are calculated after the application of the multivariate selection described in Sec. 4.4. The residual background contributions are considered as a source of systematic uncertainty in Sec. 6.3.2. The $\mathrm{m}_{K\pi\mu\mu}$ distribution of simulated signal events and the residual Class I and Class II background events are plotted in Fig. 4.5.

4.4 Multivariate selection

Reducing the amount of Class III background improves the precision with which the angular observables are measured, and reduces the sensitivity of the measurement to the

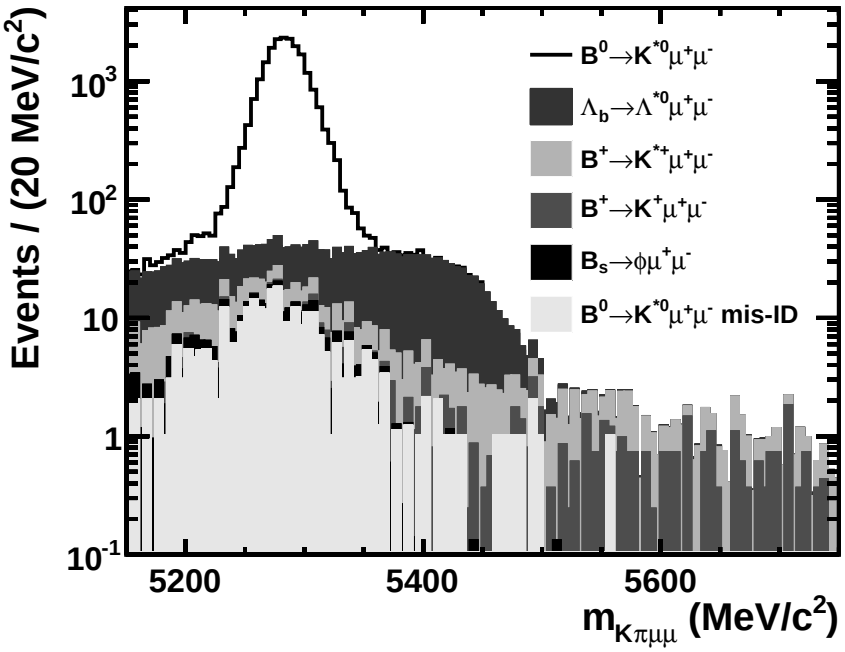


Figure 4.5: The $m_{K\pi\mu\mu}$ distribution of simulated $B^0 \to K^{*0}\mu^+\mu^-$ events, and simulated events of the exclusive decays summarised in Tab. 4.3, after the application of the selection. The contribution of $B^0 \to J/\psi\, K^{*0}$ is negligible and is not included. The contributions are stacked to show the total level of Class I and Class II backgrounds.

Channel	Full $m_{K\pi\mu\mu}$ range (%)	Signal $m_{K\pi\mu\mu}$ range (%)	Signal loss (%)
$\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$	6.89	2.41	0.07
$B^+ \to K^{*+}\mu^+\mu^-$	1.66	0.62	N/A
$B^+ \to K^+\mu^+\mu^-$	0.27	0.05	0.01
$B_s^0 \to \phi\mu^+\mu^-$	0.30	0.18	0.10
$B^0 \to K^{*0}\mu^+\mu^-$ (mis-ID)	1.48	1.05	0.06
$B^0 \to J/\psi\, K^{*0}$	0.00	0.00	0.03
Total	10.60	4.32	0.26

Table 4.3: Expected yields of the exclusive decays that form the Class I and Class II backgrounds, as fractions of the signal yield. The values are extracted in the signal and full $m_{K\pi\mu\mu}$ mass ranges. Also shown is the fraction of simulated $B^0 \to K^{*0}\mu^+\mu^-$ events that are removed by each veto.

modelling of the distribution of background events. This reduction could be achieved by imposing requirements on candidate properties, as in Sec. 4.1. A more powerful technique for separating signal and background candidates is a multivariate analysis (MVA), which exploits the separating power of correlations between a set of variables to produce a multivariate classifier. By imposing a requirement on the classifier response, background-like candidates are removed in the data sample.

The MVA algorithm used for the present analysis is described in Sec. 4.4.1. The procedure to produce the classifier, including a discussion of the input samples, is described in Sec. 4.4.2. The choice of input variables, and the tuning of the BDT parameters, is described in Sec. 4.4.3. Checks for any overtraining and any $\mathrm{m}_{K\pi\mu\mu}$ correlation are described in Sec. 4.4.4. The requirement imposed on the BDT response is optimsed with the procedure detailed in Sec. 4.4.5. The acceptance effects introduced by the BDT are described in Sec. 4.4.6. The $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to K^{*0}\mu^+\mu^-$ yields after application of the selection are extracted in Sec. 4.5.

4.4.1 Description of the MVA algorithm

A variety of MVA algorithms are available, each of which take a sample of signal and background candidates and a set of variables as input. The set of input variables define an n-dimensional variable-space. Each MVA algorithm classifies regions of variable-space as 'signal' or 'background' based on the properties of the candidates from the input samples. The various MVA algorithms differ primarily in their use of the input variables and the resulting complexity of the classifier.

The treatment of candidate weights is an important difference between the MVA algorithms. In this analysis, candidates in both the signal and background input samples have weights that differ from unity. Due to the weighting procedure detailed in Sec. 4.4.2, the weights can be negative. This limits the choice of MVA algorithm to those that can handle such weights. Each MVA algorithm also differs in the ability to exploit non-linear correlations between input variables, which are present for many of the variables considered in this analysis. The Boosted Decision Tree (BDT) algorithm is chosen as it can handle both candidates with negative weights and also exploit non-linear correlations between variables [69, 70].

A BDT is formed from the combination of many Decision Tree (DT) classifiers. A DT is a simple multivariate classifier that applies a sequence of criteria to separate regions of variable space in which there are predominantly signal or background candidates. The sequence of criteria is obtained by attempting to achieve the optimal separation of the signal and background input samples. A DT begins with a root node. A requirement is imposed that splits the variable space, and each input sample, into two parts. This

process is called 'splitting'. Each part of the variable space, and the input samples, is then represented by a new (sub-) node in the tree. One of the resulting (sub-) nodes is enriched with signal candidates and the other is enriched with background. During the splitting process, multiple requirements on each input variable are trialled. A requirement is placed on a single input variable, and is chosen to maximise the increase in separation of signal and background. This is calculated by comparing the G_{ini} index at the node to the sum of the G_{ini} indices at the two sub-nodes. The G_{ini} index is defined as

$$\mathrm{G}_{ini} = 2(purity)(1 - purity) = 2 \times \frac{s}{s+b} \times \frac{b}{s+b} = \frac{2sb}{(s+b)^2}, \tag{4.4.1}$$

where s (b) is the weighted amount of signal (background) at a given node. The largest increase in separation during a splitting corresponds to the largest reduction in the G_{ini} index, since this index tends to zero as the sample becomes dominated by either signal or background candidates. The splitting process is repeated at each node until no improvement in separation is obtained. Each node at which no splitting is beneficial is defined as a 'leaf' node.

Each leaf node of the DT represents a hypercube in variable space that is defined by the sequence of requirements that produced the leaf node. Each leaf node is identified as 'signal' or 'background' depending on the majority of candidates from the input samples at the node. The DT classifies candidates as 'signal' or 'background' depending on the leaf node that their properties determine they fall within. A signal (background) candidate in a region of the variable space that the DT classifies as background (signal) is 'misclassified'.

An advantage of the DT is that it can be written down explicitly, for example, a single DT which forms part of the BDT that is used in this analysis is shown in Fig. 4.6. Another advantage of the DT is that variables with little separation power are ignored and do not affect the performance of the DT. A disadvantage of the DT is that it is not robust against statistical fluctuations in the input data samples.

To overcome this limitation, a 'boosting' procedure is used. Boosting involves increasing the weight of each misclassified candidate in the input samples. In the MVA software used for this analysis, only the 'Adaboost' algorithm can handle candidates with negative weights, so this algorithm is used [71]. The AdaBoost algorithm calculates the error fraction err_{m} for the m^{th} DT as

$$err_{\mathrm{m}} = \frac{w_{\mathrm{mc}}}{w_{\mathrm{tot}}}, \tag{4.4.2}$$

where w_{mc} is the sum-of-weights of all misclassified candidates and w_{tot} is the sum-of-

weights of all candidates. The 'boost' associated with the m^{th} DT (α_{m}) is then calculated as

$$\alpha_{\text{m}} = \beta \log \left[\frac{1 - err_{\text{m}}}{err_{\text{m}}} \right], \tag{4.4.3}$$

where β is a tunable boosting-strength parameter. As err_{m} approaches zero (0.5), indicating the perfect (random) classification of candidates by the DT, α_{m} approaches infinity (zero). The weights of all misclassified candidates are then increased by

$$w'_{\text{i}} = w_{\text{i}} \times e^{\alpha_{\text{m}}}, \tag{4.4.4}$$

where w_{i} is the weight, and w'_{i} is the boosted weight, of the i^{th} candidate. The weights are then renormalised according to

$$w'_{\text{i}} = w'_{\text{i}} \times \frac{\sum_{i=1}^{N} w_{\text{i}}}{\sum_{i=1}^{N} w'_{\text{i}}}. \tag{4.4.5}$$

A new DT is produced using the modified set of weights. By repeatedly boosting and producing DTs, a 'forest' of DTs is generated. The BDT classifier is constructed by combining the classification of each DT in the forest. The BDT response $T(i)$ for a candidate i is calculated according to

$$T(i) = \sum_{m=1}^{N_{tree}} \alpha_{\text{m}} I_m(i), \tag{4.4.6}$$

where $I_m(i)$ is equal to 1 or -1 depending on whether the m^{th} DT classifies candidate i as signal or background. If a candidate is classified by every DT in the forest as signal (background), then $T(i)$ takes values of $+1$ (-1).

The boosting procedure effectively fluctuates the input samples to each DT, making the BDT robust against statistical fluctuations in the input samples. This improves the BDT performance when there is a small number of candidates in the input samples. As the boosting increases the influence of misclassified events in each DT production, DTs produced later in the process are more focused on correctly classifying candidates that were previously misclassified. This considerably improves the performance of a BDT compared to that of a single DT.

A disadvantage of the BDT is that the performance is determined by a large number of tunable parameters, the optimum setup of which is dependent on both the input samples and the choice of input variables. The BDT classifier is also particularly sensitive to

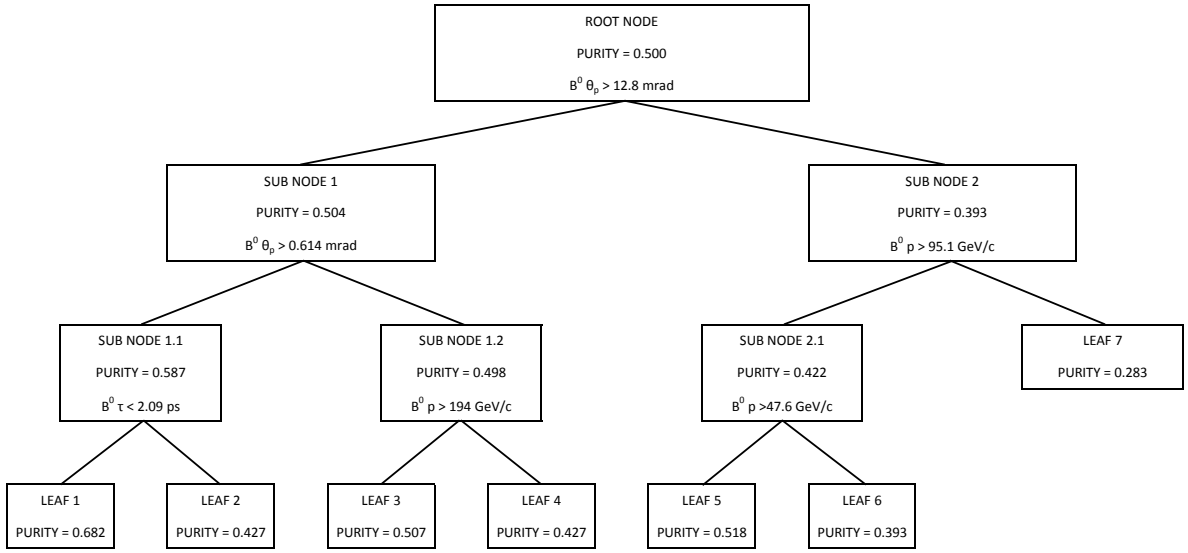


Figure 4.6: The 213^{th} decision tree of the BDT. At each node the signal fraction and cut-variable pair is indicated. At each splitting signal events move to the left side of the decision tree.

overtraining. Overtraining is caused by the properties of a classifier being determined by a small number of candidates in the input samples, rather than general properties of the signal and background candidates. Overtraining occurs when an overly complex classifier is produced with limited input samples. The BDT is particularly sensitive to this effect due to the large number of nodes in the forest. Overtraining increases the BDT performance on the input data samples, while reducing the performance on any independent sample. Furthermore, the classifier response can be correlated to quantities such as $\mathrm{m}_{K\pi\mu\mu}$ that are not input explicitly into the MVA algorithm. This arises as MVA classifiers exploit correlations between the input variables. Correlation of the BDT response to $\mathrm{m}_{K\pi\mu\mu}$ would modify the $\mathrm{m}_{K\pi\mu\mu}$ distribution and introduce a bias in the angular fit.

4.4.2 Procedure to generate the BDT

The BDT is produced using the TMVA package [72]. A sample of $B^0 \to J/\psi\, K^{*0}$ decays in the data is used for the signal input sample. The $B^0 \to J/\psi\, K^{*0}$ decay has the same topology and final state as $B^0 \to K^{*0}\mu^+\mu^-$, but with a branching fraction around one hundred times larger [1]. The $B^0 \to J/\psi\, K^{*0}$ decay differs kinematically from $B^0 \to K^{*0}\mu^+\mu^-$ due to the intermediate J/ψ resonance. A sample of $B^0 \to J/\psi\, K^{*0}$ decays in the data are isolated by removing the J/ψ and $\psi(2S)$ vetoes applied in Sec. 4.3.1 and requiring candidates to instead satisfy $3036 < \mathrm{m}_{\mu\mu} < 3156\,\mathrm{MeV}/c^2$. Treatment of background contamination in this sample is described below. Candidates in the data in the upper $\mathrm{m}_{K\pi\mu\mu}$

mass sideband are used for the background input sample.

The $B^0 \to J/\psi K^{*0}$ candidates used for the signal input sample are excluded from the angular fit and cannot introduce any bias. In contrast, the candidates that are used for the background input sample are also used in the angular fit. To avoid introducing any bias, two BDTs are produced using two separate halves of the full data sample. The BDT produced using the first (second) half, denoted BDT A (BDT B), is applied to the second (first) half. This allows all candidates in the upper $\mathrm{m}_{K\pi\mu\mu}$ mass sideband to be used both to produce the BDT and in the angular fit. The response of the two BDTs is compared in Sec. 4.4.4.

The signal input sample is taken from data and therefore contains contamination from background decays of the form $B_X \to J/\psi X$. To account for this contribution, a weight is computed that defines the influence of each candidate when producing the BDT. The weights, so-called 'sWeights', are calculated using the ${}_s\mathcal{P}lot$ technique [73], which relies on a fit to a control variable, in this case the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution, to calculate the signal probability of each candidate. The sWeights are used to modify the weight of each candidate to obtain signal-like distributions in each of the input variables. Some candidates obtain negative weights through this procedure.

The ${}_s\mathcal{P}lot$ technique requires the variables of interest, in this case the BDT input variables, to be uncorrelated with the control variable. The correlation of each BDT input variable to $\mathrm{m}_{K\pi\mu\mu}$ was extracted using candidates in the data that are distributed across $5350 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. The linear correlation coefficient $\rho_{X,Y}$ is defined for two variables X and Y as

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}, \tag{4.4.7}$$

where $\mathrm{cov}(X,Y)$ is the covariance of the two variables and σ_X is the variance of variable X. The linear correlation coefficient is calculated between each variable used to produce the BDT and $\mathrm{m}_{K\pi\mu\mu}$. Each variable is correlated to $\mathrm{m}_{K\pi\mu\mu}$ at a level of $\lesssim 1\%$, except for the B^0 p and p_T, which are correlated at the level of $\sim 6\%$. Correlations at this level should not significantly affect the production of the BDT.

The candidate weights are further modified to account for the presence of more than one candidate in each event. Multiple candidates can be produced in a single event when, for example, two candidates with different mass hypotheses for the same track both satisfy the pre-selection and trigger requirements. This occurs in about 1% of events. A weight modification of the form

$$w_\mathrm{i} \to w_\mathrm{i} \times 1/N_\mathrm{can} \tag{4.4.8}$$

is applied, where w_i is the candidate weight and N_can is the number of selected candidates in the event.

The disadvantage of the multiple candidate weight is that genuine $B^0 \to J/\psi\, K^{*0}$ candidates are given a reduced weight, and background candidates are not removed from the signal input sample. A BDT produced using an input sample that contains events with multiple candidates might not be optimal. To circumvent this, a preliminary pair of BDTs are produced using the multiple candidate weight. The response of this pair of BDTs is used to create a new input sample where only the candidate with the highest response in each event is retained. A second pair of BDTs are then produced using the reduced input samples, now applying only the sWeights and not the multiple candidate weight. It is the second pair of BDTs that are used in the remainder of the analysis.

4.4.3 BDT optimisation

The variations of the efficiency of the BDT as a function of the decay angles and q^2 are defined as the 'acceptance effects' of the BDT. The separation power of the BDT depends most strongly on the quantities given as input variables, so to correctly reproduce the BDT acceptance effects in simulated $B^0 \to K^{*0}\mu^+\mu^-$ events, the distributions and correlations of each input variable must be the same in data and in simulated events. For simplicity, and as increasing the number of input variables increases the potential for discrepancy between the BDT response in data and in simulated events, a BDT produced using only a few variables is highly desirable.

The input variables to the BDT are chosen in the following way. A BDT is produced using a small number of properties that define an initial variable set $V_{\rm ini}$. Variables, or sets of variables, are added to $V_{\rm ini}$ and a new BDT produced. The performance of the two BDTs are then compared. Any variables that do not improve the BDT performance with respect to $V_{\rm ini}$ are excluded. If a variable, or a set of variables, significantly improves the BDT performance, then the acceptance effects of the BDT with and without the variables are compared. If the variables do not increase the acceptance effects by more than a few percent, and do not introduce any structure in the acceptance effects that is difficult to model using polynomial functions, then the variable is accepted.

The size of the systematic uncertainty associated with the acceptance effects increases with the size of the acceptance effects. The initial set of variables are therefore chosen to introduce minimal acceptance effects.

The largest acceptance effects are introduced if requirements are imposed on any single particle in the candidate, so no properties of the final state particles (K, π or either μ) are included. To prevent the BDT exploiting the kinematic differences between $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to K^{*0}\mu^+\mu^-$, and therefore introducing a large acceptance effect in q^2, properties of the J/ψ resonance are not included. The properties of the K^{*0} resonance are highly correlated to $\mathrm{m}_{\mu\mu}$, so these properties are also excluded. Any variable in which

the distribution of events differs significantly in data and in simulated events is also excluded. The initial BDT input variable set V_{ini} then consists only of the following properties of the B^0 candidate: B^0 vertex quality χ^2/dof, B^0 θ_{p}, B^0 p, B^0 p_{T}, and the B^0 candidate lifetime (τ), defined as

$$\tau = \frac{\Delta_{(\text{SV}-\text{PV})} \times m_0}{p}, \tag{4.4.9}$$

where $\Delta_{(\text{SV}-\text{PV})}$ is the spatial separation of the SV and PV, m_0 is the nominal B^0 mass, and p is the B^0 momentum.

The DLL variables are trialled as potential input variables, as these are expected to have good separation between signal and background. Four BDTs are produced to explore the $\text{DLL}_{K\pi}$ variable: one using only V_{ini}, one using V_{ini} and the pion $\text{DLL}_{K\pi}$ ($V_{+\pi}$); one using V_{ini} and the kaon $\text{DLL}_{K\pi}$ (V_{+K}); and one using V_{ini} and both the pion and kaon $\text{DLL}_{K\pi}$ simultaneously ($V_{+K+\pi}$). The performance of the BDTs are compared by extracting the signal efficiency and background rejection, with rejection defined as $1 - efficiency$, as a function of the requirement imposed on the BDT response. This curve is plotted for the four BDTs in Fig. 4.7. There is an improvement in the BDT performance from adding the kaon $\text{DLL}_{K\pi}$ quantity. The pion $\text{DLL}_{K\pi}$ quantity also improves the BDT performance, but to a lesser extent. The best performance is obtained when including both the kaon and pion $\text{DLL}_{K\pi}$ variables.

In Fig. 4.8 the acceptance effect in $\cos\theta_K$ is compared for the BDTs produced using V_{ini}, V_{+K} and $V_{+K+\pi}$. The efficiency is plotted in bins of $\cos\theta_K$ having imposed the requirement on the BDT response obtained in Sec. 4.4.5. The efficiency varies at the level of $\sim$10% in $\cos\theta_K$ for the BDT produced using V_{ini}, as shown in Fig. 4.8(a). The BDT is more efficient on $B^0 \to K^{*0}\mu^+\mu^-$ candidates with $\cos\theta_K \sim \pm 1$. This feature occurs as candidates in these regions of the angular space typically contain a low momentum kaon or pion, and therefore require an unusually high B^0 p and p_{T} to be reconstructed and satisfy the pre-selection and trigger requirements.

By introducing the kaon $\text{DLL}_{K\pi}$ variable, the efficiency variation across $\cos\theta_K$ is reduced. By introducing kaon and pion $\text{DLL}_{K\pi}$ simultaneously, the efficiency variation reduces further. In both cases, the average efficiency is increased, as expected from the curves shown in Fig. 4.7. The efficiency variation across $\cos\theta_K$ is reduced by these variables as the efficiency increase occurs in the regions of $\cos\theta_K$ where previously the efficiency was lowest, as shown in Fig. 4.8. The kaon and pion $\text{DLL}_{K\pi}$ variables are therefore included into the initial variable set.

Other variables are also trialled as potential input variables for the BDT. These include: the $\text{DLL}_{\mu\pi}$ variables for each of the final state particles (the signal efficiency and

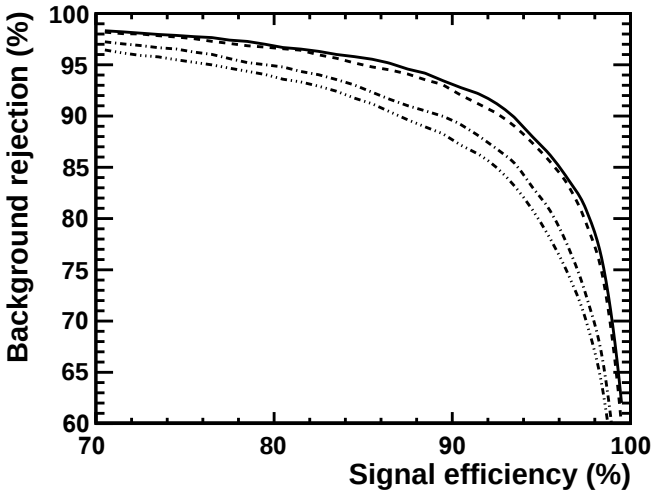
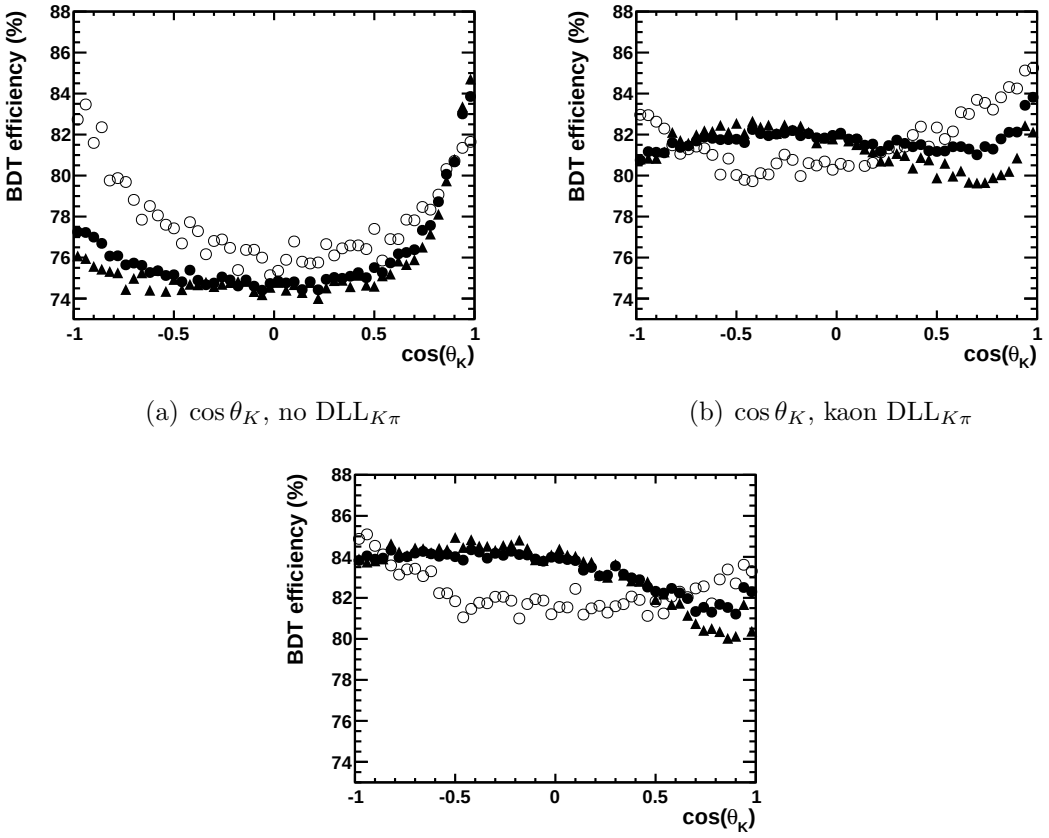


Figure 4.7: Curve of the signal efficiency and background rejection (1 - efficiency) as a function of the requirement imposed on the BDT response, for BDTs produced using: only the initial variable set (triple-dot-dashed line); the initial variable set with K $\mathrm{DLL}_{K\pi}$ added (dashed line); the initial variable set with pion $\mathrm{DLL}_{K\pi}$ added (dot-dashed line); the initial variable set with kaon and pion $\mathrm{DLL}_{K\pi}$ added (solid line). The best performing BDT corresponds to that produced with the initial variable set with kaon and pion $\mathrm{DLL}_{K\pi}$ added (solid black line).

background rejection plots are shown in App. A); the vertex separation χ^2 and vertex quality χ^2/dof of the B^0, J/ψ and K^{*0}; the IP χ^2, p and p_{T} of the B^0, J/ψ, K^{*0}, and each of the four final state particles; and the track quality χ^2/dof for each of the four final state particles. Each was rejected, either due to an insufficient improvement in the performance or due to the excessive size of the introduced acceptance effects.

The final variable set includes seven input variables: B^0 vertex quality χ^2/dof, B^0 θ_{p}, B^0 τ, B^0 p, B^0 p_{T}, kaon $\mathrm{DLL}_{K\pi}$ and pion $\mathrm{DLL}_{K\pi}$. The BDT uses properties of the B^0 candidate and the kaon and pion $\mathrm{DLL}_{K\pi}$ information. The distribution of these quantities for the signal and background input samples are displayed in Fig. 4.9 and Fig. 4.10. The variables are compared in data and in simulated events in Sec. 5.3.1.

As MVAs exploit the differences in the variable correlations between the input samples in order to separate signal and background, the linear correlation coefficient between pairs of input variables, $\rho_{X,Y}$ (defined in Eq. (4.4.7)), can indicate why a given variable enhances the separation power of the classifier. This information can only give a simplified view of the non-linear correlations present in the data, nevertheless it gives an indication of the correlations to which the BDT is sensitive. The difference between the signal and background samples is plotted in Fig. 4.11. Each variable has at least one value of $\rho_{X,Y}$ that differs by more than 5% between signal and background. The improvement in BDT performance from the kaon and pion $\mathrm{DLL}_{K\pi}$ variables is related to a significant difference



(a) $\cos\theta_K$, no $\mathrm{DLL}_{K\pi}$

(b) $\cos\theta_K$, kaon $\mathrm{DLL}_{K\pi}$

(c) $\cos\theta_K$, kaon and pion $\mathrm{DLL}_{K\pi}$

Figure 4.8: The efficiency of selecting simulated $B^0 \to K^{*0}\mu^+\mu^-$ events when imposing a requirement on BDT response in bins of $\cos\theta_K$, for BDTs produced using: only the initial variable set (a); the initial variable set with kaon $\mathrm{DLL}_{K\pi}$ added (b); the initial variable set with kaon and pion $\mathrm{DLL}_{K\pi}$ added (c). The efficiency is plotted in the q^2 ranges $0.10 < q^2 < 19\,\mathrm{GeV}^2/c^4$ (solid circles), $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$ (triangles) and $14.00 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ (open circles). The variation in efficiency across $\cos\theta_K$ is reduced, and the total efficiency is increased, by adding the kaon and pion $\mathrm{DLL}_{K\pi}$ variables.

Variable	Importance
B^0 vertex quality χ^2/dof	0.25
B^0 θ_{p}	0.17
B^0 p_{T}	0.14
kaon $\mathrm{DLL}_{K\pi}$	0.14
pion $\mathrm{DLL}_{K\pi}$	0.13
B^0 τ	0.09
B^0 p	0.09

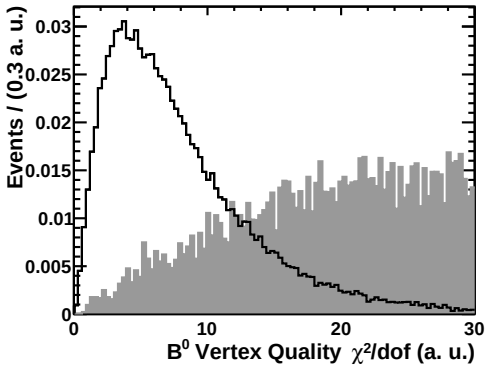
Table 4.4: Variable importance ranking for the BDT.

in $\rho_{X,Y}$ for the variable pairs (kaon $\mathrm{DLL}_{K\pi}$, B^0 p) and (pion $\mathrm{DLL}_{K\pi}$, kaon $\mathrm{DLL}_{K\pi}$), at 11% and 8% respectively.

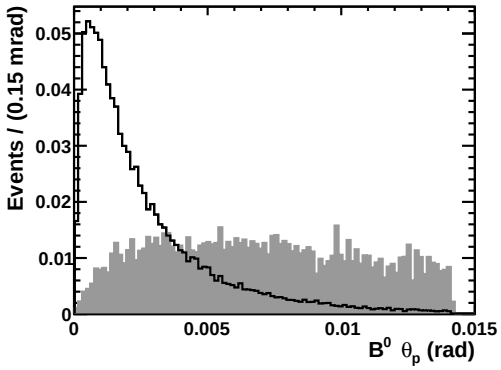
The importance of each variable in the BDT is shown in Tab. 4.4. The ranking quantity is equal to the number of splittings made on each variable weighted by the improvement in separation of the splitting. The order of the BDT input variables is consistent with the dominant source of background events originating from randomly combining tracks from other B hadron decays in the same event. This results in variables such as B^0 vertex quality χ^2/dof and B^0 θ_{p} being highly ranked, while variables such as B^0 τ are ranked lower.

Further configuration of the BDT is driven by three considerations: controlling the amount of overtraining; controlling the time taken to produce the BDT (T_{prod}) and to calculate the BDT response on a given candidate (T_{calc}); and controlling the impact of candidates with negative weights on the BDT production. The BDT configuration is chosen by trialling different values of the following five BDT parameters: NTrees, AdaBoost β, MaxDepth, NCuts and NEventsMin. Each of these parameters are discussed below. The optimal configuration is defined as that which minimises the amount of overtraining without degrading the separation performance by more than 1%, based on the area under the BDT performance curve, while keeping T_{prod} and T_{calc} within reasonable limits.

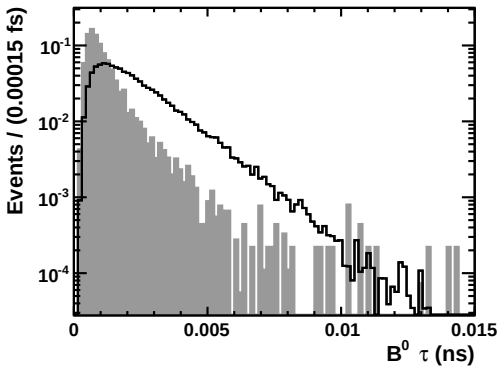
The parameter that effects the BDT performance most significantly is the number of DTs in the forest (NTrees). This parameter improves the separation power of the BDT, as more DTs are used to form the BDT classifier. However, both T_{prod} and T_{calc} increase linearly with NTrees. NTrees also directly increases the number of nodes in the forest and therefore increases the amount of overtraining. The amount of overtraining is controlled by the AdaBoost β parameter, as defined in Eq. (4.4.3). This parameter is used to reduce the strength of the boost applied to each candidate in the boosting



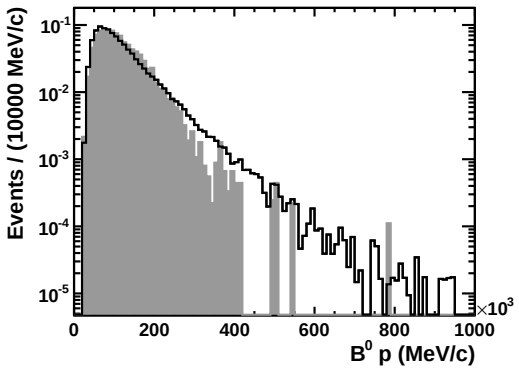
(a) B^0 vertex quality χ^2/dof



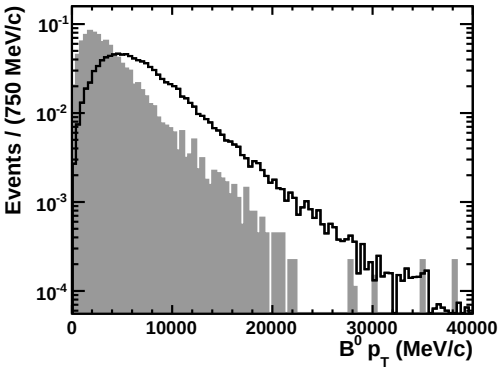
(b) B^0 $\theta_{\rm p}$



(c) B^0 τ

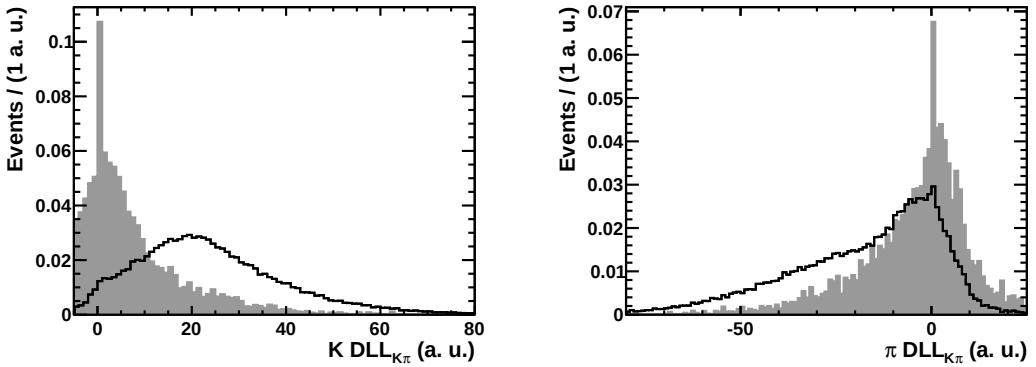


(d) B^0 p



(e) B^0 $p_{\rm T}$

Figure 4.9: Distributions of the properties of the B^0 candidate used as input variables for the BDT. Both the signal (black solid histogram) and background (grey shaded histogram) input samples are shown normalised to equal area.



(a) K DLL$_{K\pi}$ (b) π DLL$_{K\pi}$

Figure 4.10: Distributions of the kaon and pion DLL$_{K\pi}$ values used as input variables for the BDT. Both the signal (black solid histogram) and background (grey shaded histogram) input samples are shown normalised to equal area.

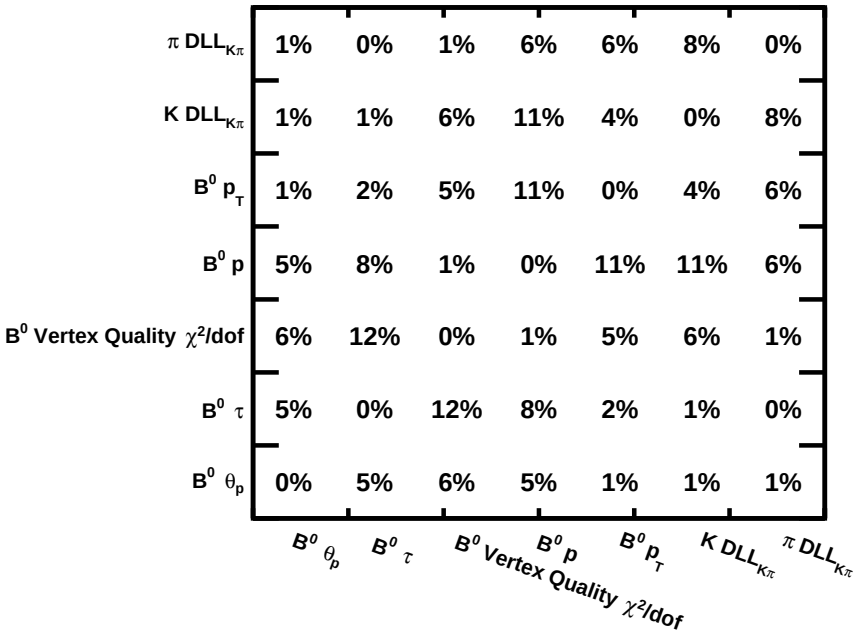


Figure 4.11: The difference in linear correlation coefficient, as defined in Eq. (4.4.7), of all pairs of input variables between then signal and background input samples.

procedure. The optimal value of these two parameters is NTrees = 400 and $\beta = 0.05$.

The BDT properties are also determined by a further two parameters, the maximum number of sequential splittings each DT can make (MaxDepth) and the number of requirements that are trialled on each input variable during the splitting procedure (NCuts). Increasing MaxDepth allows finer segmentation of the variable-space, increasing the BDT performance. However this also increases the overtraining of the BDT, as it typically introduces more nodes to the forest, and increases both T_{prod} and T_{calc}. Increasing NCuts also allows finer segmentation of the variable-space, increasing the BDT performance. Again this increases the overtraining of the BDT, as the requirements can become too specific to the candidates in the input samples. This also increases T_{prod}. The optimal value is MaxDepth = 3, which produces a relatively simple DT. This is consistent with independent work, which shows better performance with small, simple DTs than large, complex DTs in which statistically insignificant nodes are subsequently removed [72]. No significant performance gain is observed by varying NCuts, and so is set to the default value of NCuts = 20, found to give a good balance between performance and T_{prod} [72].

As the signal input sample contains candidates with negative weights, the sum-of-weights at a splitting can be negative if the splitting is made based on a small number of candidates. To avoid these issues, the number of candidates required at each leaf (NEventsMin) is calculated according to

$$\text{NEventsMin} = N_{\text{cand}}/10N_{\text{var}}^2, \tag{4.4.10}$$

where N_{cand} is the number of candidates in the input samples and N_{var} is the number of input variables, and obtains NEventsMin ≈ 240. No significant change in the BDT performance is observed when varying this parameter.

4.4.4 BDT qualification

To check for overtraining of the two BDTs, the response of each BDT on signal and background candidates is compared in two data samples: the samples used to produce the BDT; and independent samples consisting of candidates in the data that are not used to produce the BDT. The response distributions on candidates from these two samples are plotted for BDT A in Fig. 4.12. The equivalent plot for BDT B is in App. A. The distributions of signal and background events are the same in both samples, for both of the BDTs. There is no evidence of overtraining in either of the BDTs.

The response of BDT A and BDT B on a sample of simulated $B^0 \to K^{*0}\mu^+\mu^-$ events is compared in Fig. 4.13. The strong positive correlation ($\approx 96\%$) between the two distributions indicates a similar response of the two BDTs. The BDT efficiency (ϵ_{BDT}) is calcu-

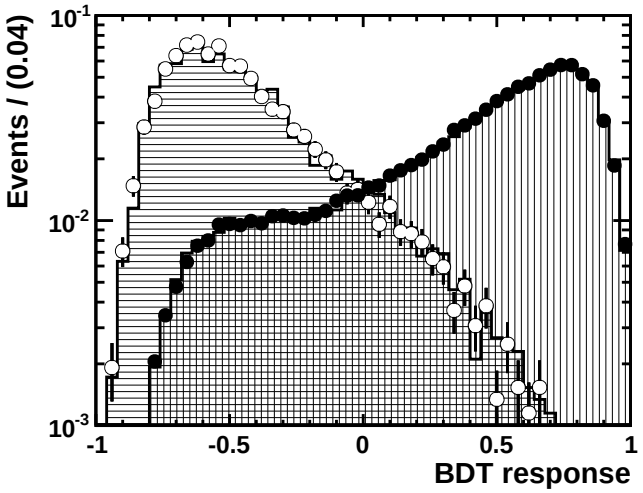


Figure 4.12: The response of BDT A on the signal (black circles) and background (open circles) input samples. The response of BDT A is also shown on signal (vertical striped histogram) and background (horizontally striped histogram) samples that are not used to produce the BDT. The distributions are normalised to equal area.

lated having applied the cut on the response of each BDT defined in Sec. 4.4.5, and is determined using simulated $B^0 \to K^{*0}\mu^+\mu^-$ events. The BDT efficiencies are $\epsilon_{\rm BDTA} \approx 83.6\%$ and $\epsilon_{\rm BDTB} \approx 82.8\%$, which are consistent at a level of $\sim 1\%$.

The correlations between input variables can result in a correlation between the BDT response and $\mathrm{m}_{K\pi\mu\mu}$. Given the definition of lifetime in Eq. (4.4.9), a correlation between BDT response and $\mathrm{m}_{K\pi\mu\mu}$ may arise through the correlations between the lifetime, vertex separation χ^2 and p of the B^0. Any correlation will modify the $\mathrm{m}_{K\pi\mu\mu}$ distribution and therefore bias the angular fit. Some correlation is expected, however, as the BDT uses variables such as B^0 p and $p_{\rm T}$, which are directly correlated to $\mathrm{m}_{K\pi\mu\mu}$. The mean of the BDT response distribution on candidates in the data is plotted in bins of the upper $\mathrm{m}_{K\pi\mu\mu}$ mass sideband to check for any correlation between BDT response and $\mathrm{m}_{K\pi\mu\mu}$. The results are shown in Fig. 4.14. There is no evidence for correlation between the BDT and $\mathrm{m}_{K\pi\mu\mu}$.

4.4.5 Using the BDT information

To reduce the number of (background) candidates input to the angular fit, a minimum requirement is imposed on the BDT response. The requirement is optimised to achieve the maximum significance of the $B^0 \to K^{*0}\mu^+\mu^-$ signal peak, defined as

$$m = s/\sqrt{s+b}, \tag{4.4.11}$$

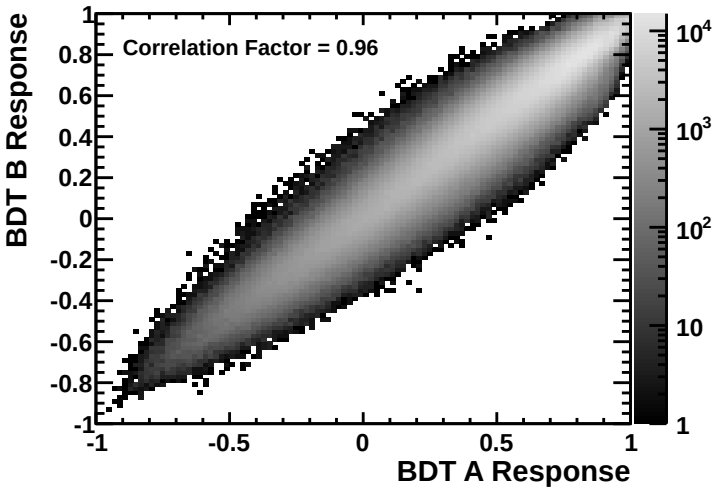


Figure 4.13: Comparison of BDT A and BDT B on simulated $B^0 \to K^{*0}\mu^+\mu^-$ events. The linear correlation coefficient of the two distributions is $\approx 96\%$.

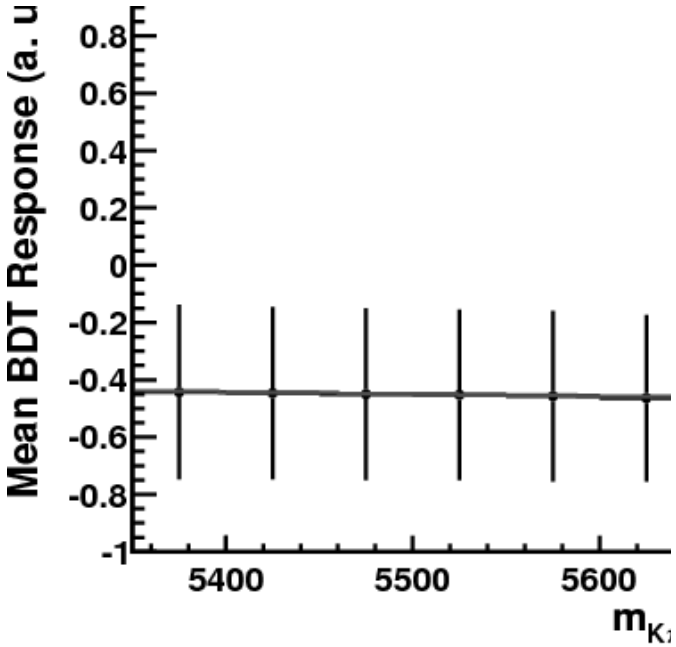
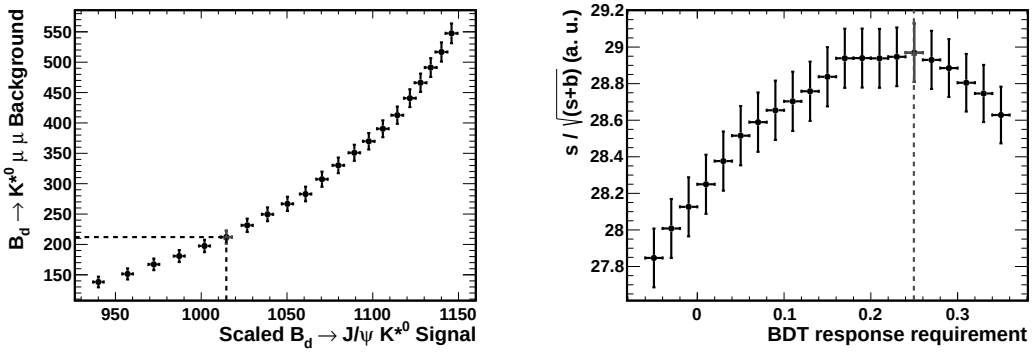


Figure 4.14: Mean of the BDT response distribution is shown in $50\,\mathrm{MeV}/c^2$ bins for candidates in the data in the upper $\mathrm{m}_{K\pi\mu\mu}$ mass sideband. The standard deviation of the BDT response distribution in the bin is taken as the uncertainty on each point.



(a) Yield curve

(b) Significance curve

Figure 4.15: Signal and background yields (a) and signal significance (b) for each trialled BDT requirement. Appropriately scaled $B^0 \to J/\psi\, K^{*0}$ candidates are used to estimate the $B^0 \to K^{*0}\mu^+\mu^-$ yield. The optimal cut value, and the corresponding (scaled) signal and background yields, are indicated.

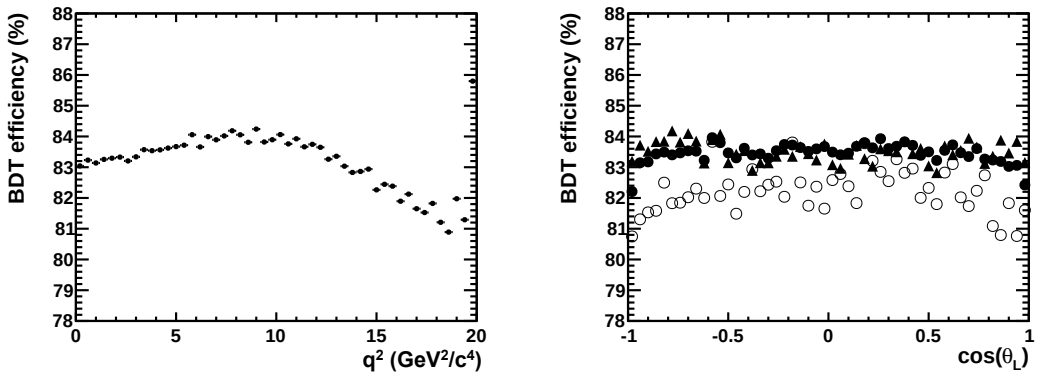
where s (b) is the number of signal (background) candidates in the signal region $5230 < \mathrm{m}_{K\pi\mu\mu} < 5330\,\mathrm{MeV}/c^2$.

The yield of $B^0 \to J/\psi\, K^{*0}$ candidates in the data is extracted using a fit to the $\mathrm{m}_{K\pi\mu\mu}$ distribution. The number of signal candidates, s, is then calculated by scaling the yield obtained from the $B^0 \to J/\psi\, K^{*0}$ fit by the ratio of measured $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to J/\psi\, K^{*0}$ branching fractions and the corresponding ratio of efficiencies in simulated events. The number of background events, b, is extracted from a fit to candidates in the data in the upper $\mathrm{m}_{K\pi\mu\mu}$ mass sideband. A range of requirements on the BDT response are trialled. The values of s and b, and the calculated significance, are shown for each requirement in Fig. 4.15. The optimum requirement is $BDT > 0.25$, which obtains $s \approx 1015$, $b \approx 210$ and $m \approx 29$.

4.4.6 BDT acceptance effects

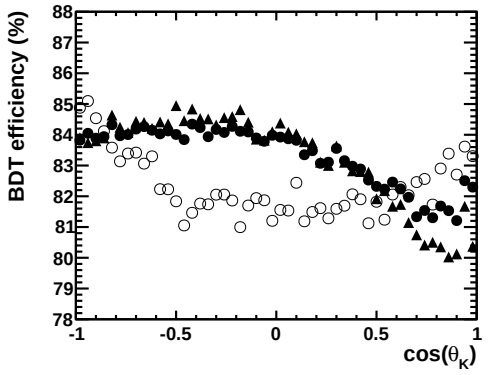
Having imposed the BDT response requirement defined in Sec. 4.4.5, the acceptance effects introduced by the BDT are shown in Fig. 4.16. The q^2 distribution displays a slow variation in efficiency at the level of 3%. The $\cos\theta_L$ and ϕ angles show a q^2 dependent effect: at low q^2 the efficiency varies less than 1%; at high q^2 the variation is around 2%. The $\cos\theta_K$ distribution shows a larger variation: 4-5% depending on the q^2 range. The variation is asymmetric in $\cos\theta_K$ due to the momentum imbalance between the kaon and pion, caused by the kaon taking a larger proportion of the K^{*0} momentum when boosing into the lab frame, since the mass of the kaon is larger than that of the pion. Since both muons have the same mass, the variation in $\cos\theta_L$ is symmetric.

The variations in efficiency have been controlled to the level of a few percent. This is

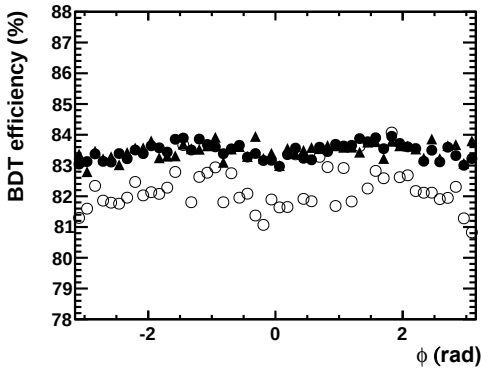


(a) q^2

(b) $\cos\theta_L$



(c) $\cos\theta_K$



(d) ϕ

Figure 4.16: The acceptance effects of BDT A for q^2 (a), $\cos\theta_L$ (b), $\cos\theta_K$ (c) and ϕ (d). For (b)-(d), the effects are shown for events where $0.10 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ (solid circles), $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$ (open circles) and $14.00 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ (triangles).

small compared to variations introduced from other sources, such as the detector geometry. Any systematic uncertainty related to the acceptance effects is therefore not sensitive the variations introduced by the BDT.

4.5 Selection results

The selection, comprising the pre-selection, trigger, and BDT response requirement, is applied to the full data sample. A fit to $B^0 \to J/\psi\, K^{*0}$ candidates, using the mass model described in Sec. 6.2.1, is shown in Fig. 4.17. The fit returns a $B^0 \to J/\psi\, K^{*0}$ yield of $\approx 102400 \pm 300$.

The same fit model is used to fit the $B^0 \to K^{*0}\mu^+\mu^-$ candidates. Only the signal fraction and background shape parameter are allowed to vary between the $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to K^{*0}\mu^+\mu^-$ fits. The $B^0 \to K^{*0}\mu^+\mu^-$ mass fits for candidates in the q^2 bins $0.10 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ and $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$ are shown in Fig. 4.18. The

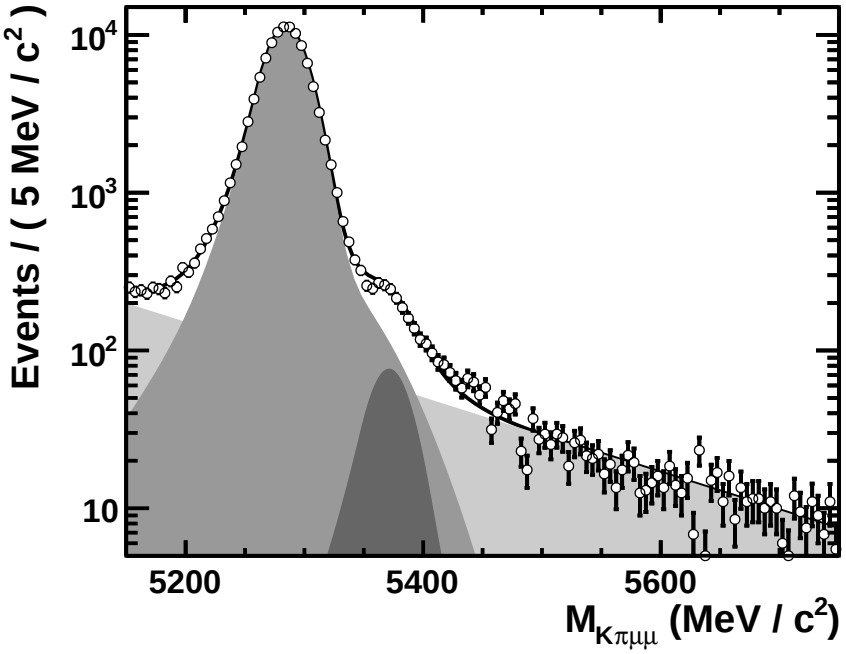
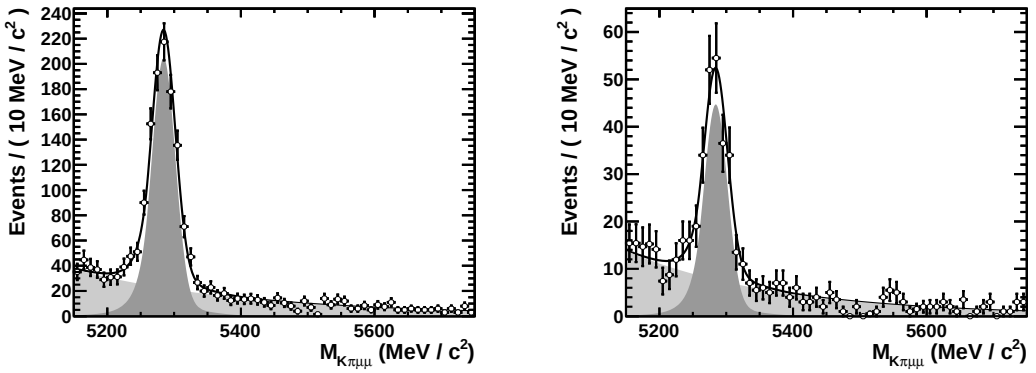


Figure 4.17: The $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of $B^0 \to J/\psi\, K^{*0}$ candidates in the data after applying the selection (black open points) in the mass range $5150 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. A fit (black solid line) to the data is overlaid, with the combinatoric (light grey area), $B^0 \to J/\psi\, K^{*0}$ (grey area) and $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ (dark grey area) components indicated. The fit model is defined in Sec. 4.4.5.

$B^0 \to K^{*0}\mu^+\mu^-$ mass fits in each analysis q^2 bin are shown in Fig. 4.19. The signal and background yields in the mass range $5230 < \mathrm{m}_{K\pi\mu\mu} < 5330\,\mathrm{MeV}/c^2$ are summarised in Tab. 4.5. The yield of $B^0 \to K^{*0}\mu^+\mu^-$ candidates is larger than that obtained at any previous experiment. The ratio of the background yield to the signal yield $b/s \approx 0.25$ indicates that the sample is also more pure than that of previous experiments.

The distribution of candidates in the $(\mathrm{m}_{K\pi\mu\mu}, \mathrm{m}_{\mu\mu})$ and $(\mathrm{m}_{K\pi\mu\mu}, m_{K^{*0}})$ planes are shown in Fig. 4.20. There is an accumulation of events at the nominal B^0 mass and across the $\mathrm{m}_{\mu\mu}$ range, and at the nominal (B^0, K^{*0}) masses.

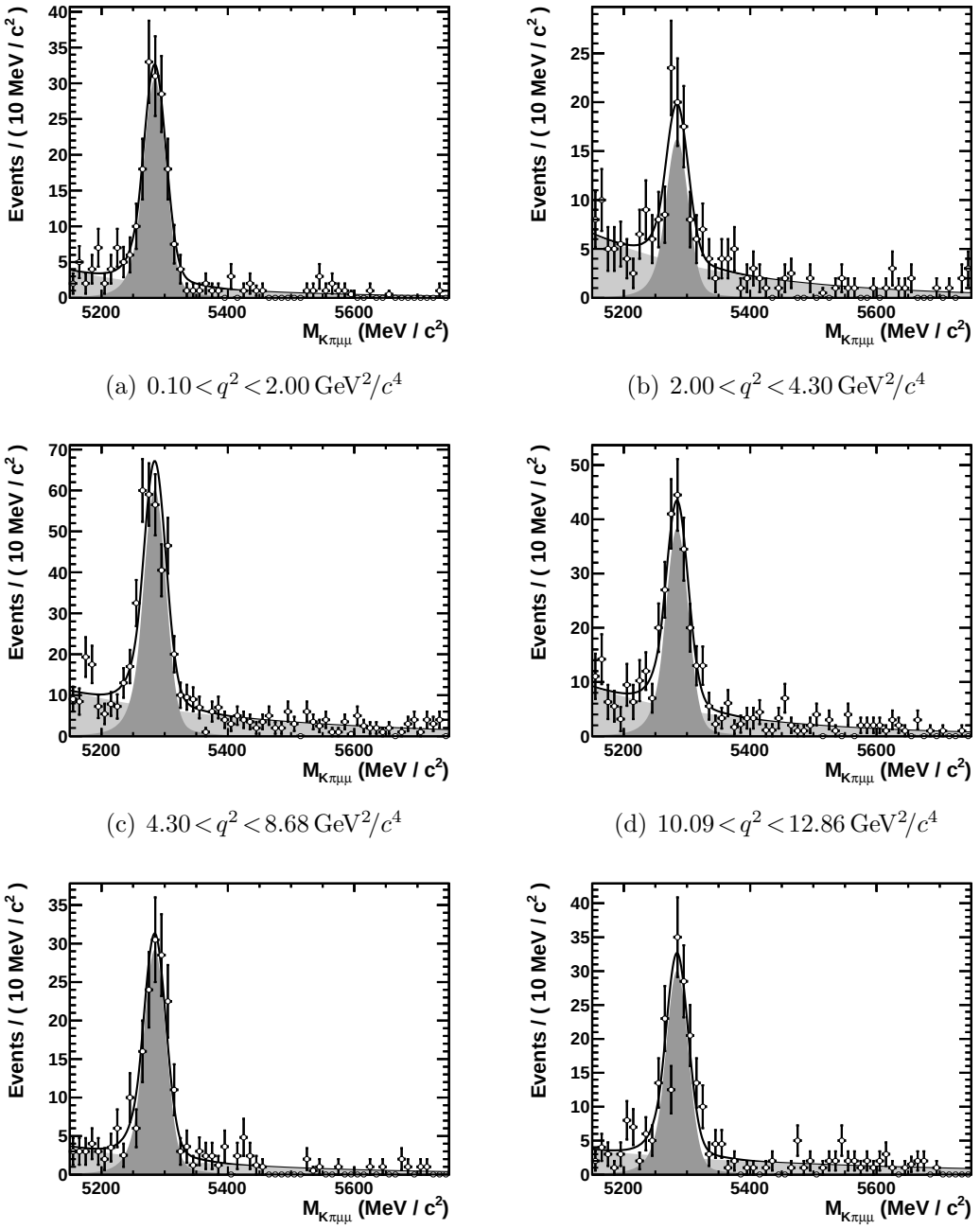


(a) $0.10 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ (b) $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$

Figure 4.18: The $\mathrm{m}_{K\pi\mu\mu}$ distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data after applying the selection (black open points) in the mass range $5150 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$ in the q^2 bins $0.10 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ (a) and $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$ (b). A fit (black solid line) to the data is overlaid, with the combinatoric (light grey area) and $B^0 \to J/\psi\, K^{*0}$ (grey area) components indicated. In each plot the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ component is too small to be visible. The fit model is defined in Sec. 4.4.5.

q^2 range (GeV^2/c^4)	Signal yield	Background yield
$0.10 < q^2 < 2.00$	141.32 ± 12.99	20.18 ± 3.11
$2.00 < q^2 < 4.30$	74.74 ± 10.76	37.20 ± 4.00
$4.30 < q^2 < 8.68$	277.25 ± 19.36	71.25 ± 6.00
$10.09 < q^2 < 12.86$	176.83 ± 15.49	52.37 ± 5.24
$14.18 < q^2 < 16.00$	134.71 ± 12.85	21.22 ± 3.32
$16.00 < q^2 < 19.00$	138.71 ± 13.43	26.03 ± 3.73
$1.00 < q^2 < 6.00$	205.98 ± 17.22	79.07 ± 5.95
$0.10 < q^2 < 19.00$	942.65 ± 35.33	229.19 ± 10.72

Table 4.5: Signal and background yields in the range $5230 < \mathrm{m}_{K\pi\mu\mu} < 5330\,\mathrm{MeV}/c^2$. The yields are obtained from the fits to the $\mathrm{m}_{K\pi\mu\mu}$ distributions shown in Fig. 4.18 and Fig. 4.19.



(a) $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$

(b) $2.00 < q^2 < 4.30\,\mathrm{GeV}^2/c^4$

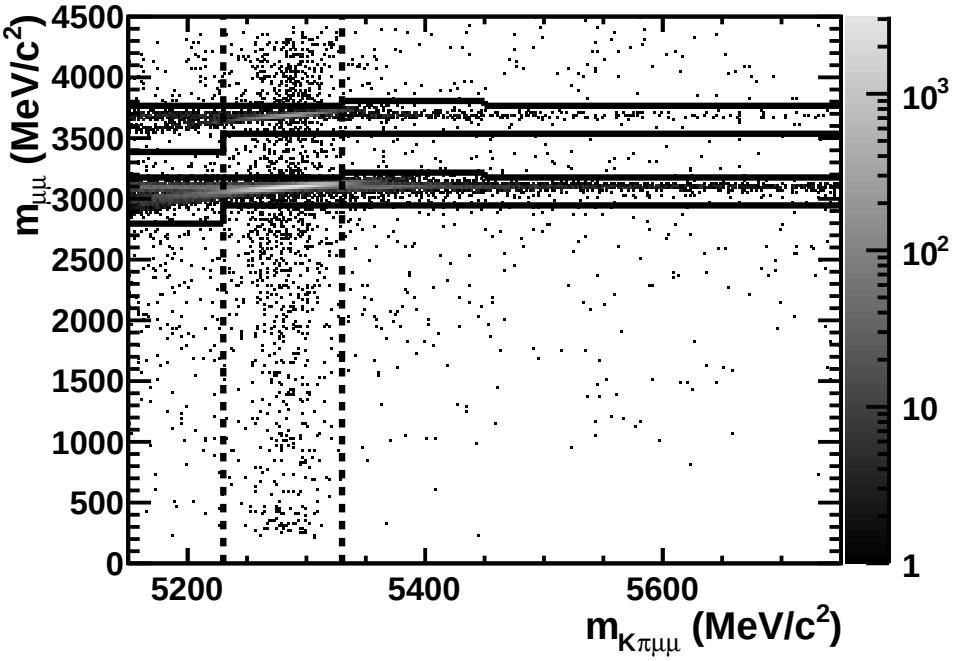
(c) $4.30 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$

(d) $10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$

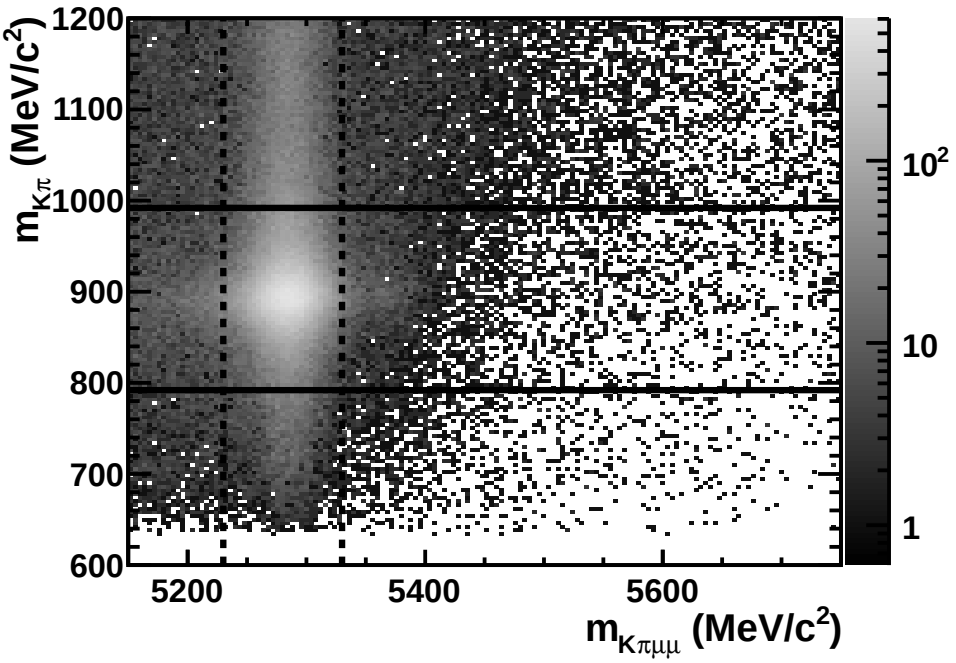
(e) $14.18 < q^2 < 16.00\,\mathrm{GeV}^2/c^4$

(f) $16.00 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$

Figure 4.19: The $\mathrm{m}_{K\pi\mu\mu}$ distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data after applying the selection (black open points) in the mass range $5150 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$ in each of the angular analysis q^2 bins. A fit (black solid line) to the data is overlaid, with the combinatoric (light grey area) and $B^0 \to K^{*0}\mu^+\mu^-$ (grey area) components indicated. In each plot the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ component is too small to be visible. The fit model is defined in Sec. 6.2.1.



(a) ($m_{K\pi\mu\mu}$, $m_{\mu\mu}$) plane



(b) ($m_{K\pi\mu\mu}$, $m_{K\pi}$) plane

Figure 4.20: The ($m_{K\pi\mu\mu}$, $m_{\mu\mu}$) (a) and ($m_{K\pi\mu\mu}$, $m_{K\pi}$) (b) planes after the selection, excluding the J/ψ and $\psi(2S)$ vetoes described in Sec. 4.3.1, is applied. The black dashed line indicates the signal region in $m_{K\pi\mu\mu}$. In (a) the solid lines indicate the J/ψ and $\psi(2S)$ vetoes described in Sec. 4.3.1. In (b) the solid lines indicate the $m_{K\pi}$ requirements defined in Tab. 4.1.

Chapter 5

Acceptance effects

An acceptance effect is a variation in the efficiency to detect, reconstruct and select $B^0 \to K^{*0}\mu^+\mu^-$ decays as a function of a given quantity. Acceptance effects in the decay angles modify the angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates, resulting in a mis-measurement of the angular observables. Since the angular observables are extracted as rate averages in bins of q^2, acceptance effects in q^2 also result in a mis-measurement of the angular observables.

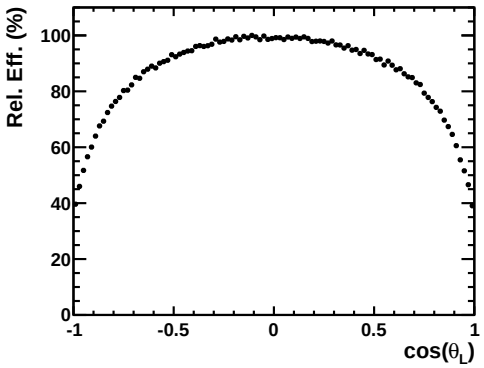
The dominant sources of acceptance effects are the detector geometry and the event reconstruction. Further effects are introduced by the requirements imposed in the pre-selection and trigger. The smallest effect is introduced by the requirement placed on the BDT response.

To accurately measure the angular observables, a correction must be applied to account for the acceptance effects. This acceptance correction (AC) is performed by adjusting the weight of each candidate by an 'AC weight'. The AC weight is defined by the efficiency to detect, reconstruct and select a candidate, given a point in q^2 and the decay angles. The AC weights are calculated by parameterising the acceptance effects in simulated events as a function of the decay angles and q^2. By calculating the efficiency as a function of these four quantities the weights are independent of the physics model used to produce the simulated events.

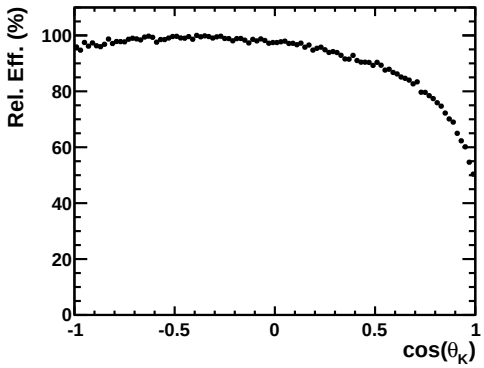
The origin of the acceptance effects are discussed in Sec. 5.1. The acceptance correction procedure is defined in Sec. 5.2 and validated in Sec. 5.3.

5.1 Description of the acceptance effects

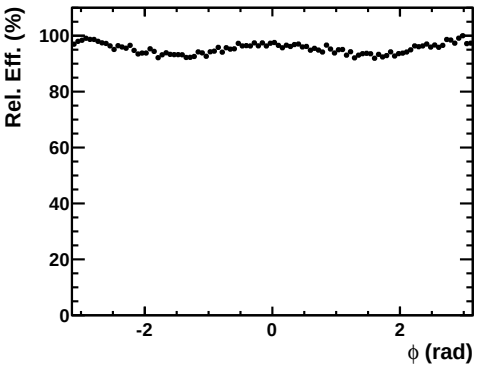
Each of the acceptance effects are dependent on the momentum of the final state particles. The acceptance effects in the decay angles are caused by the accumulation of events with low momentum particles in certain regions of the angular space, particularly at



(a) $\cos\theta_L$



(b) $\cos\theta_K$



(c) ϕ

Figure 5.1: Acceptance effects introduced by the detector geometry and event reconstruction for simulated $B^0 \to K^{*0}\mu^+\mu^-$ events. The angles $\cos\theta_L$ (a) and $\cos\theta_K$ (b) are shown for events with $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$ and ϕ (c) is shown for events with $14.18 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$.

$\cos\theta_L \sim \pm 1$ and $\cos\theta_K \sim \pm 1$. Similarly, the correlation of q^2 and the momenta of the final state particles results in acceptance effects in q^2.

The acceptance effects introduced by the detector geometry and the event reconstruction are detailed in Sec. 5.1.1. The acceptance effects introduced by the pre-selection and trigger are discussed in Sec. 5.1.2. The effects are summarised in Sec. 5.1.3.

5.1.1 Acceptance effects of the detector and event reconstruction

The efficiency to detect and reconstruct a $B^0 \to K^{*0}\mu^+\mu^-$ candidate at LHCb is approximately 3%. As shown in Fig. 5.1, this efficiency varies by $\sim$60% in $\cos\theta_L$ and $\cos\theta_K$ at low q^2, and by $\sim$10% in ϕ at high q^2. These variations are the dominant contribution to the acceptance effects.

(a) di-muon frame (b) lab frame

Figure 5.2: Schematic diagram for the muon system of a $B^0 \to K^{*0}\mu^+\mu^-$ decay at low q^2 and $\cos\theta_L \sim +1$ in the di-muon rest frame (a) and in the lab frame (b). The large momentum asymmetry when boosting from the di-muon to the lab frame is shown.

Decays with low q^2 produce low energy muons. If $\cos\theta_L \sim \pm 1$, one muon is produced approximately in the same direction as the B^0 (forward-going), and the other in the opposite direction (backward-going), in the di-muon rest frame. Upon boosting this system to the lab frame, the forward-going muon obtains a large momentum, while the backward-going momentum obtains a low momentum. In this way, a large momentum asymmetry between the muons is formed in decays with $\cos\theta_L \sim \pm 1$. This effect is shown schematically in Fig. 5.2.

At low q^2, decays with $\cos\theta_L \sim \pm 1$ tend to have one low momentum muon. Such decays are less likely to have both muons travel within the geometric acceptance of the detector, since a muon must have $p > 3\,\text{GeV}/c$ to traverse the magnetic field and reach the Muon system. Furthermore the event reconstruction uses the IsMuon criteria to identify muons, which imposes the momentum-dependent requirement detailed in Tab. 3.1. A fraction of decays at low q^2 and $\cos\theta_L \sim \pm 1$ are therefore removed by the IsMuon criteria. The resulting relative efficiency distribution is shown in Fig. 5.1(a).

An analogous situation occurs for the kaon and pion in the rest frame of the K^{*0}. Decays with $\cos\theta_K \sim +1$ have a forward-going kaon and a backward-going pion. Due to the mass difference of the kaon and pion, the kaon takes a larger portion of the momentum when the $K\pi$ system is boosted into the lab frame. This increases the momentum asymmetry between the two particles in decays with $\cos\theta_K \sim +1$, and reduces the momentum asymmetry in decays with $\cos\theta_K \sim -1$.

Particles with $p \lesssim 1\,\text{GeV}/c$ are bent out of the geometric acceptance of the detector by the magnetic field. This effect tends to remove decays with $\cos\theta_K \sim \pm 1$. Due to the kaon and pion mass difference this effect is asymmetric: the effect is exacerbated

at $\cos\theta_K \sim +1$ where decays have particularly low momentum pions, and is reduced at $\cos\theta_K \sim -1$ where the momentum asymmetry is smaller. The resulting relative efficiency distribution is shown in Fig. 5.1(b).

The geometric acceptance of the detector also removes decays with $\phi \sim \pm\pi/2$ at high q^2. In such decays the decay plane of the $K\pi$ and the di-muon system are oriented perpendicular to each other. This increases the likelihood for either the hadron system or the di-muon being produced at a large angle to the beam axis, increasing the likelihood for one of the final state particles to travel outside the angular acceptance of the detector. The resulting relative efficiency distribution is shown in Fig. 5.1(c).

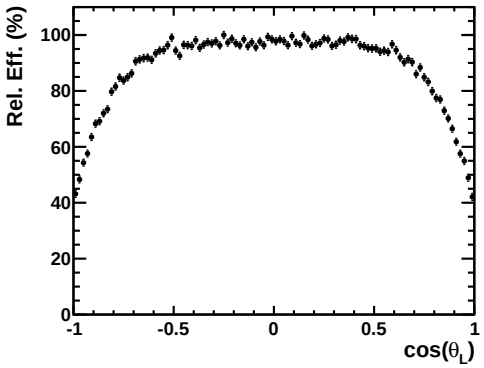
Decays are more frequently removed as q^2 tends to minimal or maximal values due to the momentum spectra of the final state particles at these limits. A typical decay at high q^2 produces high momentum muons and low momentum hadrons, while a typical decay at low q^2 produces low momentum muons and high momentum hadrons. As the acceptance effects introduced by the detector geometry and event reconstruction depend on the momentum of the final state particles, acceptance effects are therefore generated in q^2.

5.1.2 Acceptance effects of the pre-selection and trigger

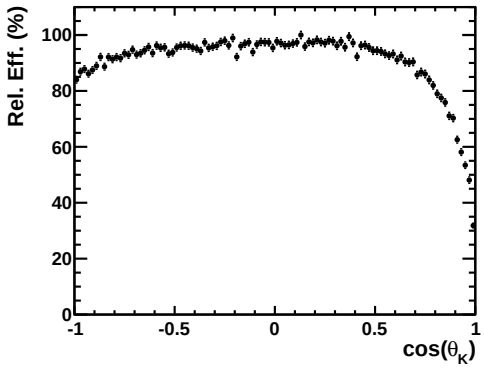
Further acceptance effects are introduced by the pre-selection and trigger, which impose requirements on the kinematic properties of the B^0 meson and the $B^0 \to K^{*0}\mu^+\mu^-$ decay products. The efficiency of the pre-selection and trigger on reconstructed $B^0 \to K^{*0}\mu^+\mu^-$ events is $\sim 30\%$. These acceptance effects are shown on reconstructed events in Fig. 5.3, and are at the level of $\sim 60\%$ in $\cos\theta_L$ and $\cos\theta_K$ at low q^2, and 10% in ϕ at high q^2.

At least one muon in the candidate is required to satisfy the p_{T} threshold imposed by the `L0Muon` trigger, which removes decays at low q^2 with $\cos\theta_L \sim \pm 1$. Furthermore, decays at low q^2, with either $\cos\theta_L \sim \pm 1$ or $\cos\theta_K \sim \pm 1$, typically contain one particle that travels approximately along the direction of flight of the B^0 and therefore has a small IP. Such decays are more frequently removed by the IP χ^2 requirement placed on each of the final state particles in the pre-selection. Furthermore, the candidate must satisfy the requirements on the momentum, p_{T}, IP and IP χ^2 imposed by the HLT1 and HLT2 trigger algorithms detailed in Sec. 4.2. The requirements imposed in the pre-selection and trigger result in the relative efficiency distribution show in Fig. 5.3.

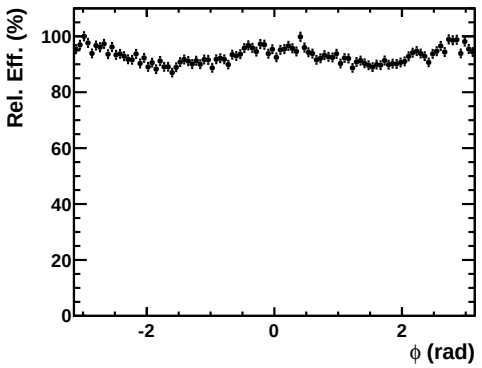
As shown in Fig. 4.5, the multivariate selection is $\sim 80\%$ efficient on candidates that satisfy the pre-selection and trigger requirements and introduces acceptance effects at the level of $\sim 3\%$. The acceptance effects are small due to the choice of input variables. Residual acceptance effects occur due to the correlation of the input variables and properties of the final state particles to which the decay angles are also correlated.



(a) $\cos\theta_L$



(b) $\cos\theta_K$



(c) ϕ

Figure 5.3: Acceptance effects introduced by imposing the pre-selection and trigger requirements on simulated $B^0 \to K^{*0}\mu^+\mu^-$ events. The angles $\cos\theta_L$ (a) and $\cos\theta_K$ (b) are shown for events with $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$ and ϕ (c) is shown for events with $14.18 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$.

5.1.3 Summary of the acceptance effects

The combined acceptance effects in the decay angles and q^2, which are shown in Figs. 5.4 and 5.5 respectively, are at the level of $\sim$85% in the angles $\cos\theta_L$ and $\cos\theta_K$ at low q^2 and $\sim$15% in ϕ at high q^2. The acceptance effects are at the level of $\sim$20% in q^2.

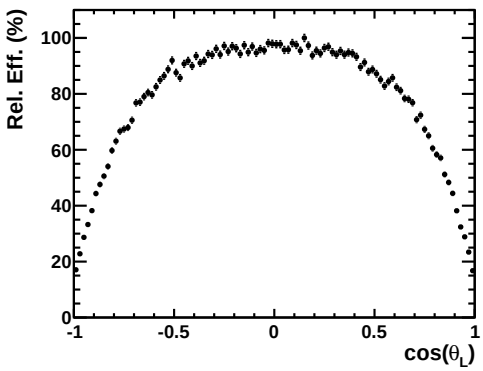
Of every 100 $B^0 \to K^{*0}\mu^+\mu^-$ events, 97 are removed by requiring that each of the final state particles traverses the detector and are subsequently reconstructed, introducing acceptance effects at the level of $\sim$60%. The modelling of these acceptance effects is dependent on the description of both the detector geometry and the momentum spectrum of the B^0, which have been verified to be accurately reproduced in simulated events, and will not be discussed further [50, 51].

The requirements imposed in the pre-selection and trigger remove two of the remaining three $B^0 \to K^{*0}\mu^+\mu^-$ events, and introduce acceptance effects at the level of 60%. The modelling of these acceptance effects is dependent on the distribution of events in each variable on which the pre-selection and trigger impose requirements. These distributions are defined by both the momentum spectrum of the B^0; and the characteristics of the pp interaction, such as the produced number of tracks. The acceptance effects of the pre-selection and trigger are therefore expected to be reproduced less well than those of the detector geometry and event reconstruction. However, after correcting the critical quantities for any difference in the distribution of $B^0 \to J/\psi\, K^{*0}$ decays in data and in simulated $B^0 \to J/\psi\, K^{*0}$ events, the acceptance effects are expected to be reproduced properly in simulated events. The BDT then removes 80% of the remaining $B^0 \to K^{*0}\mu^+\mu^-$ events. Due to the careful choice of input variables, the acceptance effects introduced by the BDT are negligible compared to those described above.

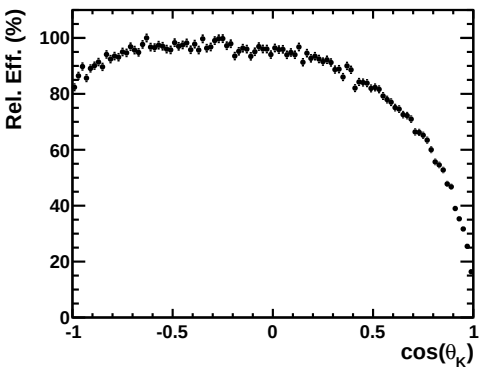
5.2 The acceptance correction

To avoid a mis-measurement of the angular observables the acceptance effects in the decay angles and q^2 must be corrected. The correction is performed by adjusting the weight of each candidate by the inverse of the efficiency of that candidate. The efficiency is calculated as a function of the decay angles and q^2 using simulated $B^0 \to K^{*0}\mu^+\mu^-$ events. By computing the efficiency as a function of all of the quantities that define the $B^0 \to K^{*0}\mu^+\mu^-$ decay, the weights are entirely independent of the physics model used to produce the simulated events. The AC weights are renormalised such that the mean weight is unity, which removes any dependency on the absolute scale of the efficiency.

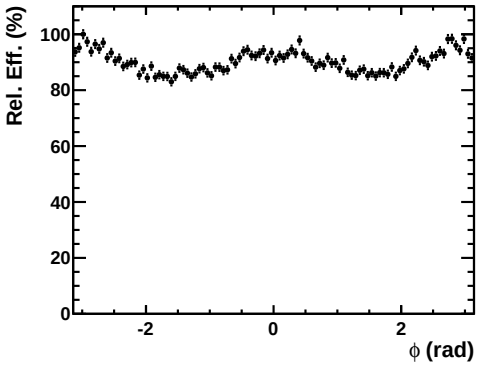
The simulated events that are used to determine the efficiency are generated with a uniform angular distribution, which ensures that the (statistical) uncertainty on the acceptance correction is uniform across the decay angles. For the acceptance correction



(a) $\cos\theta_L$



(b) $\cos\theta_K$



(c) ϕ

Figure 5.4: Combined acceptance effects for simulated $B^0 \to K^{*0}\mu^+\mu^-$ events. The angles $\cos\theta_L$ (a) and $\cos\theta_K$ (b) are shown for events with $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$ and ϕ (c) is shown for events with $14.18 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$.

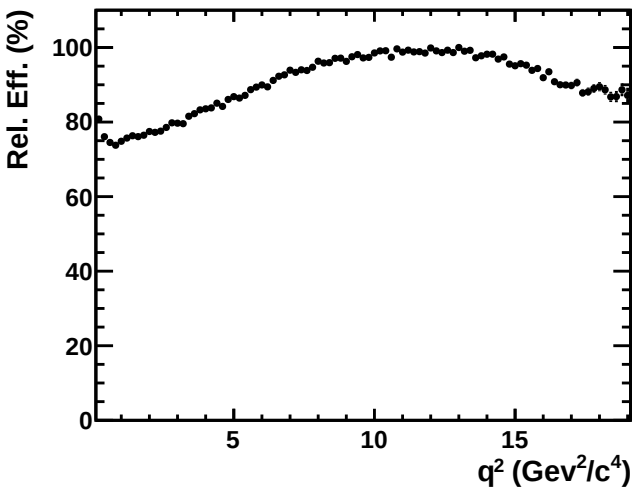


Figure 5.5: Combined acceptance effects in q^2 for simulated $B^0 \to K^{*0}\mu^+\mu^-$ events.

to be valid the simulated events must correctly reproduce the acceptance effects present in the data. This requirement is discussed in detail in Sec. 5.3.

A fit is performed to the simulated events in order to parameterise the acceptance effects in the three decay angles. Although the acceptance effects in q^2 are correlated to those in the decay angles, incorporating a q^2 dependence of the acceptance effects would make the fit considerably more complicated. Instead the fit is performed in bins of q^2, denoted $q^2_{\rm AC}$. As the acceptance effects vary quickly with q^2, on a scale that is many times smaller than the analysis q^2 bins, the $q^2_{\rm AC}$ bins are much finer than the analysis q^2 bins. A total of 60 bins are used. The variation of the acceptance effects is most rapid in the region $0.10 < q^2 < 1.00\,{\rm GeV}^2/c^4$, where a bin size of $0.10\,{\rm GeV}^2/c^4$ is used. Bins of $0.20\,{\rm GeV}^2/c^4$ are used between $1.00 < q^2 < 6.00\,{\rm GeV}^2/c^4$ and bins of $0.50\,{\rm GeV}^2/c^4$ are used in the remainder of the q^2 range. These larger bin sizes are driven by the smaller variation of the acceptance effects at high q^2. The choice of AC binning is considered as a source of systematic uncertainty in Sec. 6.3.2.

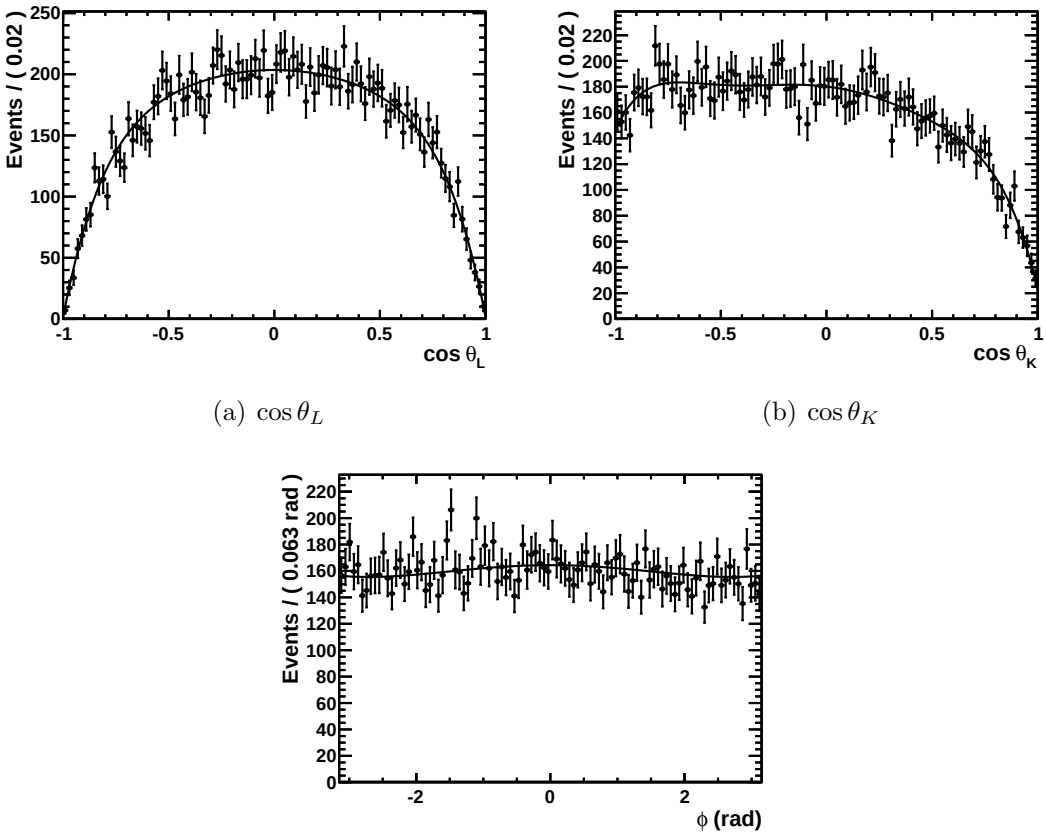
The efficiency in the three decay angles is assumed to factorise so that a 3D parameterisation of the acceptance effects is not necessary. This further reduces the complexity of the AC fit. This factorisation is such that the efficiency of a candidate

$$\epsilon(\cos\theta_L, \cos\theta_K, \phi) \equiv \epsilon(\cos\theta_L) \times \epsilon(\cos\theta_K) \times \epsilon(\phi), \tag{5.2.1}$$

where $\epsilon(\cos\theta_L, \cos\theta_K, \phi)$ is the efficiency of a candidate with respect to the three decay angles, and $\epsilon(X)$ is the efficiency of that candidate with respect to only one of the decay angles $X = \{\cos\theta_L, \cos\theta_K, \phi\}$. The validity of this assumption is examined in Sec. 5.3.

The simulated events are fitted with a $6^{\rm th}$-order polynomial function in each decay angle. The acceptance effects are assumed to be symmetric around $\cos\theta_L = 0$. This assumption is valid if there is no significant CP asymmetry between the B^0 and $\overline{B}^0$, and if there is no difference in efficiency to detect, reconstruct and select a μ^+ or a μ^-. The size of the latter effect is reduced as the magnet polarity is reversed in approximately 40% of the data. The measured amount of CP asymmetry is consistent with zero and the measured difference in the muon efficiency is $\sim$0.2% [74]. The acceptance effects are also assumed to be symmetric around $\phi = 0$, as a given distribution of final state particles can be achieved with either sign of ϕ. To enforce these symmetries in the fit the odd terms in the polynomials that parameterise the $\cos\theta_L$ and ϕ angles are fixed to zero. The acceptance effects in $\cos\theta_K$ are asymmetric due to the difference in the kaon and pion mass.

Acceptance correction pdfs (AC pdfs) are obtained from the fit. The angular distribution of simulated events with $0.50 < q^2 < 0.60\,{\rm GeV}^2/c^4$ and with $12.50 < q^2 < 13.00\,{\rm GeV}^2/c^4$



(a) $\cos\theta_L$

(b) $\cos\theta_K$

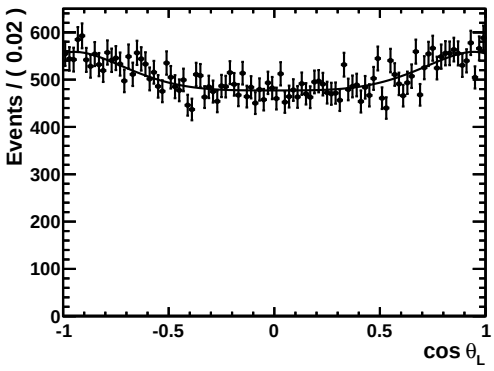
(c) ϕ

Figure 5.6: The distribution of simulated events with $0.50 < q^2 < 0.60\,\mathrm{GeV}^2/c^4$ in the decay angles $\cos\theta_L$ (a), $\cos\theta_K$ (b) and ϕ (c). The fitted AC pdf (solid blue line) is overlaid.

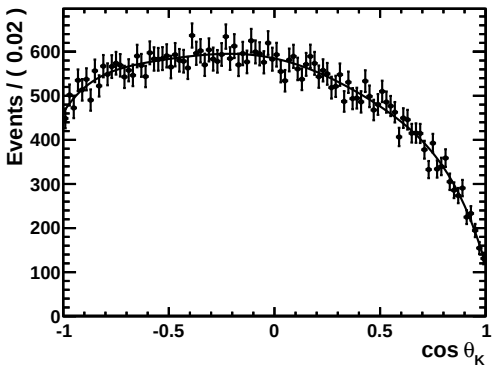
can be seen in Figs. 5.6 and 5.7 respectively. The fitted AC pdfs are indicated. The same plot in the $3.40 < q^2 < 3.60\,\mathrm{GeV}^2/c^4$ region is shown in App. B. The difference in the shape of the acceptance effects in the two q^2 regions is evident, particularly in the $\cos\theta_L$ and ϕ angles. The quality of the fit is explored in Sec. 5.3.

The total efficiency (ϵ_{tot}) of a candidate is calculated by evaluating the AC pdfs at the $\cos\theta_L$, $\cos\theta_K$, and ϕ point of the candidate, and scaling this efficiency by the average efficiency of candidates in the relevant q^2_{AC} bin. The AC weight of the candidate is then defined as the inverse of ϵ_{tot}. The AC weight therefore accounts for variations in the efficiency as a function of the decay angles and q^2. As ϵ_{tot} is typically on the order of $\sim 0.7\%$, the AC weights are on the order of ~ 140.

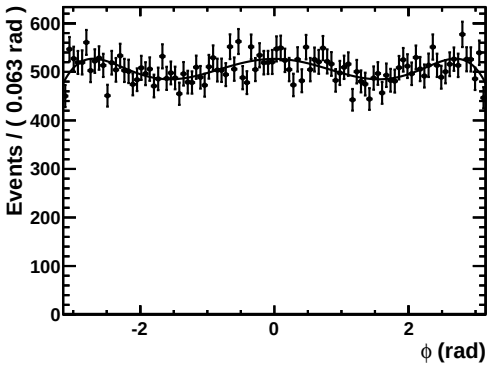
The AC weights are combined with two other weights to obtain the weight for each candidate, which are subsequently used in the measurement of the observables in Secs. 6 and 7. These other weights take into account the presence of multiple candidates in each event using Eq. (4.4.8), and the removal of Class III background by the J/ψ and $\psi(2S)$ vetoes described in Sec. 4.3.1. The weight of each candidate is dominated by the AC



(a) $\cos\theta_L$

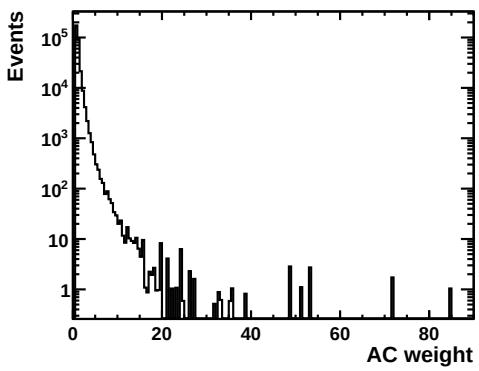


(b) $\cos\theta_K$

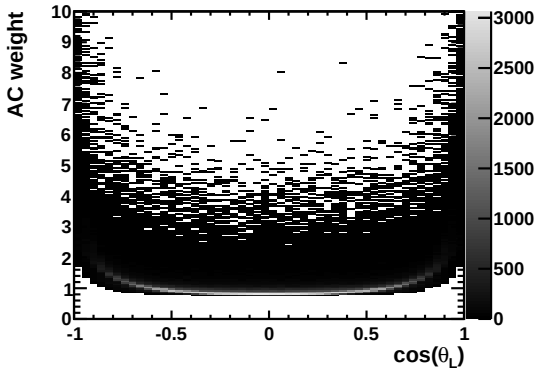


(c) ϕ

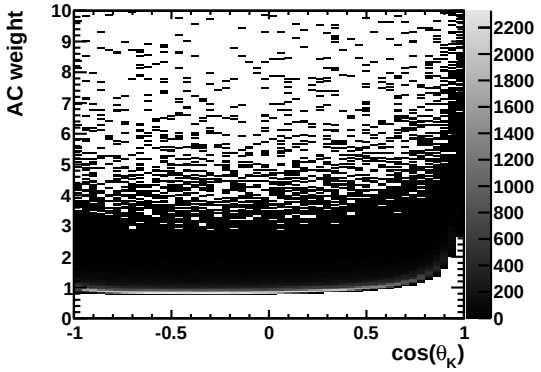
Figure 5.7: The distribution of simulated events with $12.50 < q^2 < 13.00\,\mathrm{GeV}^2/c^4$ in the decay angles $\cos\theta_L$ (a), $\cos\theta_K$ (b) and ϕ (c). The fitted AC pdf (solid blue line) is overlaid.



(a) AC weight



(b) ($\cos\theta_L$, AC weight)



(c) ($\cos\theta_K$, AC weight)

Figure 5.8: Distribution of the AC weights (a), and the AC weights as a function of the decay angles $\cos\theta_L$ (b) and $\cos\theta_K$ (c), for simulated events with $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$.

weight. In the remainder of this thesis, the 'AC weight' will refer the combination of these three weights.

The AC weights are renormalised by a scaling factor α so that the sum-of-weights of candidates in the data sample is equal to the number of candidates in the sample. The value of α is therefore equal to the average efficiency of the candidates. The renormalisation removes the sensitivity of the acceptance correction to the absolute efficiency scale; only the efficiency variations are important. In Sec. 6, α is used to calculate the differential branching fraction, as the absolute efficiency of each candidate is required.

In Fig. 5.8 the distribution of AC weights, and the AC weights as a function of the decay angles $\cos\theta_L$ and $\cos\theta_K$, are shown for simulated events with $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$. There is an increase in the AC weights for events with $\cos\theta_L \sim \pm 1$ and $\cos\theta_K \sim +1$ where the acceptance effects are largest.

5.3 Assuring the validity of the acceptance correction

The validity of the acceptance correction relies on two factors: the correct modelling of the detector geometry and event reconstruction, pre-selection and trigger in simulated events; and the validity of the assumption of the efficiency factorisation in Eq. (5.2.1). These two issues are discussed in Sec. 5.3.1 and Sec. 5.3.2 respectively.

5.3.1 Validation of the modelling of the acceptance effects

To ensure that the acceptance effects introduced by the requirements imposed in the pre-selection and trigger are accurately reproduced in simulated events, any differences in the distribution of events in data and in simulated events in each variable are corrected using the methods described below. Each correction is considered as a source of systematic uncertainty in Sec. 6.3.2.

5.3.1.1 Corrections applied to simulated events

Corrections are applied to a number of variables used in the pre-selection and trigger. The number of tracks in an event (N_{track}) is dependent on the characteristics of the pp interaction. This quantity is corrected by adjusting the weight of simulated events to obtain the same distribution of N_{track} in data and in simulated events. The tracking efficiency is dependent on the alignment of the detector and on N_{track}. This is similarly corrected by adjusting the weight of events to obtain a data-like tracking efficiency in bins of p and η. As the acceptance effects depend on the boost of the B^0, the momentum spectrum of the B^0 is also corrected by adjusting the weight of the simulated events.

The IsMuon efficiency is extracted for muons using a tag-and-probe method with $J/\psi \to \mu^+\mu^-$ candidates in the data [75]. The tag-and-probe technique takes pairs of tracks and imposes requirements on one of the tracks to isolate $J/\psi \to \mu^+\mu^-$ decays from background processes. The IsMuon efficiency is then measured on the other track in the track pair as a function of p and η. The IsMuon efficiency is corrected by adjusting the weight of the simulated events to obtain data-like IsMuon efficiencies in bins of p and η.

The distribution of DLL values depends on the behaviour of the RICH detector systems. The behaviour of this complicated system is correlated to N_{track}, the detector alignment, and the properties of the aerogel and gases. Differences are therefore expected between the DLL distributions in data and in simulated events. The distribution of DLL values is corrected using template DLL distributions in bins of p, η and N_{track}.

Template DLL distributions are obtained in the data: candidates of the $J/\psi \to \mu^+\mu^-$

decay are used to obtain templates for muons; candidates of the $D^{*-} \to \bar{D}^0(\to K^+\pi^-)\pi^-$ decay are used to obtain templates for kaons and pions; and candidates of the $\Lambda^0 \to p^+\pi^-$ decay are used to obtain templates for protons. Each decay is selected without imposing requirements on the DLL variables.

A data-like DLL value is generated for each particle by picking a DLL value at random from the relevant template distribution. As the DLL value is picked randomly, there is no correlation of the kaon $\mathrm{DLL}_{K\pi}$ and the pion $\mathrm{DLL}_{K\pi}$ value within a given event beyond that accounted for by the p, η and N_{track} binning. Since the requirements on the DLL quantities in the pre-selection remove almost no signal, these 'missing' correlations have negligible impact on the analysis.

The IP is corrected by modifying the momentum of each track such that the same distribution of IP as a function of $1/p_{\mathrm{T}}$ is obtained in data and in simulated events. This correction is required due to the inaccurate simulation of particle interactions with the material in the VELO. Unfortunately, the IP correction introduces a shift to larger values in the vertex quality χ^2/dof. To account for these shifts the IP correction is considered as a source of systematic uncertainty in Sec. 6.3.2. The systematic uncertainty from the IP correction is smaller than other systematic uncertainties, and considerably smaller than the statistical uncertainty.

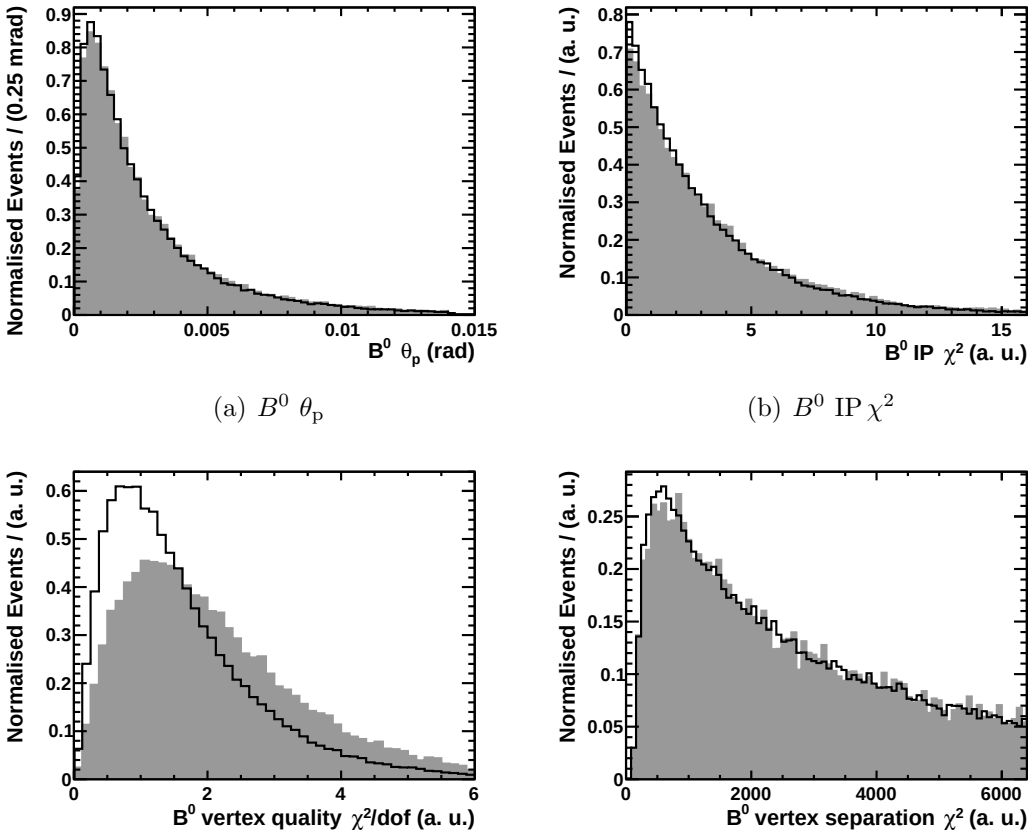
5.3.1.2 Validating the corrections using data

The variables used in the pre-selection and trigger are compared using a sample of $B^0 \to J/\psi\, K^{*0}$ candidates in the data and simulated $B^0 \to J/\psi\, K^{*0}$ events, which are generated with SM-like parameters. The residual background in the $B^0 \to J/\psi\, K^{*0}$ sample is taken into account by applying sWeights [73].

The distribution of events for each B^0 property used in the pre-selection is shown in Fig. 5.9. There are no significant discrepancies in any of the B^0 quantities except the vertex quality χ^2/dof. This discrepancy arises due to the IP correction applied to the simulated events. The distribution of events for other quantities used in the pre-selection are also compared. Any quantity on which the imposed requirement is $\sim$100% efficient on simulated events is not shown, otherwise, the distribution of events is compared in App. B. There are no significant discrepancies in any of these pre-selection quantities.

The `L0Muon` trigger decision is based on a threshold imposed on the p_{T} of a muon in the event. Furthermore, the muon p_{T} is also exploited in the algorithms of the HLT that accept the majority of the events that are used in this analysis. The distribution of events for the p_{T} of the μ^+, and for the μ^-, is shown in Fig. 5.10. There is no discrepancy between the distribution of events in data and in simulated events.

To ensure that discrepancies in the BDT response in data and in simulated events



(a) B^0 $\theta_{\rm p}$

(b) B^0 IP χ^2

(c) B^0 vertex quality χ^2/dof

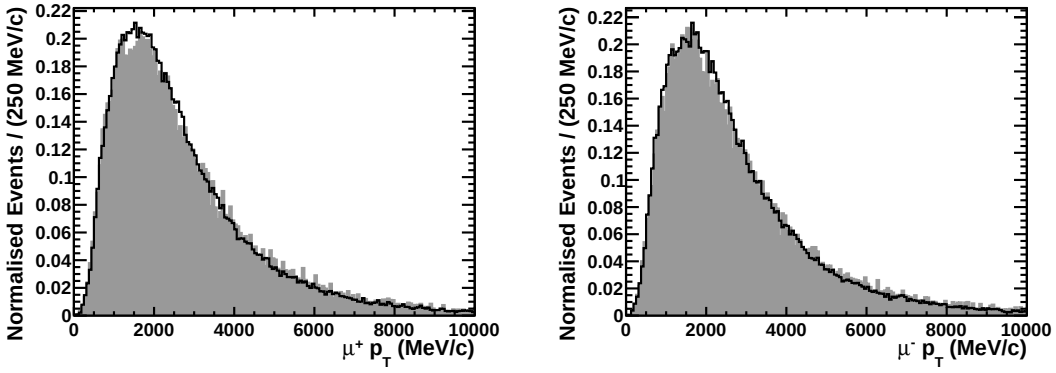
(d) B^0 vertex separation χ^2

Figure 5.9: Comparison of the B^0 variables used in the pre-selection in $B^0 \to J/\psi\, K^{*0}$ candidates in the data (black solid histogram) and in simulated $B^0 \to J/\psi\, K^{*0}$ events (grey shaded histogram), normalised to equal area.

cannot impact the analysis, the distribution of events for the B^0 properties and DLL quantities that are used in the BDT, but are not used in the pre-selection, are shown in Fig. 5.11 and Fig. 5.12 respectively. Residual discrepancies are observed at low B^0 $p_{\rm T}$ and in the DLL distributions. These discrepancies are considered as sources of systematic uncertainty in Sec. 6.3.2. The response of the BDT is also compared, as shown in Fig. 5.13. Imposing a $BDT > 0.25$ requirement introduces a difference in the absolute efficiency of $\sim$4%, which is primarily a result of the IP and DLL corrections. This discrepancy causes a negligible $\sim$0.1% difference in the combined acceptance effects.

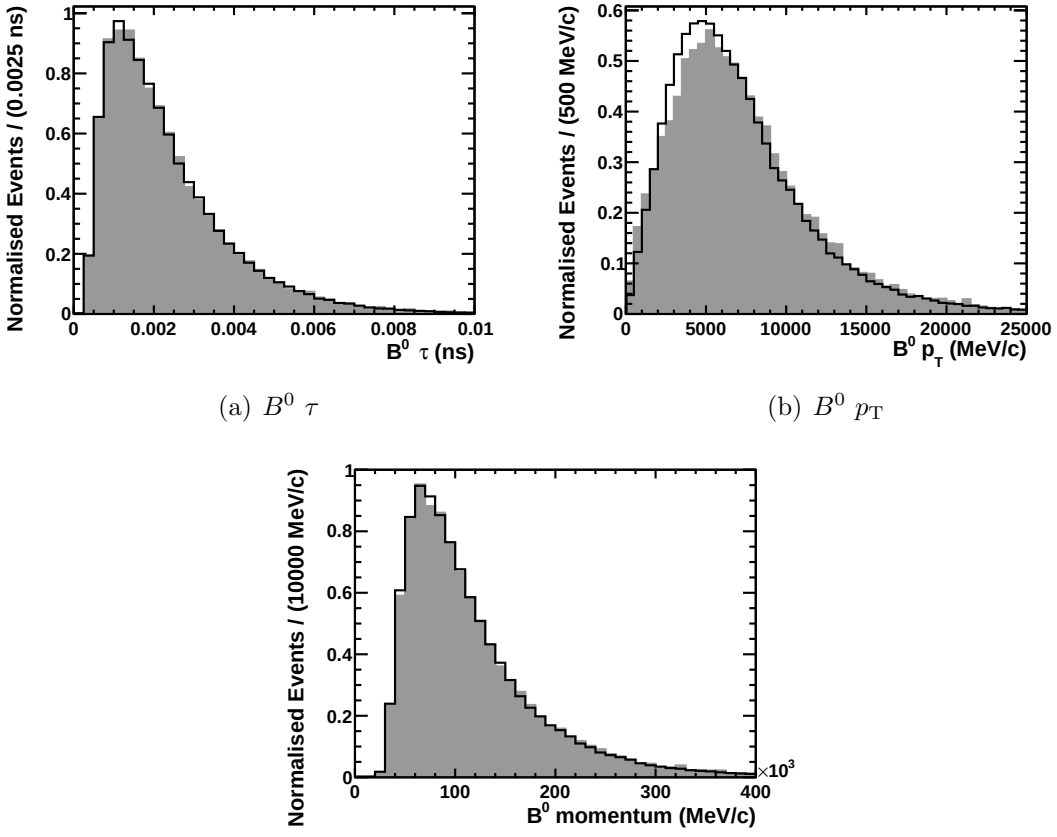
5.3.2 Validation of the efficiency factorisation

The acceptance correction procedure assumes that the angular efficiencies for a given candidate factorise as in Eq. (5.2.1). If this assumption is valid, applying the AC weights to the sample of simulated $B^0 \to K^{*0}\mu^+\mu^-$ events will result in a uniform angular distri-



(a) μ^+ $p_{\rm T}$ (b) μ^- $p_{\rm T}$

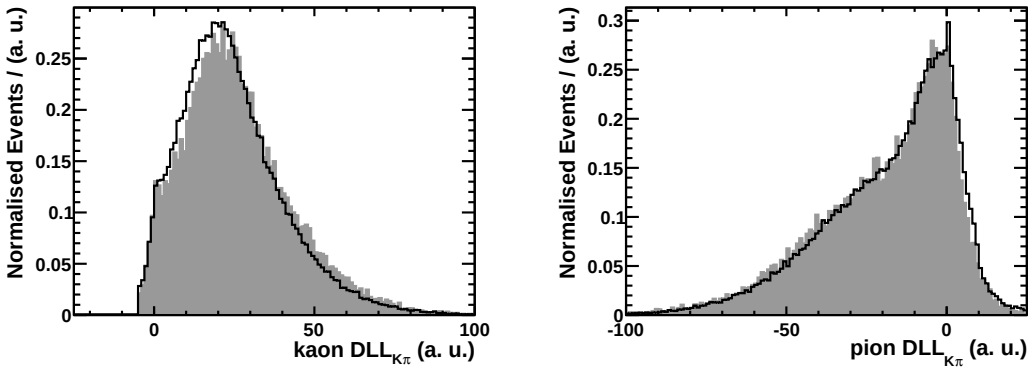
Figure 5.10: Comparison of the $p_{\rm T}$ of the muons in $B^0 \to J/\psi\, K^{*0}$ candidates in the data (black solid histogram) and in simulated $B^0 \to J/\psi\, K^{*0}$ events (grey shaded histogram).



(a) B^0 τ (b) B^0 $p_{\rm T}$

(c) B^0 momentum

Figure 5.11: Comparison of the B^0 variables used to produce the BDT, but not used in the pre-selection, in $B^0 \to J/\psi\, K^{*0}$ candidates in the data (black solid histogram) and in simulated $B^0 \to J/\psi\, K^{*0}$ events (grey shaded histogram), normalised to equal area.



(a) kaon $\mathrm{DLL}_{K\pi}$ (b) pion $\mathrm{DLL}_{K\pi}$

Figure 5.12: Comparison of the DLL variables used to produce the BDT, in $B^0 \to J/\psi\, K^{*0}$ candidates in the data (black solid histogram) and in simulated $B^0 \to J/\psi\, K^{*0}$ events (grey shaded histogram), normalised to equal area.

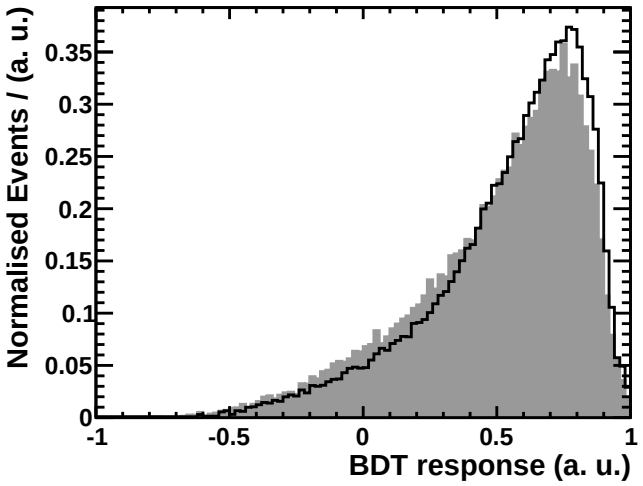
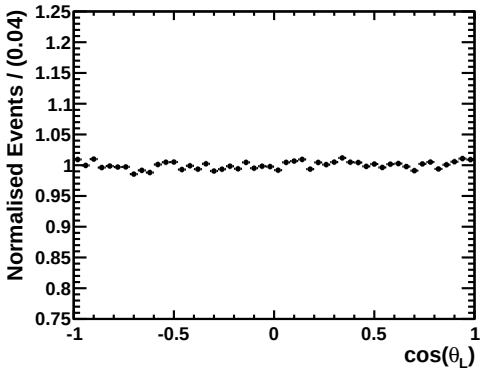
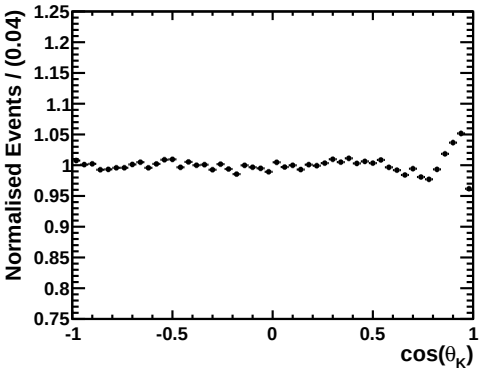


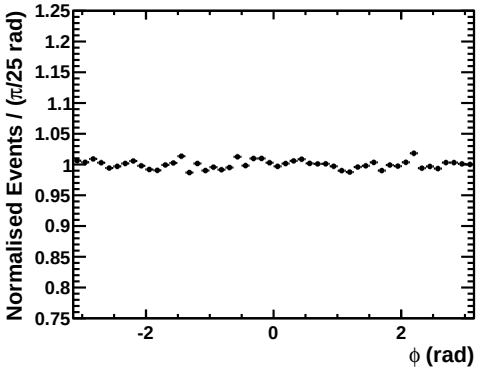
Figure 5.13: Comparison of the BDT response, in $B^0 \to J/\psi\, K^{*0}$ candidates in the data (black solid histogram) and in simulated $B^0 \to J/\psi\, K^{*0}$ events (grey shaded histogram), normalised to equal area.



(a) $\cos\theta_L$ distribution of events



(b) $\cos\theta_K$ distribution of events



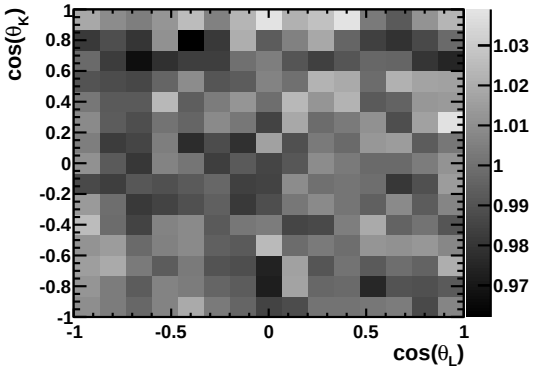
(c) ϕ distribution of events

Figure 5.14: Distribution of acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $0.10 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ in the decay angles $\cos\theta_L$ (a), $\cos\theta_K$ (b) and ϕ (c), normalised to obtain a mean of unity.

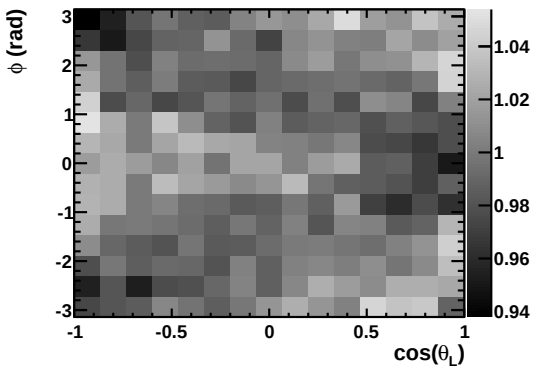
bution of the events.

The angular distribution of acceptance-corrected simulated events with $0.10 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ is shown in Fig. 5.14. The events are uniformly distributed in the $\cos\theta_L$ and ϕ angles. The events are also uniformly distributed between $-1 < \cos\theta_K < 0.8$. However, there is an excess of events at the level of 5% in the remainder of the $\cos\theta_K$ range. The excess in this region is caused by events obtaining weights that are too large at low q^2 and $\cos\theta_L \sim \pm 1$. This is caused by the combination of large acceptance effects in both the $\cos\theta_L$ and $\cos\theta_K$ angles simultaneously. The excess is indicative of either a non-factorisable efficiency, or is a feature of the fit due to the small number of simulated events in this region of the $\cos\theta_L$ and $\cos\theta_K$ angular space. The excess is considered as a source of systematic uncertainty in Sec. 6.3.2.

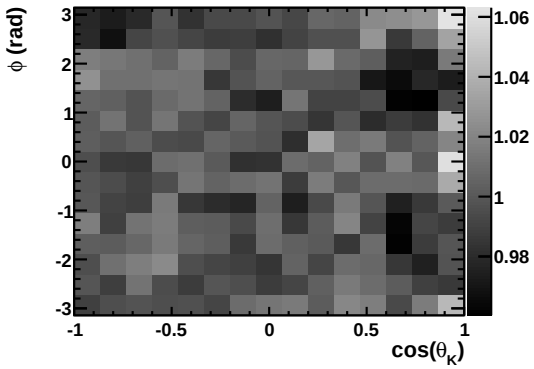
The distribution of acceptance-corrected simulated events in the ($\cos\theta_L$, $\cos\theta_K$), ($\cos\theta_L$, ϕ) and ($\cos\theta_K$, ϕ) 2D projections are shown in Fig. 5.15 for events with $0.10 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$. The events are uniformly distributed across the projections ($\cos\theta_L$,



(a) $(\cos\theta_L,\, \cos\theta_K)$ distribution of events



(b) $(\cos\theta_L,\, \phi)$ distribution of events



(c) $(\cos\theta_K,\, \phi)$ distribution of events

Figure 5.15: Projections of the acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $0.10 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$, for $(\cos\theta_L,\, \cos\theta_K)$ (a), $(\cos\theta_L,\, \phi)$ (b) and $(\cos\theta_K,\, \phi)$ (c).

$\cos\theta_K$) and $(\cos\theta_K,\, \phi)$. There is a non-uniformity in the $(\cos\theta_L,\, \phi)$ projection at the level of 5%, as shown in Fig. 5.15(b), which is indicative of a non-factorisable efficiency. This non-uniformity is present in events with $4 \lesssim q^2 \lesssim 13\,\mathrm{GeV}^2/c^4$ and is considered as a source of systematic uncertainty in Sec. 6.3.2.

The deviation from a uniform distribution of events indicates that the acceptance correction does not properly account for the acceptance effects present in the decay angles. The impact of these inadequacies on the overall validity of the acceptance correction is quantified by comparing the distribution of simulated events to the uniform distribution expected for a perfect acceptance correction in regions of the angular space. This study is presented in App. C, and concludes that the simulated events deviate from a uniform distribution at the level of 5%, across the three decays angles and q^2. This deviation is considered as a source of systematic uncertainty in Sec. 6.3.2.

5.3.3 Summary of the acceptance correction validation

The acceptance effects introduced by the detector and event reconstruction have been verified to be accurately reproduced in simulated events elsewhere. The acceptance effects introduced by the requirements imposed in the pre-selection and trigger are dependent only on the distribution of events in a number of variables. There are no significant discrepancies in the majority of the variables. Any residual discrepancies are considered to be sources of systematic uncertainty. Any discrepancies in the BDT input variables are also taken as sources of systematic uncertainty. The $\sim$4% discrepancy in the efficiency of the BDT response requirement introduces a negligible $\sim$0.1% discrepancy in the total acceptance effect.

The quality of the AC fit is examined by plotting the distribution of acceptance-corrected events in each decay angle. The distributions are uniform, except for an excess of events in $0.8 < \cos\theta_K < 1.0$. The quality of the AC fit is further examined by plotting the distribution of acceptance-corrected events in each 2D projection of the angular distribution. These distributions are also uniform, except for the $(\cos\theta_L,\, \phi)$ projection for events with $4 \lesssim q^2 \lesssim 13\,\mathrm{GeV}^2/c^4$. The discrepancy of the simulated events from a uniform distribution is quantified in App. C across the three decay angles and q^2. The acceptance correction is shown to be valid at the level of $\sim$5%. The non-uniformity is considered as a source of systematic uncertainty in Sec. 6.3.2.

Chapter 6

Measurement of the differential branching fraction

The differential branching fraction of the $B^0 \to K^{*0}\mu^+\mu^-$ decay ($d\mathcal{B}/dq^2$) is sensitive to contributions from NP particles. To reduce the experimental uncertainties on the determination of $d\mathcal{B}/dq^2$, the measurement is normalised to the control channel $B^0 \to J/\psi\, K^{*0}$ by forming a ratio of branching fractions. Such a normalisation causes the largest experimental uncertainties to cancel, allowing a more precise measurement of the $d\mathcal{B}/dq^2$ to be made. The branching fraction of the $B^0 \to J/\psi\, K^{*0}$ decay is measured to a precision of about $\sim$5% [76, 77].

The $d\mathcal{B}/dq^2$ is extracted using a fit to the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data, and is normalised using a fit to $B^0 \to J/\psi\, K^{*0}$ candidates in the data. The AC weights from Sec. 5 are used to account for the efficiency of each candidate in the fit. The confidence intervals returned from the fit do not account for fluctuations of the $d\mathcal{B}/dq^2$ due to the candidate weights, so corrected confidence intervals are computed.

The $d\mathcal{B}/dq^2$ is derived in section 6.1. The fitting procedure is discussed in section 6.2. The measurement of $d\mathcal{B}/dq^2$, including a discussion of the statistical and systematic uncertainties, is detailed in Sec. 6.3.

6.1 Derivation of the differential branching fraction

The number of $B^0 \to K^{*0}\mu^+\mu^-$ events ($N_{\rm sig}$) in a data set that corresponds to a given integrated luminosity ($\mathcal{L}_{\rm int}$) is expressed as

$$N_{\rm sig} = 2\,\mathcal{L}_{\rm int}\,\sigma^{b\bar{b}}\,f_d\,\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)\mathcal{B}(K^{*0} \to K^+\pi^-)\,\varepsilon_{\rm sig}, \tag{6.1.1}$$

where $\sigma^{b\bar{b}}$ is the cross section to produce a $b\bar{b}$ pair, f_d is the fraction of b quarks that hadronise to a B^0 meson, $\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$ is the branching fraction of the $B^0 \to K^{*0}\mu^+\mu^-$ decay, $\mathcal{B}(K^{*0} \to K^+\pi^-)$ is the branching fraction of the $K^{*0} \to K^+\pi^-$ decay and ε_{sig} is the efficiency to detect, reconstruct and select a $B^0 \to K^{*0}\mu^+\mu^-$ decay. The factor 2 arises as b quarks are produced in pairs, and the $d\mathcal{B}/dq^2$ is calculated using both B^0 and $\overline{B}^0$ mesons. Eq. (6.1.1) can be rearranged for the branching fraction

$$\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-) = \frac{N_{\text{sig}}}{2\mathcal{L}_{\text{int}}\,\sigma^{b\bar{b}}\,f_d\,\mathcal{B}(K^{*0} \to K^+\pi^-)\varepsilon_{\text{sig}}}. \tag{6.1.2}$$

If the branching fraction was calculated using the above expression, there would be large contributions to the uncertainty from $\mathcal{L}_{\text{int}}$, $\sigma^{b\bar{b}}$ and f_d. Each of these quantities has been measured: $\sigma^{b\bar{b}} = \sigma(pp \to b\bar{b}X) = 288 \pm 4 \pm 48\,\mu\text{b}$ [78], which incorporates a 3.5% uncertainty from the measurement of $\mathcal{L}_{\text{int}}$ [79]; and $f_d = (33.9 \pm 3.9)\%$ [80]. The uncertainty on $\sigma^{b\bar{b}}$ would introduce a $\sim$16% uncertainty on the branching fraction, and given the N_{sig} and ε_{sig} expected with the $1\,\text{fb}^{-1}$ data set, would be the dominant uncertainty.

Normalising the measurement to the $B^0 \to J/\psi\,K^{*0}$ control channel results in the cancellation of the uncertainties on $\mathcal{L}_{\text{int}}$, $\sigma^{b\bar{b}}$ and f_d, resulting in a more precise measurement of $d\mathcal{B}/dq^2$. The ratio of the $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to J/\psi\,K^{*0}$ branching fractions is

$$\frac{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to J/\psi\,K^{*0})} = \frac{N_{\text{sig}}}{N_{\text{norm}}}\frac{\varepsilon_{\text{norm}}}{\varepsilon_{\text{sig}}}\mathcal{B}(J/\psi \to \mu^+\mu^-), \tag{6.1.3}$$

where: N_{norm} is the number of $B^0 \to J/\psi\,K^{*0}$ decays in the data set; $\varepsilon_{\text{norm}}$ is the efficiency to detect, reconstruct and select a $B^0 \to J/\psi\,K^{*0}$ decay; $\mathcal{B}(B^0 \to J/\psi\,K^{*0})$ is the branching fraction of the $B^0 \to J/\psi\,K^{*0}$ decay; and $\mathcal{B}(J/\psi \to \mu^+\mu^-)$ is the branching fraction for the $J/\psi \to \mu^+\mu^-$ decay.

The $d\mathcal{B}/dq^2$ is extracted in the analysis q^2 bins to avoid incorporating the q^2 dependence of the branching fraction into the fit model, which would make the fit too complicated given the limited number of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data. In a q^2 bin i the $d\mathcal{B}/dq^2$ is defined as

$$\frac{d\mathcal{B}(i)}{dq^2} = \frac{1}{q^2_{\text{max};i} - q^2_{\text{min};i}}\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-; i), \tag{6.1.4}$$

where $q^2_{\text{max};i}$ and $q^2_{\text{min};i}$ are the maximum and minimum of the q^2 bin.

In Eq. (6.1.3) the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction is expressed in terms of the efficiency of the $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to J/\psi\,K^{*0}$ decays, which simulated events indicate are $\sim$0.7% and $\sim$1.0% respectively. As the AC weights give a model-independent estimate of the efficiency, they can be used to account for the efficiencies directly in the fit.

However, for the fit to return confidence intervals that correctly account for fluctuations in the number of events in the data the mean candidate weight must equal unity. In Sec. 6.3.1 the confidence intervals are adjusted so that they also account for fluctuations in the candidate weights. The factor required to normalise the AC weights from the 'efficiency scale' of $\sim$0.7% to a mean of unity is the α factor introduced in Sec. 5. As such, the fit is performed with the normalised AC weights applied to the candidates, while the factor α is incorporated into the calculation of the $d\mathcal{B}/dq^2$ to account for the absolute scale of the efficiency. As the weight of any given candidate is different from unity, even though the mean is unity, a correction of the confidence intervals is necessary. The $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction is therefore calculated according to

$$\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-; i) = \mathcal{B}(B^0 \to J/\psi\, K^{*0})\, \mathcal{B}(J/\psi \to \mu^+\mu^-)\, \frac{N'_{\rm sig}(i)}{N'_{\rm norm}}\, \frac{\alpha_{\rm norm}}{\alpha_{\rm sig}(i)}, \qquad (6.1.5)$$

where $\alpha_{\rm sig}(i)$ is the $B^0 \to K^{*0}\mu^+\mu^-$ efficiency in q^2 bin i and $\alpha_{\rm norm}$ is the $B^0 \to J/\psi\, K^{*0}$ efficiency in the q^2 range $9.21 < q^2 < 9.96\,{\rm GeV}^2/c^4$. The quantities obtained from the fit are $N'_{\rm sig}(i)$, the yield of acceptance-corrected $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the q^2 bin i, and $N'_{\rm norm}$, the yield of acceptance-corrected $B^0 \to J/\psi\, K^{*0}$ candidates with $9.21 < q^2 < 9.96\,{\rm GeV}^2/c^4$.

6.2 Fitting for the differential branching fraction

A fit is performed to the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of acceptance-corrected $B^0 \to J/\psi\, K^{*0}$ candidates in the data to obtain $N'_{\rm norm}$ and determine the value of the parameters of the mass model. A suitably modified version of the mass model is then used to fit the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates to obtain $N'_{\rm sig}$. These two quantities are used to calculate $d\mathcal{B}/dq^2$ in each analysis q^2 bin.

6.2.1 Definition of the mass model

The $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of $B^0 \to J/\psi\, K^{*0}$ candidates in the data is formed from contributions of $B^0 \to J/\psi\, K^{*0}$ decays, $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ decays and Class III background events. The $B^0 \to J/\psi\, K^{*0}$ contribution is parameterised by a Crystal Ball (CB) function [81]. This function incorporates a Gaussian pdf to account for the peaking structure of the decay plus an power-law tail that accounts for the contribution of radiative $B^0 \to J/\psi\, K^{*0}$

decays. The CB is defined as

$$M(m|\bar{m},\sigma,n,a) = \begin{cases} \exp\left(-\dfrac{(m-\bar{m})^2}{2\sigma^2}\right) & \text{, for } \dfrac{m-\bar{m}}{\sigma} > -a \\ \left(\dfrac{n}{|a|}\right)^n \exp\left(-\dfrac{|a|^2}{2}\right) \cdot \left(\dfrac{n}{|a|} - |a| - \dfrac{m-\bar{m}}{\sigma}\right)^{-n} & \text{, for } \dfrac{m-\bar{m}}{\sigma} \leq -a \end{cases} \tag{6.2.1}$$

where m is the $\mathrm{m}_{K\pi\mu\mu}$ value of a candidate, $\bar{m}$ is the mean of the Gaussian component, σ is the mass resolution (the width of the Gaussian component), n is the size of the power-law component and a determines the transition point between the Gaussian and the power-law component. The value of a is required to be positive to ensure that the power-law component exists only for $m < \bar{m}$. A second CB is incorporated with a mass resolution (denoted σ_{w}) that is required to be larger than that of the first CB (denoted σ_{n}) to account for a second population of signal candidates that are poorly reconstructed. These two components share the same $\bar{m}$, n and a parameters and form the 'double-CB' function $M_{\mathrm{B^0}}$.

Contributions from $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ are incorporated into the fit by a further double-CB function $M_{\mathrm{B^0_s}}$. This shares the same values of σ_{n}, σ_{w}, n and a as $M_{\mathrm{B^0}}$. Furthermore, $\bar{m}$ is constrained to be

$$\bar{m}_{B^0_s} = \bar{m}_{B^0} + 87\,\mathrm{MeV}/c^2$$

to account for the difference in the B^0_s and B^0 masses [1, 82]. Requiring the same values of σ_{n} and σ_{w} for both $M_{\mathrm{B^0}}$ and $M_{\mathrm{B^0_s}}$ is considered a negligible source of systematic uncertainty as no significant variation in the mass resolution is expected between $B^0 \to J/\psi\, K^{*0}$ and $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ decays. The contribution of $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ is calculated in Sec. 4.3.2 to be $(0.88 \pm 0.17)\%$ of the $B^0 \to J/\psi\, K^{*0}$ yield. A Gaussian pdf is used to constrain the $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ yield within the measurement uncertainty. The calculation assumes that the ratio of efficiencies of the two channels is unity. It is measured to be 0.97 ± 0.01 [67]. The difference is small compared to the uncertainty on the $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ yield, and so this assumption is considered a negligible source of systematic uncertainty.

The Class III background is modelled by a falling exponential $E_{bkg}(m|p_0)$ that is parameterised by an exponent p_0. This is an empirical model that accurately describes the shape of the Class III background events in data.

6.2.1.1 The $B^0 \to K^{*0}\mu^+\mu^-$ mass model

The $B^0 \to K^{*0}\mu^+\mu^-$ fit model is an appropriately modified version of the fit model described above. Firstly, the ratio of branching fractions of the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ decays is not measured. The ratio of the relevant CKM factors of the two

decays, and the value of f_s/f_d, gives an estimate of the ratio of yields as

$$\frac{f_s}{f_d}\frac{|V_{td}|^2}{|V_{ts}|^2} \approx \frac{f_s}{f_d}\frac{|V_{cd}|^2}{|V_{cs}|^2} \approx 1\%. \tag{6.2.2}$$

The $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ yield is therefore constrained by a Gaussian pdf centred at 1% with a width of 1% of the $B^0 \to K^{*0}\mu^+\mu^-$ yield. The large uncertainty is assumed to account for potential NP contributions to the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ decay.

The contribution of Class I and Class II background events is accounted for in the $B^0 \to K^{*0}\mu^+\mu^-$ fit. The $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of simulated Class I and Class II background events with $0.10 < q^2 < 19.00\,\mathrm{GeV}/c^2$ is shown in Fig. 4.5. The equivalent histogram in each analysis q^2 bin is used to model the Class I and Class II background contribution. The combined yield of these backgrounds is 10.6% of the $B^0 \to K^{*0}\mu^+\mu^-$ yield (see Tab. 4.3). This fit assumes that the relative amounts of Class III and Class II background events and $B^0 \to K^{*0}\mu^+\mu^-$ decays is the same in each analysis q^2 bin. The branching fraction of $\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$ is assumed to be $\mathcal{B}(\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-) \sim 5 \times 10^{-7}$, given the approximate branching fraction scale of other $b \to s$ FCNC decays of $\sim 10^{-6}$ and the measured $\mathcal{B}(\Lambda^{*0} \to p^+K^-) = (45 \pm 1)\%$. The assumed $\mathcal{B}(\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-)$, and the variation of $\mathcal{B}(\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-)$ in q^2, are considered as sources of systematic uncertainty in Sec. 6.3.2.

The log-likelihood for the fit to $B^0 \to K^{*0}\mu^+\mu^-$ candidates is

$$\begin{aligned}
-\log\mathcal{L} = & -\sum_{i=0}^{N} w_{\mathrm{i}} \log \\
& \left[\frac{N'_{B^0}}{(1 + f_\phi + f_{B^0_s})N'_{B^0} + N'_{bkg}} M_{B^0}(m|\bar{m}, \sigma_{\mathrm{n}}, \sigma_{\mathrm{w}}, a, n) \right. \\
& + \frac{f_{B^0_s} N'_{B^0}}{(1 + f_\phi + f_{B^0_s})N'_{B^0} + N'_{bkg}} M_{B^0_s}(m|\bar{m}, \sigma_{\mathrm{n}}, \sigma_{\mathrm{w}}, a, n) \\
& + \frac{f_\phi N'_{B^0}}{(1 + f_\phi + f_{B^0_s})N'_{B^0} + N'_{bkg}} M_\phi(m) \\
& \left. + \frac{N'_{bkg}}{(1 + f_\phi + f_{B^0_s})N'_{B^0} + N'_{bkg}} E_{bkg}(m|p_0) \right] \\
& - \log\mathcal{P}(N|(1 + f_\phi + f_{B^0_s})N'_{B^0} + N'_{bkg}) \,,
\end{aligned} \tag{6.2.3}$$

where: m is the $\mathrm{m}_{K\pi\mu\mu}$ value of a candidate; w_{i} is the weight of candidate i; N'_{B^0} is the $B^0 \to K^{*0}\mu^+\mu^-$ yield; $f_{B^0_s}$ is the ratio of the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ yields;

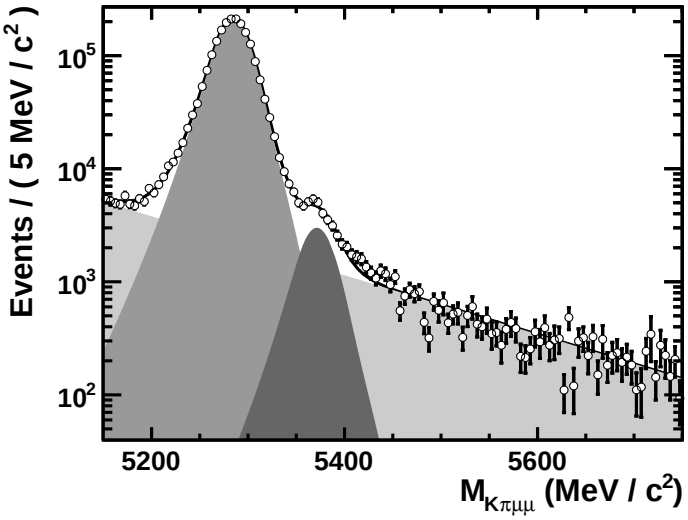


Figure 6.1: The $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of acceptance-corrected $B^0 \to J/\psi\, K^{*0}$ candidates in the data (black open points) in the mass range $5150 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. A fit (black solid line) to the data is overlaid, with the combinatoric (light grey area), $B^0 \to J/\psi\, K^{*0}$ (grey area) and $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ (dark grey area) components indicated.

M_ϕ is a histogram that parameterises the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of Class I and Class II background events; f_ϕ is the ratio of the Class I and Class II background yield to N'_{B^0}; and N'_{bkg} is the Class III background yield. The final term is a Poisson pdf that accounts for statistical fluctuations in the number of events.

6.3 Extracting the differential branching fraction

The $d\mathcal{B}/dq^2$ is measured using an unbinned extended maximum-likelihood fit to the $B^0 \to J/\psi\, K^{*0}$ candidates in the data to extract N'_{norm}, and to the $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data in each analysis q^2 bin to extract $N'_{\mathrm{sig}}(i)$. A correction of the confidence intervals returned from the fit is required due to the candidate weights differing from unity. The correction of the confidence intervals is described in Sec. 6.3.1. The systematic uncertainties of the $d\mathcal{B}/dq^2$ measurement are discussed in Sec. 6.3.2.

The fit to the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of acceptance-corrected $B^0 \to J/\psi\, K^{*0}$ candidates in the data is shown on a logarithmic scale in Fig. 6.1. The same plot is shown on a linear scale in App. D. This fit yields $N'_{\mathrm{norm}} = 103850 \pm 340$. The mass resolution is $\sigma_{\mathrm{n}} \sim 16\,\mathrm{MeV}$, which is consistent with that measured in simulated events.

The value of $N'_{\mathrm{sig}}(i)$ is obtained by fitting the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of acceptance-corrected $B^0 \to K^{*0}\mu^+\mu^-$ candidates in each analysis q^2 bin. The fit parameters for $M_{\mathrm{B^0}}$ and $M_{\mathrm{B^0_s}}$ are fixed to the values obtained in the fit to the $B^0 \to J/\psi\, K^{*0}$ candidates. In

q^2 range (GeV^2/c^4)	Signal yield	Background yield	$d\mathcal{B}/dq^2$ ($10^{-7}\,\mathrm{GeV}^{-2}c^4$)
$0.10 < q^2 < 2.00$	125.85 ± 12.25	22.16 ± 3.43	0.62 ± 0.06
$2.00 < q^2 < 4.30$	64.42 ± 10.49	40.44 ± 4.34	0.30 ± 0.05
$4.30 < q^2 < 8.68$	262.28 ± 18.80	70.55 ± 6.23	0.55 ± 0.04
$10.09 < q^2 < 12.86$	160.21 ± 14.61	47.85 ± 4.92	0.47 ± 0.04
$14.18 < q^2 < 16.00$	131.41 ± 12.42	17.86 ± 3.31	0.55 ± 0.05
$16.00 < q^2 < 19.00$	142.09 ± 13.27	22.70 ± 3.79	0.37 ± 0.03
$1.00 < q^2 < 6.00$	180.17 ± 16.51	83.87 ± 6.34	0.36 ± 0.03

Table 6.1: Results of the differential branching fraction measurement. The acceptance-corrected signal and Class III background yields in the mass range $5230 < \mathrm{m}_{K\pi\mu\mu} < 5330\,\mathrm{MeV}/c^2$ and $d\mathcal{B}/dq^2$ are shown in each analysis q^2 bin, where the uncertainties are statistical only.

Sec. 6.3.2 the mass resolution is considered as a source of systematic uncertainty, and is found to be small compared to both the statistical uncertainty and the other systematic uncertainties. Only the yield of each component and the exponent p_0 is allowed to vary between the $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to K^{*0}\mu^+\mu^-$ fits. The branching fraction in each q^2 bin is then calculated using Eq. (6.1.4) and Eq. (6.1.5) with

$$\mathcal{B}(B^0 \to J/\psi\, K^{*0}) = (1.31 \pm 0.06) \times 10^{-3}\ [76] \tag{6.3.1}$$

and

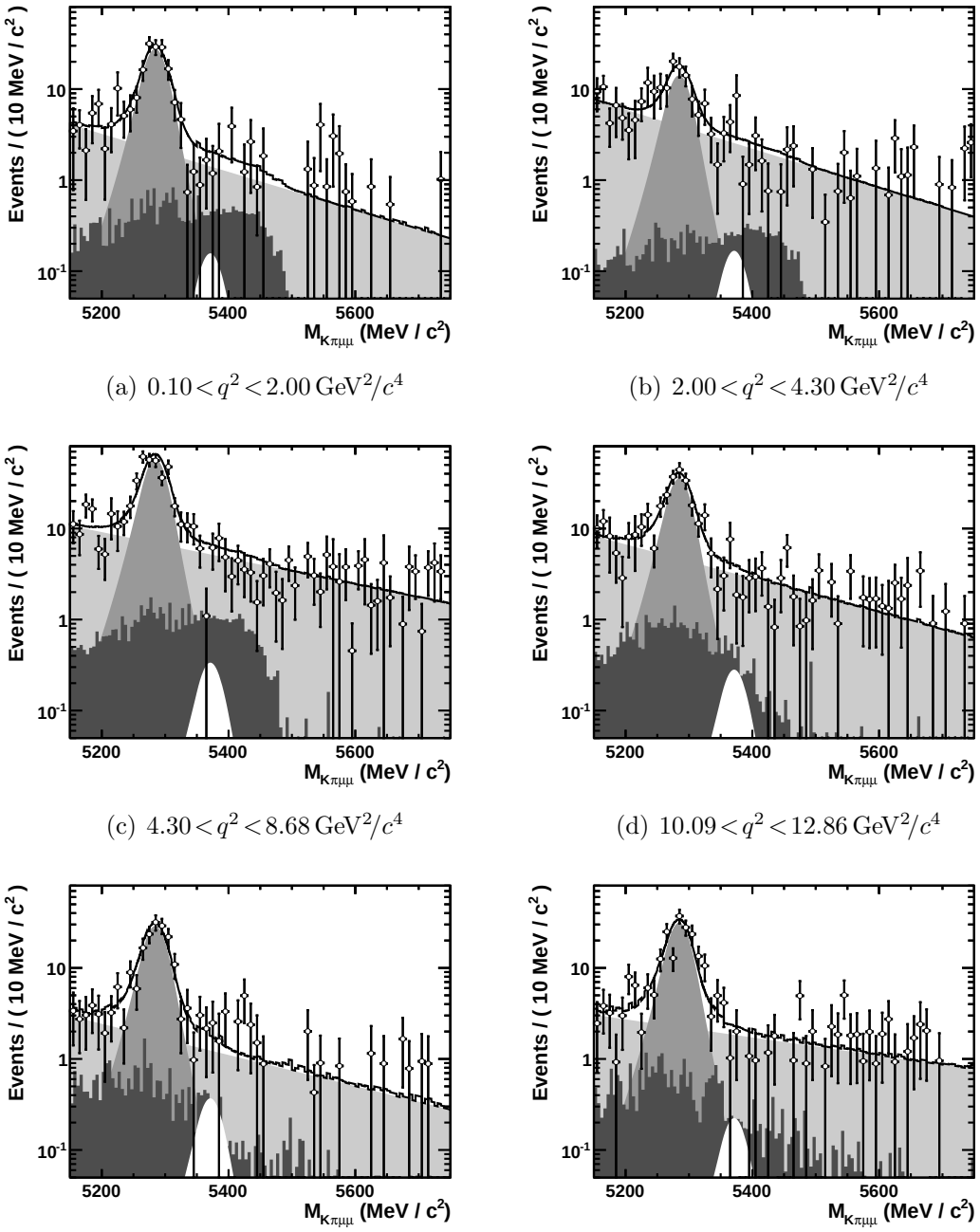
$$\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.93 \pm 0.06) \times 10^{-2}\ [1]. \tag{6.3.2}$$

The fits are shown on a logarithmic scale in Figs. 6.2 and 6.3 and the results are summarised in Tab. 6.1. The same plots are shown on a linear scale in App. D.

6.3.1 Statistical uncertainties

The confidence intervals returned from the fit, detailed in Tab. 6.1, do not account for fluctuations in the data due to the candidate weights and will therefore under-cover the true $d\mathcal{B}/dq^2$ value. The confidence intervals from the fit are adjusted so that the correct confidence intervals are obtained.

The correction factor is calculated in each analysis q^2 bin by generating an ensemble of simulated data sets using the fit model. The number of candidates in each simulated data set is drawn from a Poisson distribution with a mean equal to the number of candidates in the data. The values of the parameters of the fit model are obtained from the fit to $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data.



(a) $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$

(b) $2.00 < q^2 < 4.30\,\mathrm{GeV}^2/c^4$

(c) $4.30 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$

(d) $10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$

(e) $14.18 < q^2 < 16.00\,\mathrm{GeV}^2/c^4$

(f) $16.00 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$

Figure 6.2: The $\mathrm{m}_{K\pi\mu\mu}$ distribution of acceptance-corrected $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data (open points) in the mass range $5150 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$ in each of the analysis q^2 bins. A fit (black solid line) to the data is overlaid, with the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ (white area); Class III background (light grey area); $B^0 \to K^{*0}\mu^+\mu^-$ (grey area); and Class I and Class II background (dark grey histogram) components indicated. The fit model is defined in Sec. 6.3.

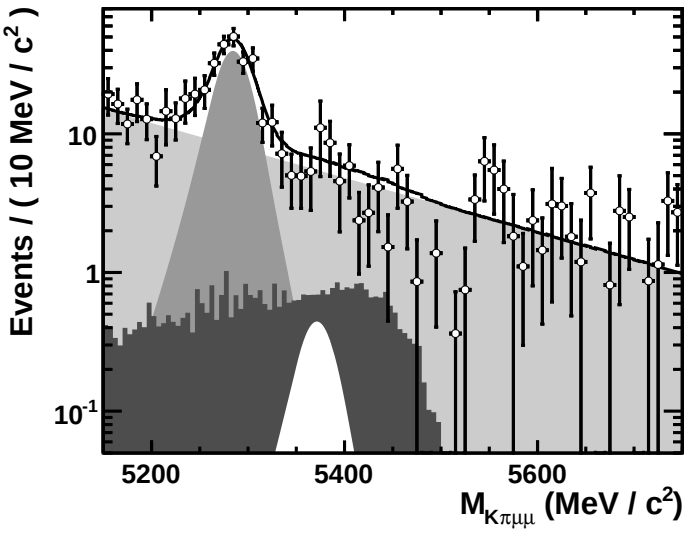


Figure 6.3: The $m_{K\pi\mu\mu}$ distribution of acceptance-corrected $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$ (black open points) in the mass range $5150 < m_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. A fit (black line) to the data is overlaid, with the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ (white area); Class III background (light grey area); $B^0 \to K^{*0}\mu^+\mu^-$ (grey area); and Class I and Class II background (dark grey histogram) components indicated. The fit model is defined in Sec. 6.3.

To mimic the acceptance effects in the simulated events, the events are accept-rejected using an AC pdf. An angular distribution for the simulated candidates is generated according to the $B^0 \to K^{*0}\mu^+\mu^-$ angular distribution in the SM using Eq. (2.5.38). A cross-check is performed using a uniform distribution of events across the decay angles with no significant change in the result. The AC pdf is chosen according to the average q^2 value of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data in the relevant analysis q^2 bin. An AC weight is applied to each event to mimic the acceptance correction.

Each simulated data set is fitted to obtain a distribution of $N'_{\rm sig}(i)$. The pull of each $N'_{\rm sig}(i)$ is defined as

$$\text{pull} = \frac{1}{\sigma_{N'_{\rm sig}(i)}} \left(N'_{\rm sig}(i) - \overline{N'_{\rm sig}}(i) \right) \tag{6.3.3}$$

where $\sigma_{N'_{\rm sig}(i)}$ is the confidence interval on $N'_{\rm sig}(i)$ obtained from the fit and $\overline{N'_{\rm sig}}(i)$ is the number of acceptance-corrected $B^0 \to K^{*0}\mu^+\mu^-$ events in the simulated data set. In the absence of candidate weights, the pulls are distributed as a Gaussian with a width of unity. As $\sigma_{N'_{\rm sig}(i)}$ does not account for fluctuations in the candidate weights, the width of the distribution of pulls is larger than unity.

The width of the distribution of pulls is obtained from a Gaussian fit. The deviation of the width of the pull distribution from unity is used to correct the confidence intervals obtained from the fit to the data. If the width of the fit is consistent with unity within

q^2 range (GeV^2/c^4)	Fit Uncertainty	Pull Width	Corrected Uncertainty
$0.10 < q^2 < 2.00$	0.06	1.11 ± 0.01	0.07
$2.00 < q^2 < 4.30$	0.05	1.10 ± 0.01	0.05
$4.30 < q^2 < 8.68$	0.04	1.07 ± 0.01	0.04
$10.09 < q^2 < 12.86$	0.04	1.05 ± 0.01	0.04
$14.18 < q^2 < 16.00$	0.05	1.01 ± 0.01	0.05
$16.00 < q^2 < 19.00$	0.03	1.00 ± 0.01	0.03
$1.00 < q^2 < 6.00$	0.03	1.08 ± 0.01	0.03

Table 6.2: Correction of the statistical component of the confidence interval for the differential branching fraction measurement. The confidence intervals obtained from the fit, the width and uncertainty of the pull distribution, and the corrected confidence intervals are shown.

the uncertainty of the Gaussian fit no correction is applied. The confidence interval from the fit, the width of the distribution of pulls for each analysis q^2 bin, and the corrected confidence interval are summarised in Tab. 6.2.

6.3.2 Sources of systematic uncertainty

The systematic uncertainties associated with the measurement of $d\mathcal{B}/dq^2$ are detailed below. The systematic uncertainties related to the AC weights are detailed in Secs. 6.3.2.1 and 6.3.2.2; those related to the trigger efficiency are detailed in Sec. 6.3.2.3; those related to the corrections that are applied to the simulated events when producing the AC pdfs are detailed in Secs. 6.3.2.4 to 6.3.2.8; and those related to the fit model are detailed in Secs. 6.3.2.9 to 6.3.2.12. For each systematic variation the fit is repeated and the $d\mathcal{B}/dq^2$ is calculated. The absolute difference in the $d\mathcal{B}/dq^2$ between the nominal and systematically varied fit is taken as the uncertainty associated with that variation.

6.3.2.1 The AC weights

The size of the AC weights are considered as a source of systematic uncertainty due to the non-uniform distribution of events in Fig. 5.14(b) and the 5% shift in the total pull distribution obtained in App. C. These uncertainties arise due to the limited number of simulated events used to calculate the AC weights. To conservatively explore the impact

of uncertainties in the AC weights, the AC weights are varied according to:

$$w_{\rm i} \rightarrow w_{\rm i} \times (1 \pm 0.1 \times |\cos\theta_{L;{\rm i}}|); \tag{6.3.4}$$

$$w_{\rm i} \rightarrow w_{\rm i} \times (1 \pm 0.1 \times |\cos\theta_{K;{\rm i}}|); \tag{6.3.5}$$

$$w_{\rm i} \rightarrow w_{\rm i} \times (1 \pm 0.1 \times |\cos\theta_{L;{\rm i}}|)(1 \pm 0.1 \times |\cos\theta_{K;{\rm i}}|); \tag{6.3.6}$$

$$w_{\rm i} \rightarrow w_{\rm i} \times (1 \pm 0.1 \times |\cos\theta_{L;{\rm i}}|)(1 \mp 0.1 \times |\cos\theta_{K;{\rm i}}|); \tag{6.3.7}$$

where $w_{\rm i}$ is the weight of candidate i. The 0.1 factor gives a 5% fluctuation when averaged across the decay angle.

The non-uniform distribution of events in Fig. 5.15(b) indicates the extent of non-factorisable acceptance effects. To conservatively explore the impact of such effects the AC weights are varied according to

$$w_{\rm i} \rightarrow w_{\rm i} \times (1 \pm 0.1 \times \sin(\pi \cos\theta_{L;{\rm i}}) \sin(\pi \cos\theta_{K;{\rm i}})). \tag{6.3.8}$$

6.3.2.2 Acceptance correction binning

The acceptance effects are averaged over each $q^2_{\rm AC}$ bin when fitting for the AC pdfs. To explore the impact of the choice of binning, an AC pdf that differs from the nominal AC pdf by ± 1 $q^2_{\rm AC}$ bin is used to compute the AC weights.

6.3.2.3 Trigger efficiency

The trigger efficiency is considered a source of systematic uncertainty due to the $\sim$3% difference in the efficiency of the L0Muon algorithm measured in data and in simulated events (see Fig. 3.8). To explore the impact of this uncertainty, a set of AC pdfs are generated where the efficiency of the L0Muon trigger algorithm is varied by $\pm$3% for muons with $p < 10\,{\rm GeV}/c$.

6.3.2.4 DLL binning scheme

The binning scheme used when generating DLL values for the simulated events is considered as a source of systematic uncertainty due to the residual differences in the DLL distributions of data and simulated events (see Fig. 5.12). To conservatively explore the impact of this uncertainty, a set of AC pdfs are generated where the simulated events are assigned a DLL value from the template that corresponds to a higher (lower) bin in p.

6.3.2.5 IsMuon efficiency

The IsMuon efficiency is corrected to account for the different IsMuon efficiency measured in data and in simulated events (see Fig. 3.7(b)). To explore the impact of this uncertainty, a set of AC pdfs are generated where the efficiency of the IsMuon criteria of each track is fluctuated by $\pm\sigma_{\text{IsMuon}}$ if the track satisfies $p<10\,\text{GeV}/c$ and $\mp\sigma_{\text{IsMuon}}$ otherwise, where σ_{IsMuon} is the uncertainty on the measured IsMuon efficiency in the relevant bin of p and η.

6.3.2.6 Tracking efficiency

The tracking efficiency is corrected to account for the different trigger efficiency measured in data and in simulated events (see Fig. 3.4). To explore the impact of this uncertainty, a set of AC pdfs are generated where the tracking efficiency of each track is fluctuated by $\pm\sigma_{\text{Tracking}}$ if the track satisfies $p<10\,\text{GeV}/c$ and $\mp\sigma_{\text{Tracking}}$ otherwise, where σ_{Tracking} is the uncertainty on the measured tracking efficiency in the relevant bin of p and η.

6.3.2.7 B^0 momentum spectrum

The reweighting of the momentum spectrum of the B^0 corrects the symptom, rather than the cause, of the different B^0 momentum spectrum in data and in simulated events and is therefore considered as a source of systematic uncertainty. To conservatively explore the impact of this uncertainty, a set of AC pdfs are generated without reweighting the B^0 momentum spectrum.

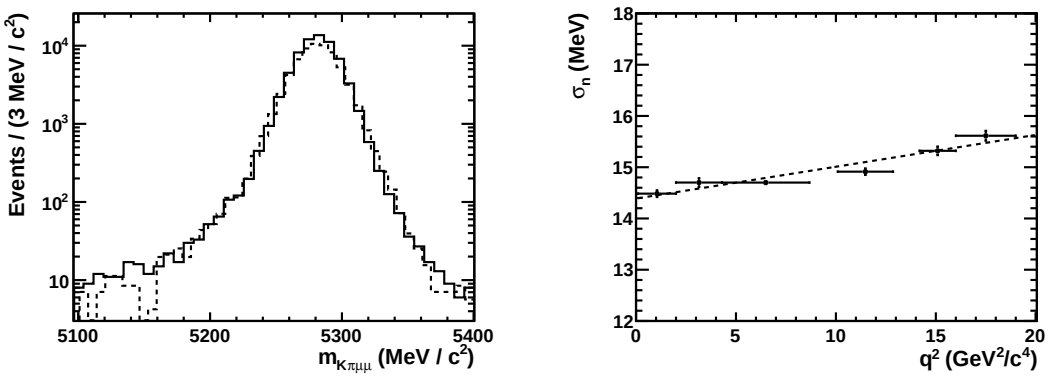
A set of AC pdfs is also generated where the B^0 p_{T} spectrum is reweighted instead of the momentum spectrum, in order to account for the difference in the B^0 p_{T} spectrum at low p_{T} (see Fig. 5.11(b)).

6.3.2.8 IP correction

The IP correction procedure accounts for the symptom, rather than the cause, of the different distribution of the IP as a function of $1/p_{\text{T}}$ in data and in simulated in events (see Fig. 3.5(c)). This is considered as a source of systematic uncertainty, which is explored by generating a set of AC pdfs without correcting the IP of the simulated events.

6.3.2.9 Mass resolution

The mass resolution σ_{n} is considered as a source of systematic uncertainty as it is fixed in the fits to $B^0\to K^{*0}\mu^+\mu^-$ candidates, despite it varying in q^2 (see Fig. 6.4(a)). The variation in q^2 is measured by fitting the $\text{m}_{K\pi\mu\mu}$ mass distribution of simulated $B^0\to K^{*0}\mu^+\mu^-$



(a) $\mathrm{m}_{K\pi\mu\mu}$ (b) $\sigma_{\mathrm{n}}(q^2)$

Figure 6.4: The $\mathrm{m}_{K\pi\mu\mu}$ distribution of simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$ (solid black histogram) and with $16.00 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ (dashed black histogram) in the mass range $5100 < \mathrm{m}_{K\pi\mu\mu} < 5400\,\mathrm{MeV}/c^2$ (a). The value of σ_{n} as a function of q^2, extracted from a fit to the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of simulated $B^0 \to K^{*0}\mu^+\mu^-$ events in each analysis q^2 bin (b). A fit to these data is indicated (black dashed line).

events in each analysis q^2 bin. The straight line fit shown in Fig. 6.4(b) indicates a 5% variation in σ_{n} as a function of q^2. To explore the impact of this variation, the fit is performed having fluctuated σ_{n} of $M_{\mathrm{B^0}}$ by $\pm 5\%$.

6.3.2.10 Class I and Class II backgrounds

The fit model incorporates the contribution of Class I and Class II backgrounds to the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution, however the (differential) branching fraction of each contribution has not been measured. The size of each contribution is therefore considered as a source of systematic uncertainty. The analysis of Ref. [83] measures the branching fraction of $B^0_s \to \phi\mu^+\mu^-$ with an uncertainty of 40%, so the contribution from $B^0_s \to \phi\mu^+\mu^-$ is fluctuated within the measurement uncertainty. The assumed branching fraction for $\Lambda^0_b \to \Lambda^{*0}(1520)\mu^+\mu^-$ is conservatively fluctuated to 1×10^{-7} and 1×10^{-6} to account for the unknown branching fraction of this decay and the unknown variation in q^2.

6.3.2.11 Multiple candidates

Due to the small signal branching fraction and the efficient rejection of background events, only one candidate is expected in each event. Multiple candidates in a single event are taken into account by adjusting the weight of each candidate according to Eq. (4.4.8). The equal treatment of each candidate as a result of this weighting is considered as a source of systematic uncertainty. To conservatively explore the impact of this uncertainty, the fit is performed having removed all events that contain more than one candidate.

6.3.2.12 Contributions from an S-wave component

The presence of an S-wave contribution to the $B^0 \to K^{*0}\mu^+\mu^-$ decay is neglected in the fit and is therefore considered as a source of systematic uncertainty. An S-wave contribution is present in the $B^0 \to J/\psi K^{*0}$ decay at the level of about 7% [41, 84]. To explore the impact of potential S-wave contributions to the $B^0 \to K^{*0}\mu^+\mu^-$ decay, the $d\mathcal{B}/dq^2$ is recalculated having reduced the signal yield by 7%. This uncertainty is included in Tab. 6.3, however, as it can only reduce $d\mathcal{B}/dq^2$, it is not included in the aggregated systematics shown in Tab. 6.4. It is included as a separate systematic in Tab. 6.5.

6.3.2.13 Summary of the systematic uncertainties

The systematic uncertainties are summarised in Tab. 6.3. The largest systematic uncertainty is on the AC weights, all the other systematic uncertainties are relatively small. The systematic uncertainty of each source (Secs. 6.3.2.1 to 6.3.2.12) is calculated by selecting the largest uncertainty associated with that source. This is performed in each analysis q^2 bin separately. These aggregated uncertainties are summed in quadrature to obtain the total systematic uncertainty. The aggregated and total uncertainties are summarised in Tab. 6.4. In each analysis q^2 bin the systematic uncertainty is approximately the same size as the statistical uncertainty.

6.3.3 Result of the differential branching fraction measurement

The result of the $d\mathcal{B}/dq^2$ measurement is shown in Tab. 6.5. The measurements are compared to the SM prediction, binned in the analysis q^2 bins, in Fig. 6.5. No theoretical prediction is shown between $8.68 < q^2 < 10.09\,\text{GeV}^2/c^4$ since the $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ decays dominate the decay rate in this region of q^2. Similarly, no measurement is performed between $8.68 < q^2 < 10.09\,\text{GeV}^2/c^4$ and $12.86 < q^2 < 14.18\,\text{GeV}^2/c^4$ as these regions of q^2 are removed to reject contributions of the $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ decays.

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	0.622	0.295	0.549	0.465	0.546	0.374	0.363
AC $\cos\theta_L$ Up	-0.037	-0.011	-0.023	-0.012	-0.006	0.004	-0.017
AC $\cos\theta_L$ Down	0.004	-0.000	0.006	0.016	0.030	0.031	0.003
AC $\cos\theta_K$ Up	-0.035	-0.019	-0.033	-0.017	-0.007	0.003	-0.022
AC $\cos\theta_K$ Down	0.001	0.010	0.018	0.022	0.031	0.032	0.009
AC $\cos\theta_L$ Up $\cos\theta_K$ Down	-0.053	-0.024	-0.046	-0.030	-0.024	-0.009	-0.031
AC $\cos\theta_L$ Down $\cos\theta_K$ Up	0.023	0.016	0.034	0.038	0.051	0.047	0.019
AC Non-factorisable Up	-0.015	-0.003	-0.007	0.004	0.012	0.015	-0.007
AC Non-factorisable Down	-0.017	-0.007	-0.009	0.001	0.014	0.020	-0.007
q^2 binning	-0.001	-0.000	0.001	0.000	0.000	-0.001	-0.001
q^2 binning	-0.001	-0.001	0.000	-0.000	0.001	-0.001	-0.001
Trigger efficiency Up	-0.003	-0.000	-0.000	0.001	0.002	0.001	-0.001
Trigger efficiency Down	0.004	0.001	0.001	-0.000	-0.001	-0.001	0.001
DLL binning +1	-0.008	-0.004	-0.005	-0.001	0.004	0.007	-0.005
DLL binning -1	-0.019	-0.007	-0.011	0.000	0.010	0.016	-0.010
IsMuon efficiency Up	0.005	0.003	0.004	-0.001	-0.005	-0.005	0.005
IsMuon efficiency Down	-0.006	-0.001	-0.003	0.002	0.005	0.006	-0.004
Tracking efficiency Up	-0.004	-0.001	-0.001	0.000	0.002	0.004	-0.001
Tracking efficiency Down	0.005	0.002	0.002	0.001	-0.001	-0.003	0.002
B^0 p_{T} re-weighting	0.001	0.001	0.002	0.000	0.000	0.002	0.001
No B^0 p re-weighting	0.000	0.001	0.003	0.001	0.001	0.001	0.001
IP Correction	0.007	0.003	0.006	0.005	0.006	0.004	0.004
Mass resolution Down	-0.003	-0.004	-0.005	-0.003	-0.003	-0.003	-0.004
Mass resolution Up	0.003	0.003	0.004	0.003	0.002	0.003	0.003
Peaking background Down	-0.005	-0.003	-0.004	-0.001	-0.000	-0.000	-0.002
Peaking background Up	0.001	0.001	0.001	0.000	0.000	0.000	0.001
No multiple candidates	-0.005	-0.003	-0.012	-0.010	-0.006	-0.009	-0.005
S-wave component	-0.043	-0.021	-0.038	-0.033	-0.038	-0.026	-0.025

Table 6.3: Systematic uncertainties on the differential branching fraction measurement in each analysis q^2 bin. The 'Up' and 'Down' labels correspond to an upward or downward fluctuation of the relevant uncertainty.

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	0.622	0.295	0.549	0.465	0.546	0.374	0.363
AC weights	-0.053	-0.024	-0.046	0.038	0.051	0.047	-0.031
Trigger efficiency	0.004	0.001	0.001	0.001	0.002	0.001	0.001
DLL binning	-0.019	-0.007	-0.011	-0.001	0.010	0.016	-0.010
IsMuon efficiency	-0.006	0.003	0.004	0.002	0.005	0.006	0.005
Tracking efficiency	0.005	0.002	0.002	0.001	0.002	0.004	0.002
B^0 weighting	0.001	0.001	0.003	0.001	0.001	0.002	0.001
IP Correction	0.007	0.003	0.006	0.005	0.006	0.004	0.004
Mass resolution	-0.003	-0.004	-0.005	-0.003	-0.003	-0.003	-0.004
Peaking background	-0.005	-0.003	-0.004	-0.001	-0.000	0.000	-0.002
No multiple candidates	-0.005	-0.003	-0.012	-0.010	-0.006	-0.009	-0.005
Total	0.058	0.026	0.050	0.040	0.054	0.051	0.034

Table 6.4: Aggregated systematic uncertainties on the differential branching fraction measurement in each analysis q^2 bin.

q^2 range (GeV^2/c^4)	$d\mathcal{B}/dq^2$ ($10^{-7}\,\text{GeV}^{-2}c^4$)
$0.10 < q^2 < 2.00$	$0.62 \pm 0.07(\text{stat}) \pm 0.04(\mathcal{B}) \pm 0.06(\text{syst})^{+0.00}_{-0.04}(\text{S-wave})$
$2.00 < q^2 < 4.30$	$0.30 \pm 0.05(\text{stat}) \pm 0.02(\mathcal{B}) \pm 0.03(\text{syst})^{+0.00}_{-0.02}(\text{S-wave})$
$4.30 < q^2 < 8.68$	$0.55 \pm 0.04(\text{stat}) \pm 0.03(\mathcal{B}) \pm 0.05(\text{syst})^{+0.00}_{-0.04}(\text{S-wave})$
$10.09 < q^2 < 12.86$	$0.47 \pm 0.04(\text{stat}) \pm 0.03(\mathcal{B}) \pm 0.04(\text{syst})^{+0.00}_{-0.03}(\text{S-wave})$
$14.18 < q^2 < 16.00$	$0.55 \pm 0.05(\text{stat}) \pm 0.03(\mathcal{B}) \pm 0.05(\text{syst})^{+0.00}_{-0.04}(\text{S-wave})$
$16.00 < q^2 < 19.00$	$0.37 \pm 0.03(\text{stat}) \pm 0.02(\mathcal{B}) \pm 0.05(\text{syst})^{+0.00}_{-0.03}(\text{S-wave})$
$1.00 < q^2 < 6.00$	$0.36 \pm 0.03(\text{stat}) \pm 0.02(\mathcal{B}) \pm 0.03(\text{syst})^{+0.00}_{-0.03}(\text{S-wave})$

Table 6.5: Results of the differential branching fraction measurement in each analysis q^2 bin.

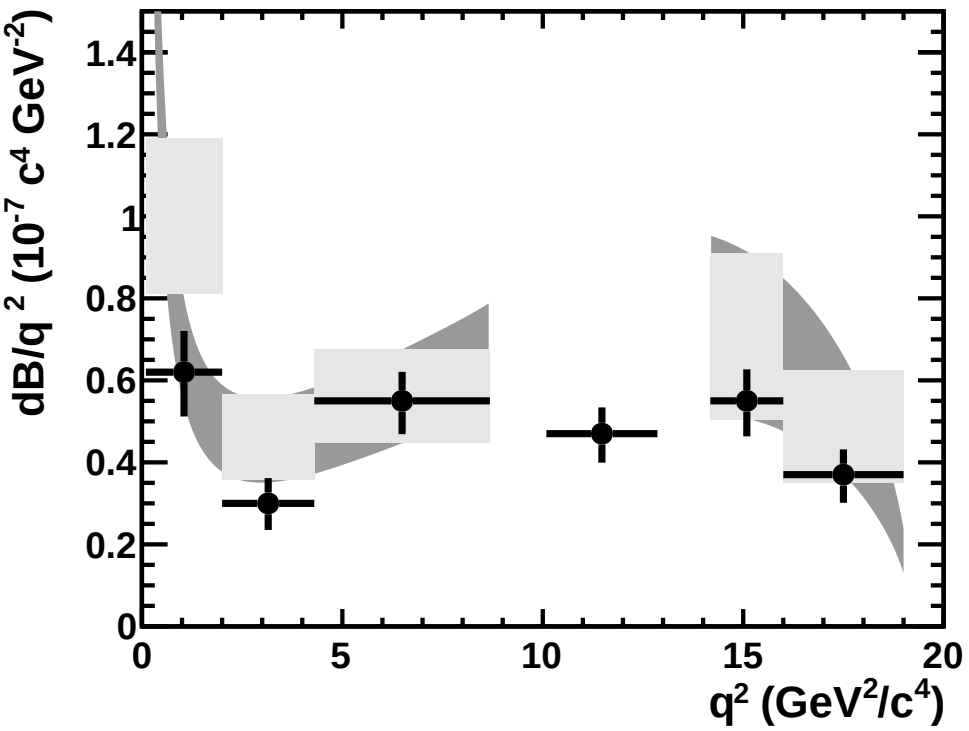


Figure 6.5: Results of the differential branching fraction measurement including statistical and systematic uncertainties (black points). The SM prediction (dark grey) and the SM prediction binned in the analysis q^2 bins (light grey) are indicated.

Chapter 7

Measurement of the angular observables

An unbinned maximum likelihood fit to the distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data is used to extract the $B^0 \to K^{*0}\mu^+\mu^-$ angular observables. The distribution of candidates in $\mathrm{m}_{K\pi\mu\mu}$, $\cos\theta_L$, $\cos\theta_K$ and ϕ is fitted in each analysis q^2 bin. As the $B^0 \to K^{*0}\mu^+\mu^-$ candidates are weighted, the Feldman-Cousins (FC) technique is used to calculate the correct confidence intervals [85].

The angular fit model is discussed in Sec. 7.1. The validity of the fit model is tested, including checks for bias, in Sec. 7.2. The fit results are presented in Sec. 7.3, including discussion of the FC technique, the systematic uncertainties, and the p-value of the data assuming the SM hypothesis.

7.1 The angular fit model

Only $B^0 \to K^{*0}\mu^+\mu^-$ candidates with $5150 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{GeV}/c^2$ are used in the fit. To fit a given process that contributes in this mass range, a model for both the $\mathrm{m}_{K\pi\mu\mu}$ and angular distribution is required. The angular observables are extracted in the analysis q^2 bins so that a q^2 dependence does not need to be incorporated into the fit model. Substantially more $B^0 \to K^{*0}\mu^+\mu^-$ candidates would be required for such a parameterised fit to work.

The contributions to the relevant $\mathrm{m}_{K\pi\mu\mu}$ mass region are: $B^0 \to K^{*0}\mu^+\mu^-$ decays; $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ decays; and Class I, II and III background events. The $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of the $B^0 \to K^{*0}\mu^+\mu^-$ decays is parameterised by the M_{B^0} function described in Sec. 6.2.1, denoted the 'signal mass pdf', while their angular distribution is parameterised according to Eq. (2.5.34) or Eq. (2.5.38), denoted the 'signal angular pdf'.

The $\mathrm{m}_{K\pi\mu\mu}$ mass distribution of Class III background events is parameterised by the

exponential function $E_{bkg}(m|p_0)$, denoted the 'background mass pdf', while their angular distribution is parameterised by second-order polynomial functions in each decay angle, denoted the 'background angular pdf'. The linear component of each polynomial is able to model an asymmetric distribution of Class III background events in a given decay angle. This is particularly important for the measurement of $A_{\rm FB}$ ($A_{\rm T}^{\rm R}$) as this observable is related to an asymmetry in the $\cos\theta_L$ distribution of $B^0 \to K^{*0}\mu^+\mu^-$ decays (see Eq. (2.5.34)). The quadratic component of each polynomial is able to model any peaking structure at $\cos\theta_L \sim \pm 1$ and $\cos\theta_K \sim \pm 1$ and any modulations in the ϕ distribution of Class III background events. The latter is particularly important for the measurement of S_3 ($A_{\rm T}^2$), S_9 ($A_{\rm T}^{\rm I}$) and A_9 as these observables are related to sinusoidal modulations in the ϕ distribution (see Eq. (2.5.34)). The decay angles for Class III background events are assumed to factorise. The quality of the background angular fit, including the assumption of factorisation of the decay angles, is tested in Sec. 7.2.2.1. Since the goodness-of-fit of the background angular model depends on the candidate weights, the variation of the fit under fluctuations of the candidate weights is considered as a source of systematic uncertainty in Sec. 7.3.2.1.

The probability distribution of the background angular pdf is

$$\begin{aligned} &P_{bkg}(\cos\theta_L, \cos\theta_K, \phi) \\ &\quad = P^L_{bkg}(\cos\theta_L) P^K_{bkg}(\cos\theta_K) P^\phi_{bkg}(\phi) \\ &\quad = \left(\sum_{k=0}^{n} c_k^L T_k(\cos\theta_L)\right)\left(\sum_{k=0}^{n} c_k^K T_k(\cos\theta_K)\right)\left(\sum_{k=0}^{n} c_k^\phi T_k(\phi)\right), \end{aligned} \tag{7.1.1}$$

where T_k is a $k^{\rm th}$ order Chebychev polynomial of the first kind; and c_k^X are the coefficients associated with the polynomials for $X = \{\cos\theta_L, \cos\theta_K, \phi\}$. For the present analysis $n = 2$. The goodness-of-fit of the background angular pdf is tested in Sec. 7.2.2.

The contributions from Class I and Class II background candidates to the signal $\mathrm{m}_{K\pi\mu\mu}$ mass region and to the full $\mathrm{m}_{K\pi\mu\mu}$ mass region are estimated to be at the level of $\sim$4% and $\sim$10% of the $B^0 \to K^{*0}\mu^+\mu^-$ yield respectively (see Tab. 4.3). In general the differential branching fraction and angular distribution of these backgrounds has not been measured and so cannot be accurately incorporated into the fit model. These contributions are therefore neglected in the fit and are considered as a source of systematic uncertainty in Sec. 7.3.2. The contribution from $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ is estimated to be at the level of $\sim$1% of the $B^0 \to K^{*0}\mu^+\mu^-$ yield in Sec. 4.3.2. This contribution is neglected in the fit due to the small number of $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ candidates expected in the data. This introduces a negligible systematic uncertainty.

The product of the signal mass pdf and the signal angular pdf forms the four dimen-

sional 'signal pdf'. By forming this product, the angular distribution is assumed to be uncorrelated to the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution. The angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ decays does not depend on $\mathrm{m}_{K\pi\mu\mu}$, and any difference caused by the detector, trigger and selection in the $50\,\mathrm{MeV}/c^2$ signal window (over 95% of $B^0 \to K^{*0}\mu^+\mu^-$ decays are within the signal window) is verified to be negligible using simulated events. This assumption again contributes a negligible systematic uncertainty. The product of the background mass pdf and background angular pdf forms the four dimensional 'background pdf'. Again by forming this product the angular distribution is assumed to be uncorrelated to the $\mathrm{m}_{K\pi\mu\mu}$ mass distribution. This assumption is tested in Sec. 7.2.2.

The sum of the signal pdf and the background pdf forms the angular fit pdf. The probability distribution of the angular fit pdf is

$$\begin{aligned} &P(m, \cos\theta_L, \cos\theta_K, \phi) \\ &\quad = f_{\mathrm{sig}}\left[M_{\mathrm{B^0}} \times P_{sig}(\cos\theta_L, \cos\theta_K, \phi)\right] \\ &\quad + (1 - f_{\mathrm{sig}})\left[E_{bkg}(m|p_0) \times P_{bkg}(\cos\theta_L, \cos\theta_K, \phi)\right] \end{aligned} \tag{7.1.2}$$

where: m is the $\mathrm{m}_{K\pi\mu\mu}$ value of a candidate; $P_{sig}(\cos\theta_L, \cos\theta_K, \phi)$ is the signal angular pdf; $P_{bkg}(\cos\theta_L, \cos\theta_K, \phi)$ is the background angular pdf; and f_{sig} is the signal fraction, which is the relative contribution of the signal pdf to the full pdf. There are twelve free parameters in the angular fit: the four angular observables; the signal fraction f_{sig}; the six coefficients of the background angular model c_k^L, c_k^K and c_k^ϕ for $k = 1, 2$; and the exponent p_0.

7.1.1 Experimental issues with the signal angular pdf

The signal angular pdf is expressed in terms of four angular observables. These observables are not independent of each other, as they are expressed in terms of the same transversity amplitudes. Also, the J_{6s} and J_9 angular terms change sign if the flavour of the B^0 meson is misidentified. The misidentification of the B^0 flavour is discussed in Sec. 7.1.1.1. The impact of the correlated observables is discussed in Sec. 7.1.1.2.

7.1.1.1 Misidentification of the B^0 flavour

The decays of B^0 and $\overline{B}^0$ mesons are distinguished by the charge of the kaon, with a K^+ indicating a B^0 and K^- indicating a $\overline{B}^0$. When the K^+ from a $B^0 \to K^{*0}\mu^+\mu^-$ decay is misidentified as a π^+ and the π^- from the same decay is misidentified as a K^-, then the B^0 flavour is also misidentified. If a B^0 decay is misidentified as a $\overline{B}^0$ decay the wrong angular definition is used, which results in the observables A_{FB} ($A_{\mathrm{T}}^{\mathrm{R}}$), S_9 ($A_{\mathrm{T}}^{\mathrm{I}}$) and

A_9 flipping sign. This 'dilutes' the measurement of the angular observables, biasing the measured value towards zero. The dilution is accounted for by transforming

$$O_{\text{-ve}} \to O_{\text{-ve}}(1 - 2\omega_{\text{ID}}) \tag{7.1.3}$$

in the signal angular pdf, where $O_{\text{-ve}}$ is any of the observables indicated above and ω_{ID} is the fraction of $B^0 \to K^{*0}\mu^+\mu^-$ decays that are misidentified as described above. The fraction of misidentified $B^0 \to K^{*0}\mu^+\mu^-$ decays is estimated using simulated events to be $\omega_{\text{ID}} = (1.05 \pm 0.11)\%$ (see Tab. 4.3). The uncertainty on ω_{ID} is due to the limited number of simulated $B^0 \to K^{*0}\mu^+\mu^-$ events and is considered as a source of systematic uncertainty in Sec. 7.3.2. The variation of ω_{ID} as a function of q^2 is about 1% of the value of ω_{ID} (see App. G). This variation is a negligible source of systematic uncertainty.

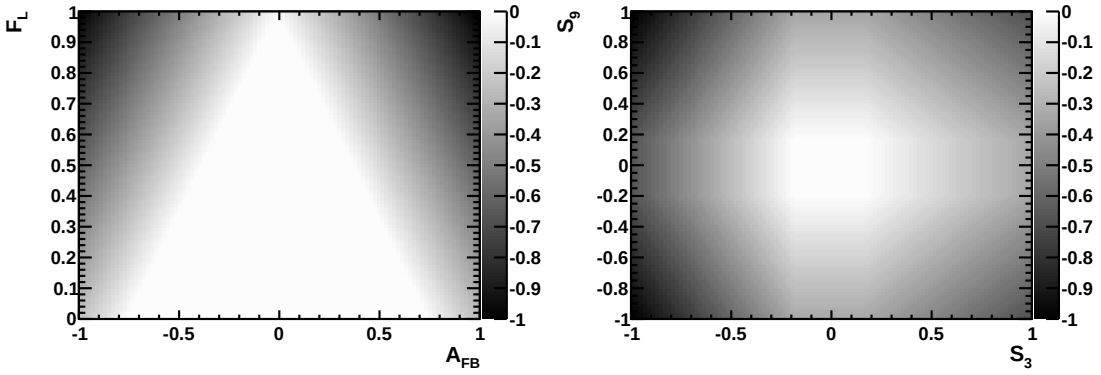
7.1.1.2 Boundaries in the signal angular pdf

The normalisation and correlation of the angular observables results in 'physical' and 'mathematical' boundaries in the signal angular pdf. Physical boundaries refer to the physically valid range of the angular observables, for example A_{FB} is defined in the range $-1.0 < A_{\text{FB}} < 1.0$ and F_{L} in the range $0.0 < F_{\text{L}} < 1.0$. These boundaries arise as the observables are normalised to the total rate.

Mathematical boundaries refer to the signal angular pdf becoming non-positive in some region of the angular space due to certain combinations of values of the angular observables. This feature is caused by the correlations between the angular observables. The 'mathematically allowed region' refers to all combinations of angular observables for which the signal angular pdf is positive in all of the angular space. If the fraction of longitudinally polarised decays, F_{L}, is large, then (by definition) most of the $B^0 \to K^{*0}\mu^+\mu^-$ decays are not transversely polarised. As $F_{\text{L}} \to 1$ the number of transversely polarised $B^0 \to K^{*0}\mu^+\mu^-$ decays goes to zero, and by definition the FFD observables A_{FB}, S_3 and (S_9, A_9) must therefore also go to zero: there can be no asymmetry in transversely polarised decays when there are no transversely polarised decays. The relationship between F_{L} and each transverse FFD observable is

$$\begin{aligned} |A_{\text{FB}}| &\leq \tfrac{3}{4}(1 - F_{\text{L}}), \\ |S_3| &\leq \tfrac{1}{2}(1 - F_{\text{L}}), \\ |S_9| &\leq \tfrac{1}{2}(1 - F_{\text{L}}). \end{aligned} \tag{7.1.4}$$

Since only transversely polarised $B^0 \to K^{*0}\mu^+\mu^-$ decays provide sensitivity to the transverse angular observables, these correlations also imply that when F_{L} is large, the preci-



(a) ($A_{\rm FB}$, $F_{\rm L}$) (b) (S_3, S_9)

Figure 7.1: Mathematical boundaries between the observables ($A_{\rm FB}$, $F_{\rm L}$) and (S_3, S_9). Grey regions of the plot indicate that the signal angular pdf has become negative, where increasingly darker shades of grey indicate an increasingly negative pdf. The colour scale is normalised such that the most negative point in the 2D projection appears black.

sion of the measurement of the remaining FFD observables is reduced. There are more complicated relationships than those of Eq. (7.1.4) which cannot be expressed so simply, such as the relationship between S_3, S_9 and $F_{\rm L}$ in Eq. (2.5.41).

Due to the limited number of candidates in the data the maximum likelihood obtained in the fit can be outside of the physical and mathematical boundaries. The signal angular pdf is then negative in some region of the angular space. The extent of the negative region $pdf_{\text{-ve}}$ can be calculated for given values of the angular observables by evaluating the pdf at an ensemble of points in the angular space. To calculate $pdf_{\text{-ve}}$ the value of the pdf is summed at each point the pdf is less than zero, *i.e.*,

$$pdf_{\text{-ve}} = \sum_{i=0}^{i=i_{\rm max}} \text{eval}(pdf; i) I(pdf; i), \tag{7.1.5}$$

where i is a point in the ensemble, $i_{\rm max}$ is the number of points in the ensemble, eval($pdf; i$) is the value of the pdf at point i and $I(pdf; i)$ is 1 when eval($pdf; i$) < 0 and zero otherwise.

The extent of the negative region of the pdf is plotted as a function of the pairs of angular observables ($A_{\rm FB}$, $F_{\rm L}$) and (S_3, S_9) in Fig. 7.1. The remaining observables are set to the SM prediction for the $1.00 < q^2 < 6.00\,{\rm GeV}^2/c^4$ analysis q^2 bin. In each plot $pdf_{\text{-ve}}$ is normalised to $-1 < pdf_{\text{-ve}} < 0$ and darker grey regions indicate larger $pdf_{\text{-ve}}$. The mathematical boundaries in Eq. (7.1.4) are visible, with $pdf_{\text{-ve}}$ becoming more negative as the angular observables move away from the mathematically allowed region.

Fitting without constraining the angular observables to the physically and mathematically allowed regions will ensure an unbiased maximum likelihood estimator. Since a pdf

that becomes negative cannot be correctly normalised, and to avoid problems generating simulated events, both the physical and mathematical boundaries are incorporated into the fit. The physical boundaries are incorporated simply as limits on the values of each angular observable. The mathematical boundaries are incorporated by introducing a penalty term in the likelihood that is equal to $e^{pdf\text{-ve}}$. This increase in the likelihood when outside the mathematical boundaries results in the fit converging within the mathematically allowed region. This same technique is also used to enforce that the background angular pdf is positive, which ensures that the angular fit pdf is positive.

Constraining the fit in this way biases the maximum likelihood estimator for the angular observables. This bias is discussed in App. E. Using the FC technique ensures that the correct confidence intervals are obtained.

7.2 The validity of the fit model

Any bias in the fit causes the mis-measurement of the angular observables and the calculation of incorrect confidence intervals. Bias can be introduced as a result of the fit model inaccurately describing the distribution of candidates in the data. This bias is distinct from that of the biased maximum likelihood estimator discussed in the previous section.

In Sec. 6 the signal mass pdf and background mass pdf are shown to accurately describe the $\mathrm{m}_{K\pi\mu\mu}$ distribution of $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to J/\psi\, K^{*0}$ candidates in the data. The signal angular pdf is derived theoretically and is assumed to accurately describe the angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ decays. Finally, the background angular pdf must accurately describe the angular distribution of Class III background events.

The background angular pdf factorises the decay angles, *i.e.*

$$f(\cos\theta_L, \cos\theta_K, \phi) \equiv f(\cos\theta_L) \times f(\cos\theta_K) \times f(\phi), \tag{7.2.1}$$

where $f(\cos\theta_L, \cos\theta_K, \phi)$ is the angular distribution of Class III background candidates, and $f(\cos\theta_L)$, $f(\cos\theta_K)$ and $f(\phi)$ is the distribution of those candidates in each decay angle (each projection of the angular space). This factorisation must be valid to avoid introducing bias in the fit, since the decay angles are not factorised in the signal angular pdf. The decay angles are also factorised with $\mathrm{m}_{K\pi\mu\mu}$. If this factorisation is not valid the background angular pdf will not describe the shape of the Class III and Class II backgrounds in the signal window. This would result in the mis-measurement of the angular observables. The validity of this factorisation is sensitive to the contribution of Class I and Class II background events, as these events do not contribute above $\mathrm{m}_{K\pi\mu\mu} \sim 5450\,\mathrm{MeV}/c^2$.

A powerful validation of the total fit is provided by the $B^0 \to J/\psi K^{*0}$ decay. The $B^0 \to J/\psi K^{*0}$ angular observables can be calculated using the measured $B^0 \to J/\psi K^{*0}$ decay amplitudes measured by the BaBar collaboration, allowing comparison to the results of the angular fit to $B^0 \to J/\psi K^{*0}$ candidates in the data. This validation is sensitive not only to any bias caused by the inaccurate description of the data by the angular fit pdf, but also to the AC weights calculated in Sec. 5.

A check for bias in the angular fit procedure is detailed in Sec. 7.2.1. A test of the validity of the background angular model is detailed in Sec. 7.2.2. The fit is validated using $B^0 \to J/\psi K^{*0}$ candidates in the data in Sec. 7.2.3.

7.2.1 Checking for bias in the angular fit

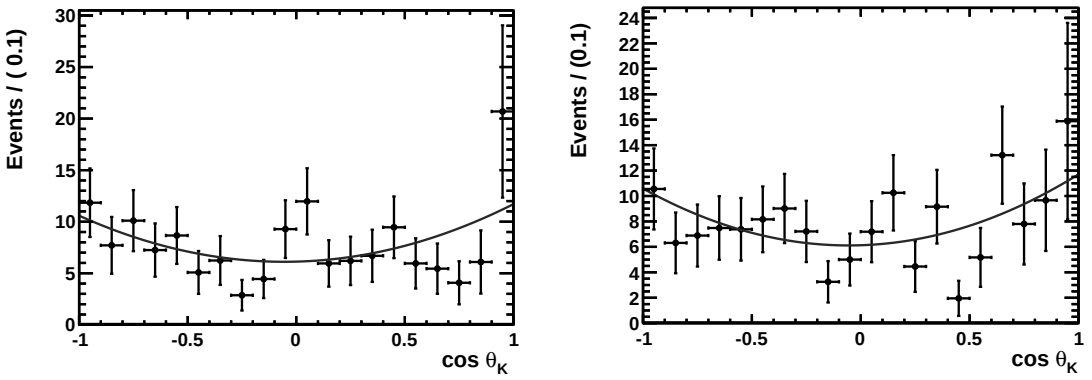
Checks for fit bias can be performed by fitting simulated events and checking for shifts in the mean of the pull distributions of each observable. Simulated data sets are generated according to the procedure outlined in Sec. 7.2.1.1. The checks for fit bias are described in Sec. 7.2.1.2.

7.2.1.1 Generating simulated data sets

Simulated data sets are generated using the angular fit pdf. The parameters of the signal and background mass pdfs are taken from the differential branching fraction fits shown in Fig. 6.2. These fits also define the number of events in each data set, and the fraction of those events that are generated from the signal or background pdf. The parameters of the signal angular pdf are set to the values predicted in the SM. The parameters of the background angular pdf are set to zero, such that the background events are distributed uniformly in the decay angles.

To mimic the acceptance effects the simulated events are accept-rejected using an AC pdf. The AC pdf is chosen according to the average q^2 value of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data in the relevant analysis q^2 bin. An AC weight is calculated for, and applied to, each event to mimic the acceptance correction.

The $\cos\theta_K$ distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $5450 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$ is shown in Fig. 7.2(a) with a second order polynomial fir overlaid. A simulated data set generated from the fitted polynomial, and weighted according to the scheme detailed above, is shown in Fig. 7.2(b). In both plots the highest bin has a large uncertainty. This indicates that the events in this bin have large weights, as expected at $\cos\theta_K \sim +1$. The (qualitative) similarity of the plots indicates that the procedure described above mimics the acceptance effects and acceptance correction in simulated events.



(a) data (b) simulation

Figure 7.2: The $\cos\theta_K$ distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $5450 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$ with the background angular fit pdf overlaid (a). The $\cos\theta_K$ distribution of a simulated events that are generated according to the same fit pdf, which are weighted according to the procedure detailed in Sec. 7.2.1.1 (b).

7.2.1.2 Checking for bias in the fit

To check for any bias in the fit (which is not a result of the physical boundaries or the mathematical boundaries) an ensemble of simulated data sets is generated with the predicted values of the FFI observables in the SM. Each data set in the ensemble corresponds to $20\,\mathrm{fb}^{-1}$ of integrated luminosity. This large integrated luminosity is used to negate the effects of the physical boundaries. The SM predicted values allow the fit bias to be examined at different points within the physical and mathematical boundaries. Extracting the FFI observables removes the correlations expressed in Eq. (7.1.4) and so removes the large mathematical boundaries associated with the FFD observables. The SM predicts $A_{\mathrm{T}}^{\mathrm{R}} \sim 1$ in the $14.18 < q^2 < 16.00\,\mathrm{GeV}^2/c^4$ analysis q^2 bin. If no bias is observed in this observable and analysis q^2 bin, then $20\,\mathrm{fb}^{-1}$ is sufficient to negate any bias introduced by the physical boundaries.

The simulated data sets are fitted to obtain a distribution of values of each angular observable in each analysis q^2 bin. The pull of each value can be calculated from the value of the SM prediction. A Gaussian fit is used to extract the mean and width of each corresponding pull distribution. The results are shown in Tab. 7.1. The mean of the pull distribution of $A_{\mathrm{T}}^{\mathrm{R}}$ in the $14.18 < q^2 < 16.00\,\mathrm{GeV}^2/c^4$ analysis q^2 bin is consistent with zero, indicating that the number of events in the data sets is sufficiently large to negate the effects of the physical boundaries. The mean of each of the pull distributions is consistent with zero. With a data set corresponding to $20\,\mathrm{fb}^{-1}$ of integrated luminosity and SM values of the angular observables, there is no evidence for bias in the fit.

The confidence intervals returned from the fit are calculated using the profile likelihood

q^2 range (GeV^2/c^4)	quantity	$A_{\mathrm{T}}^{\mathrm{R}}$	F_{L}	A_{T}^{2}	$A_{\mathrm{T}}^{\mathrm{I}}$
$0.10 < q^2 < 2.00$	mean	0.01±0.04	-0.02±0.03	0.01±0.02	0.02±0.02
$0.10 < q^2 < 2.00$	width	1.47±0.05	1.28±0.03	0.90±0.01	0.88±0.01
$2.00 < q^2 < 4.30$	mean	0.01±0.02	0.02±0.02	-0.00±0.02	0.01±0.02
$2.00 < q^2 < 4.30$	width	1.03±0.02	1.07±0.02	0.91±0.01	0.93±0.02
$4.30 < q^2 < 8.68$	mean	-0.01±0.02	-0.01±0.02	-0.00±0.02	-0.01±0.02
$4.30 < q^2 < 8.68$	width	0.95±0.02	1.08±0.02	0.96±0.02	0.94±0.02
$10.09 < q^2 < 12.86$	mean	0.02±0.02	0.01±0.02	0.00±0.02	0.01±0.02
$10.09 < q^2 < 12.86$	width	0.93±0.02	1.10±0.02	0.95±0.02	0.99±0.02
$14.18 < q^2 < 16.00$	mean	0.02±0.02	0.00±0.02	0.01±0.02	0.01±0.02
$14.18 < q^2 < 16.00$	width	0.97±0.02	1.10±0.02	0.94±0.02	0.97±0.02
$16.00 < q^2 < 19.00$	mean	-0.02±0.02	0.03±0.02	0.02±0.02	0.01±0.02
$16.00 < q^2 < 19.00$	width	0.97±0.02	1.11±0.02	0.94±0.02	0.94±0.02
$1.00 < q^2 < 6.00$	mean	0.04±0.02	-0.03±0.02	-0.04±0.02	-0.01±0.02
$1.00 < q^2 < 6.00$	width	1.02±0.02	1.10±0.02	0.92±0.02	0.93±0.02

Table 7.1: Measurements of the mean and width of the pull distribution for each FFI angular observable for each analysis q^2 bin.

$\mathcal{L}_{\mathrm{p}}$. The coverage of these confidence intervals is incorrect when the fit uses weighted events. This effect is seen in Tab. 7.1. Each pull width deviates from unity due to the candidate weights. The candidate weights are dominated by the AC weights and are therefore largest in regions of the angular space where the acceptance effects are largest. The large acceptance effects in $\cos\theta_K$, which vary slowly in q^2, lead to the confidence intervals on F_{L} being too small, causing the width of the pull distribution to be greater than unity. The large acceptance effects in $\cos\theta_L$ at low q^2, which reduce rapidly as q^2 increases, lead to the confidence intervals for $A_{\mathrm{T}}^{\mathrm{R}}$ in the $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$ and $2.00 < q^2 < 4.30\,\mathrm{GeV}^2/c^4$ analysis q^2 bins being too small. Conversely the uncertainties for A_{T}^{2} and $A_{\mathrm{T}}^{\mathrm{I}}$ are too large, causing the width of the pull distributions to be less than unity. This indicates that these observables are constrained by an additional mathematical boundary that is not shown in Eq. (7.1.4).

7.2.2 Testing the background angular model

The goodness-of-fit of the background angular pdf is tested in each analysis q^2 bin using a point-to-point dissimilarity test (DT_{ptp}) [86, 87]. This test is described in App. F. The validation of the background angular pdf is described in Sec. 7.2.2.1.

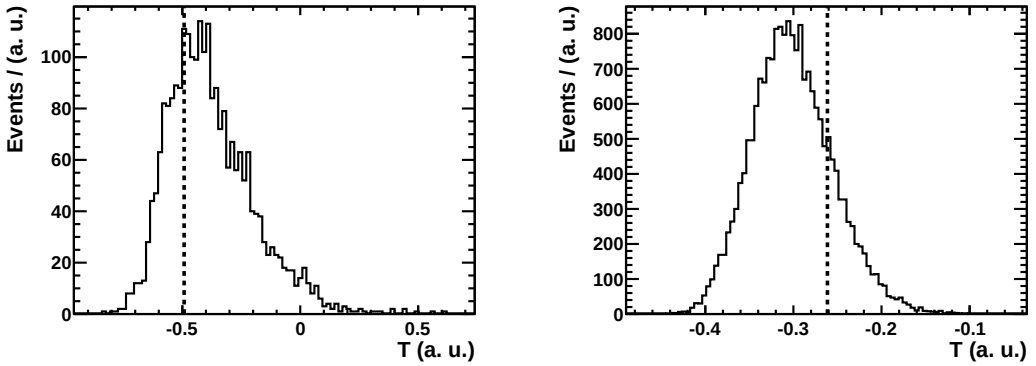
q^2 range (GeV^2/c^4)	p-value
$0.10 < q^2 < 2.00$	0.47
$2.00 < q^2 < 4.30$	0.72
$4.30 < q^2 < 8.68$	0.38
$10.09 < q^2 < 12.86$	0.16
$14.18 < q^2 < 16.00$	0.51
$16.00 < q^2 < 19.00$	0.93
$1.00 < q^2 < 6.00$	0.73
$9.21 < q^2 < 9.96$	0.20

Table 7.2: The p-value for the goodness-of-fit of the background angular pdf to $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $5450 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. The p-value is calculated in the analysis q^2 bins and $9.21 < q^2 < 9.96\,\mathrm{GeV}^2/c^4$, which contains Class III candidates in the $B^0 \to J/\psi\, K^{*0}$ upper $\mathrm{m}_{K\pi\mu\mu}$ mass sideband.

7.2.2.1 Validation of the background angular model

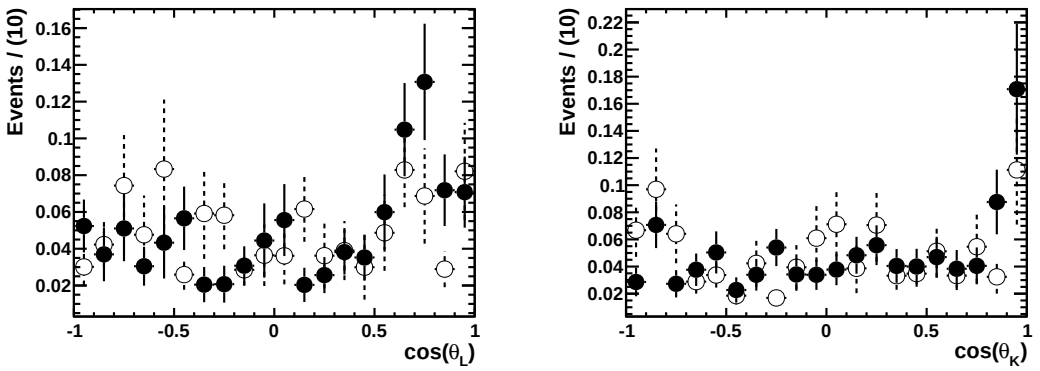
The goodness-of-fit of the background angular pdf to the angular distribution of Class III candidates is calculated using a DT_{ptp} with $\vec{x} = \{\cos\theta_L,\ \cos\theta_K,\ \phi\}$ and $\bar{\sigma} = 0.05$ (so that on average Ψ covers 5% of $\vec{x}$). The Class III candidates are required to satisfy $5450 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. The goodness-of-fit is calculated for each of the analysis q^2 bins, as well as for $9.21 < q^2 < 9.96\,\mathrm{GeV}^2/c^4$ to examine the Class III candidates in the $B^0 \to J/\psi\, K^{*0}$ upper mass sideband. The results of the DT_{ptp} are shown in Tab. 7.2. The background angular pdf accurately models the angular distribution of Class III background candidates in each of the q^2 regions. The distribution of $\vec{T}_{\mathrm{sim}}$ using candidates with $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$ and $9.21 < q^2 < 9.96\,\mathrm{GeV}^2/c^4$ are shown, with the corresponding value of T_{data} indicated, in Fig. 7.3.

The fit uses candidates in the data with $5150 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. No variation of the background angular pdf in $\mathrm{m}_{K\pi\mu\mu}$ is accounted for in the fit model, as the background angular pdf and background mass pdf are factorised. Such a factorisation may not be valid due to contributions of Class I and Class II background events, which do not contribute above $\mathrm{m}_{K\pi\mu\mu} \sim 5450\,\mathrm{MeV}/c^2$. The validity of the factorisation is tested by comparing the angular distribution of Class III background candidates with $5150 < \mathrm{m}_{K\pi\mu\mu} < 5220\,\mathrm{MeV}/c^2$ and $5450 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. The distributions in each decay angle are shown in Fig. 7.4. A binned χ^2 test on the two distributions returns p-values of: 0.21 for $\cos\theta_L$; 0.27 for $\cos\theta_K$ and 0.37 for ϕ. There is no evidence for a different angular distribution of Class III background events in the two regions of $\mathrm{m}_{K\pi\mu\mu}$.

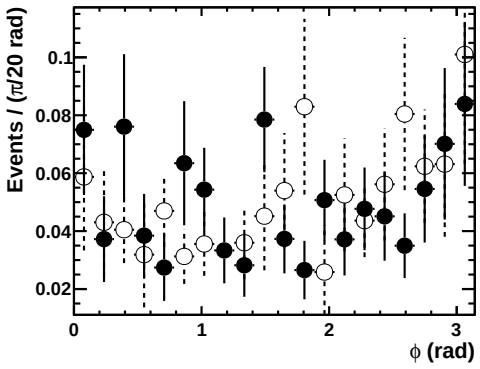


(a) $B^0 \to K^{*0}\mu^+\mu^-$ (b) $B^0 \to J/\psi\, K^{*0}$

Figure 7.3: Distribution of $\vec{T}_{\rm sim}$ generated from the DT$_{ptp}$ when performed using $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $1.00 < q^2 < 6.00\,{\rm GeV}^2/c^4$ and $5450 < {\rm m}_{K\pi\mu\mu} < 5750\,{\rm MeV}/c^2$ (a) and $B^0 \to J/\psi\, K^{*0}$ candidates in the data with $9.21 < q^2 < 9.96\,{\rm GeV}^2/c^4$ and $5450 < {\rm m}_{K\pi\mu\mu} < 5750\,{\rm MeV}/c^2$ (b). The corresponding values of $T_{\rm data}$ are indicated by the dashed line.



(a) $\cos\theta_L$ (b) $\cos\theta_K$



(c) ϕ

Figure 7.4: Angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data in the decay angles $\cos\theta_L$ (a), $\cos\theta_K$ (b) and ϕ (c), with $5450 < {\rm m}_{K\pi\mu\mu} < 5750\,{\rm MeV}/c^2$ (circles) and with $5150 < {\rm m}_{K\pi\mu\mu} < 5220\,{\rm MeV}/c^2$ (open circles).

Observable	BaBar Measurement	This Measurement
$A_{\mathrm{T}}^{\mathrm{R}}$	N/A	0.007 ± 0.007
F_{L}	0.556 ± 0.013	0.559 ± 0.003
A_{T}^{2}	0.050 ± 0.037	0.046 ± 0.015
$A_{\mathrm{T}}^{\mathrm{I}}$	-0.428 $\pm$ 0.096	-0.369 $\pm$ 0.018

Table 7.3: Results of the BaBar measurements of the decay amplitudes of the $B^0 \to J/\psi K^{*0}$ decay, having been converted to the angular observables F_{L}, A_{T}^{2} and $A_{\mathrm{T}}^{\mathrm{I}}$. There is no term corresponding to $A_{\mathrm{T}}^{\mathrm{R}}$ in the $B^0 \to J/\psi K^{*0}$ angular distribution. The results of the angular fit to $B^0 \to J/\psi K^{*0}$ candidates in the data are also shown. The uncertainties are statistical only.

7.2.3 Testing the angular model using $B^0 \to J/\psi K^{*0}$

The validity of the angular fit is established by measuring the angular observables of the $B^0 \to J/\psi K^{*0}$ decay and comparing the results to those of previous experiments.

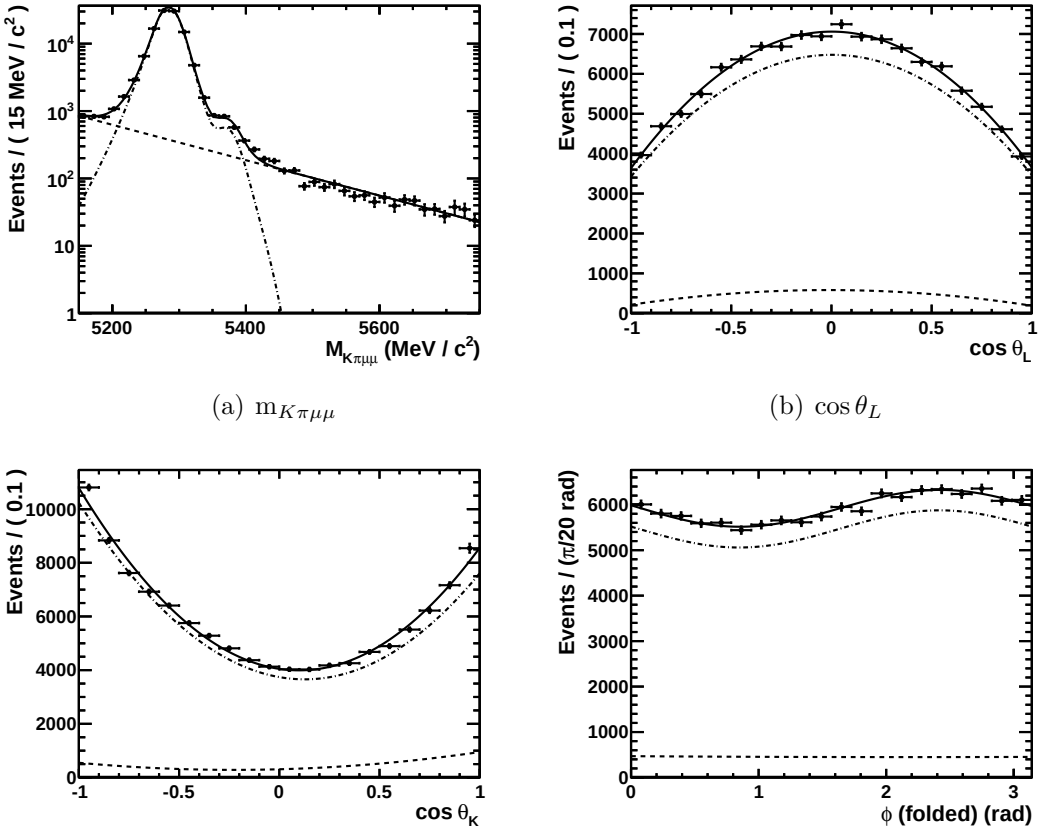
The most precise measurements of the $B^0 \to J/\psi K^{*0}$ transverse amplitudes are published by the BaBar collaboration [88]. The angular observables are calculated using these measurements and the expressions in Sec. 2.5.2. The calculated values are shown in Tab. 7.3. There is no value for $A_{\mathrm{T}}^{\mathrm{R}}$ as the relevant term is not present in the $B^0 \to J/\psi K^{*0}$ angular distribution [88].

Two modifications of the fit pdf are required to accurately model the distribution of $B^0 \to J/\psi K^{*0}$ candidates in the data. To account for the contribution of $\overline{B}{}_s^0 \to J/\psi K^{*0}$ events an additional component is incorporated to the fit pdf. The $M_{\mathrm{B_s^0}}$ function described in Sec. 6.2.1 is used as the mass pdf. The $\overline{B}{}_s^0 \to J/\psi K^{*0}$ angular distribution is assumed to be uniform to avoid further complicating the fit. An S-wave contribution to the $B^0 \to J/\psi K^{*0}$ decay, which constitutes about 7% of $B^0 \to J/\psi K^{*0}$ decays, has been measured by the BaBar collaboration [41, 84]. An S-wave component is incorporated to the fit pdf by replacing the nominal signal angular pdf with Eq. (2.5.48).

The measurement of the $B^0 \to J/\psi K^{*0}$ angular observables, accounting for statistical uncertainties only, is summarised in Tab. 7.3. The angular distribution of the $B^0 \to J/\psi K^{*0}$ candidates is shown in Fig. 7.5 with the fit overlaid. The measurements are consistent with those of the BaBar experiment. There is no evidence for bias in the fit or any other mis-measurement of the angular observables. This also further validates the AC weights calculated in Sec. 5.

7.2.4 Summarising the validity of the angular fit model

A number of checks are performed to ensure that the fit model is valid and that an accurate measurement of the angular observables is obtained. There is no evidence for bias



(a) $\mathrm{m}_{K\pi\mu\mu}$ (b) $\cos\theta_L$

(c) $\cos\theta_K$ (d) ϕ

Figure 7.5: The $\mathrm{m}_{K\pi\mu\mu}$ and angular distribution of $B^0 \to J/\psi\, K^{*0}$ candidates in $\mathrm{m}_{K\pi\mu\mu}$ (a) and the decay angles $\cos\theta_L$ (b), $\cos\theta_K$ (c) and ϕ (d), for $B^0 \to J/\psi\, K^{*0}$ candidates in the data. The total fit pdf (solid black line), the combined $B^0 \to J/\psi\, K^{*0}$ and $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ component (blue dot-dashed line) and the Class III background component (red dashed line) are indicated.

q^2 range (GeV^2/c^4)	A_{FB}	F_{L}	S_3	S_9	A_9	$A_{\mathrm{T}}^{\mathrm{R}}$	A_{T}^2	$A_{\mathrm{T}}^{\mathrm{I}}$
$0.10 < q^2 < 2.00$	-0.11	0.26	-0.05	-0.04	0.12	-0.20	-0.15	-0.10
$2.00 < q^2 < 4.30$	-0.11	0.82	-0.06	-0.02	0.08	-0.82	-0.61	-0.22
$4.30 < q^2 < 8.68$	0.17	0.63	0.06	0.01	-0.12	0.61	0.32	0.05
$10.09 < q^2 < 12.86$	0.28	0.47	-0.20	-0.04	0.08	0.72	-0.77	-0.17
$14.18 < q^2 < 16.00$	0.47	0.34	-0.10	-0.00	-0.07	0.94	-0.30	-0.01
$16.00 < q^2 < 19.00$	0.28	0.41	-0.16	-0.03	0.06	0.62	-0.52	-0.10
$1.00 < q^2 < 6.00$	-0.08	0.62	0.03	0.07	-0.01	-0.29	0.15	0.36

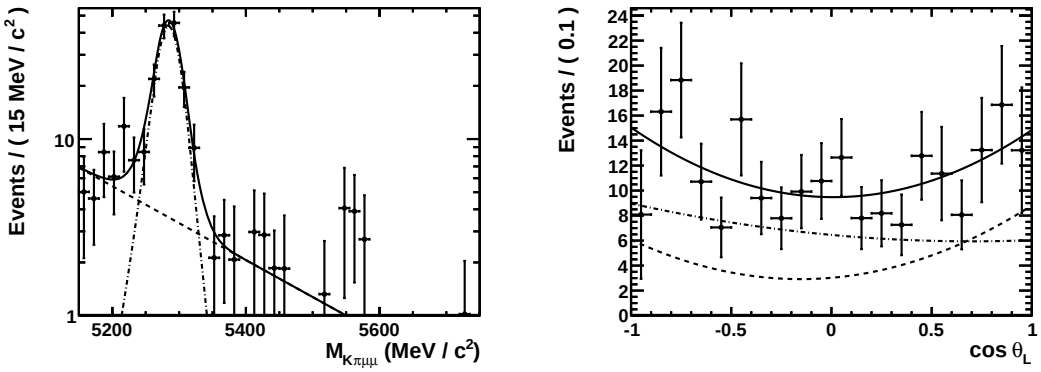
Table 7.4: Central values of the measurements of the $B^0 \to K^{*0}\mu^+\mu^-$ angular observables.

in the fit. The background angular model accurately describes the angular distribution of Class III background candidates. There is no evidence for significant variation of the angular distribution of Class III background candidates with $\mathrm{m}_{K\pi\mu\mu}$. The $B^0 \to J/\psi\, K^{*0}$ angular observables are measured using candidates in the data, and the results are consistent with those of the BaBar collaboration. These checks show that the fit accurately measures the $B^0 \to K^{*0}\mu^+\mu^-$ angular observables.

7.3 Extracting the angular observables

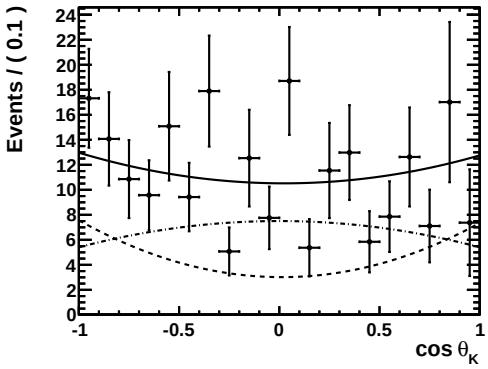
The angular observables are extracted using an unbinned maximum likelihood fit to $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data. As the candidates are weighted, the correct confidence intervals are obtained using the FC technique in Sec. 7.3.1. The systematic uncertainties on the angular observables are discussed in Sec. 7.3.2.

The measured central value of each angular observable, in each analysis q^2 bin, is summarised in Tab. 7.4. The $\mathrm{m}_{K\pi\mu\mu}$ and angular distribution of the $B^0 \to K^{*0}\mu^+\mu^-$ candidates in each analysis q^2 bin are shown with the angular fit pdf overlaid in Figs. 7.6 to 7.12.

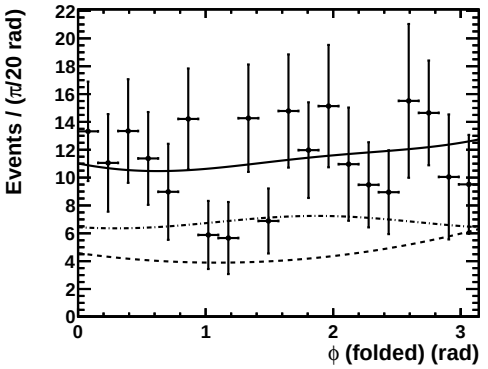


(a) $\mathrm{m}_{K\pi\mu\mu}$

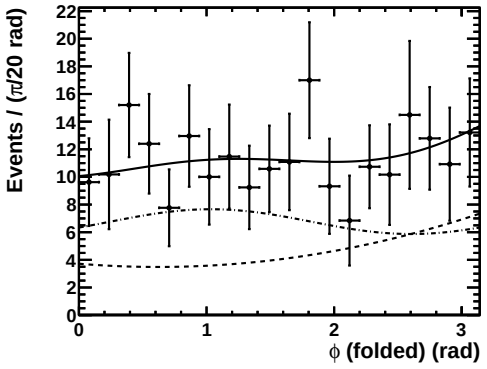
(b) $\cos\theta_L$



(c) $\cos\theta_K$

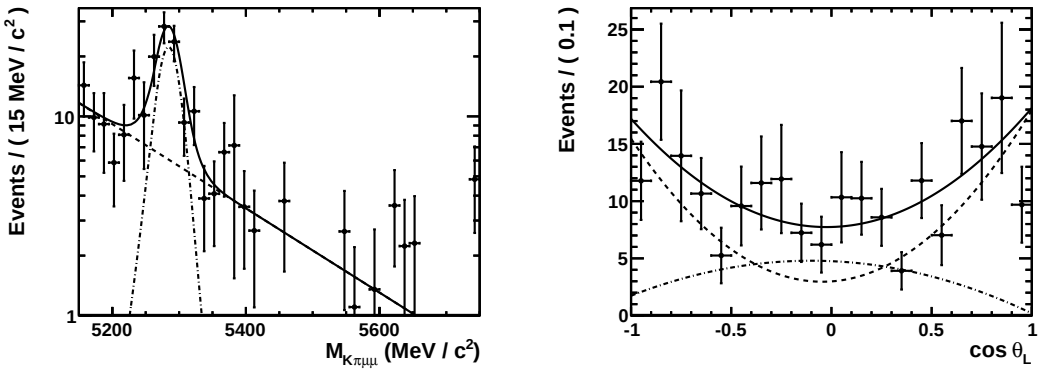


(d) ϕ for S_9



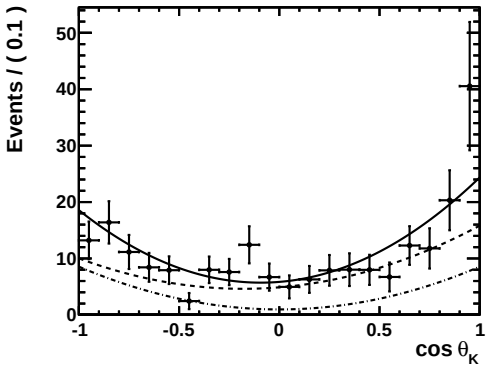
(e) ϕ for A_9

Figure 7.6: Distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$ in $\mathrm{m}_{K\pi\mu\mu}$ (a) and the decay angles $\cos\theta_L$ (b), $\cos\theta_K$ (c) and ϕ that corresponds to S_9 (d) and A_9 (e). The total fit pdf (solid black line), the $B^0 \to K^{*0}\mu^+\mu^-$ component (dot-dashed line) and the Class III background component (dashed line) are indicated.

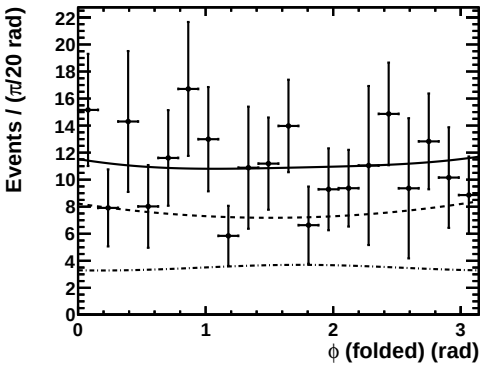


(a) $\mathrm{m}_{K\pi\mu\mu}$

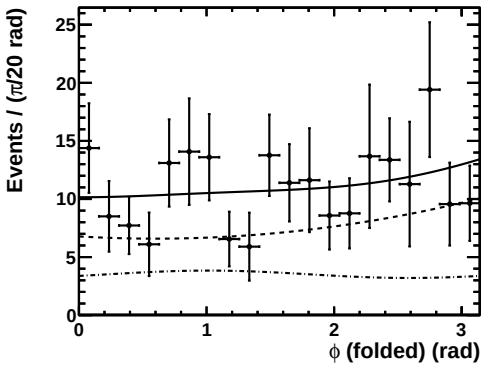
(b) $\cos\theta_L$



(c) $\cos\theta_K$

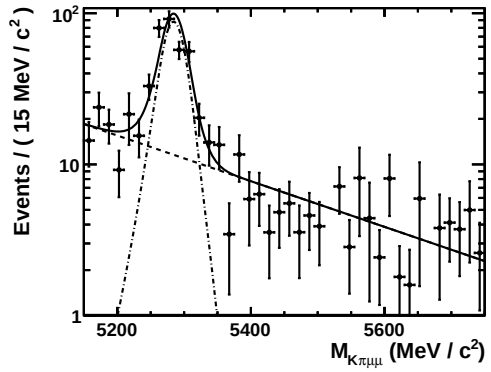


(d) ϕ for S_9

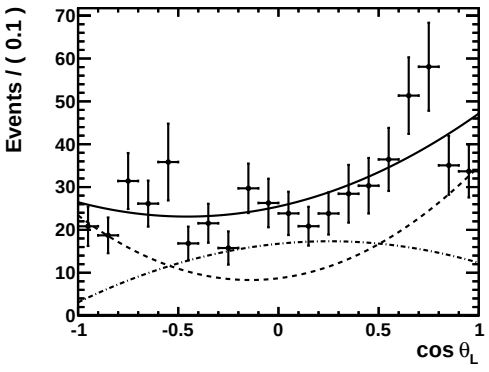


(e) ϕ for A_9

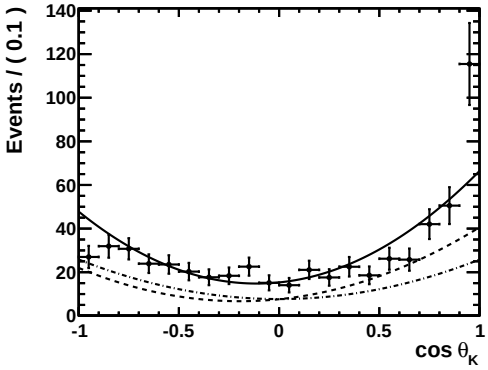
Figure 7.7: Distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $2.00 < q^2 < 4.30\,\mathrm{GeV}^2/c^4$ in $\mathrm{m}_{K\pi\mu\mu}$ (a) and the decay angles $\cos\theta_L$ (b), $\cos\theta_K$ (c) and ϕ that corresponds to S_9 (d) and A_9 (e). The total fit pdf (solid black line), the $B^0 \to K^{*0}\mu^+\mu^-$ component (dot-dashed line) and the Class III background component (dashed line) are indicated.



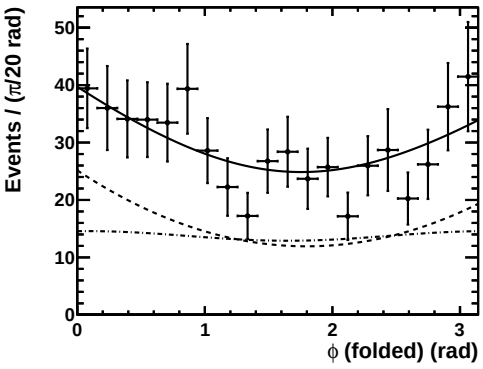
(a) $m_{K\pi\mu\mu}$



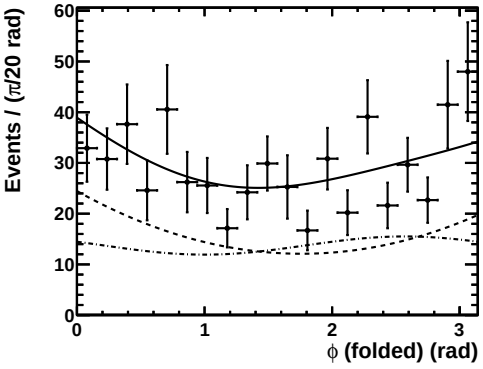
(b) $\cos\theta_L$



(c) $\cos\theta_K$

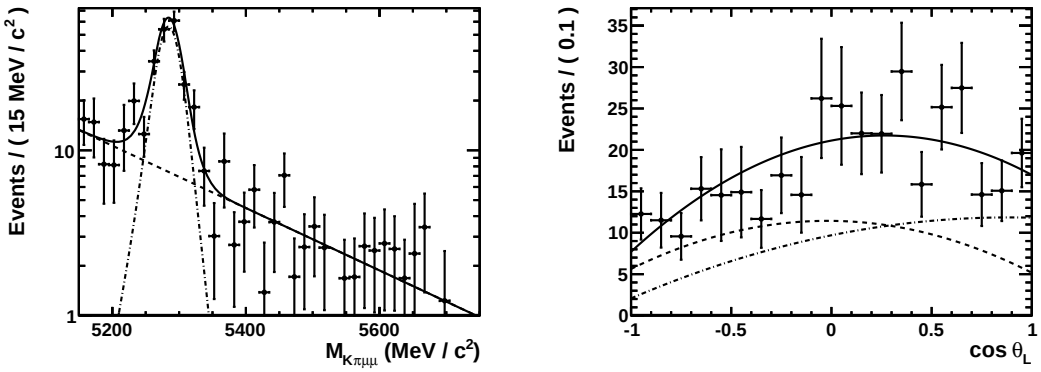


(d) ϕ for S_9



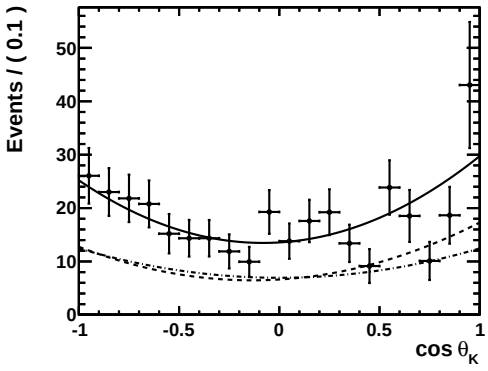
(e) ϕ for A_9

Figure 7.8: Distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $4.30 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$ in $m_{K\pi\mu\mu}$ (a) and the decay angles $\cos\theta_L$ (b), $\cos\theta_K$ (c) and ϕ that corresponds to S_9 (d) and A_9 (e). The total fit pdf (solid black line), the $B^0 \to K^{*0}\mu^+\mu^-$ component (dot-dashed line) and the Class III background component (dashed line) are indicated.

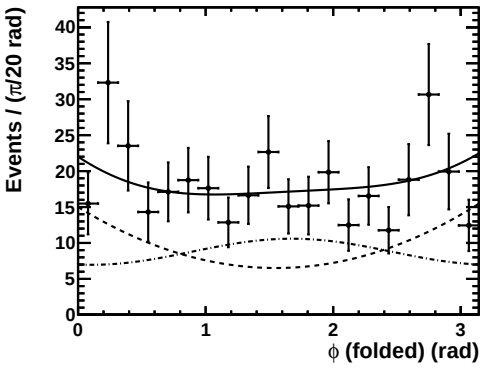


(a) $\mathrm{m}_{K\pi\mu\mu}$

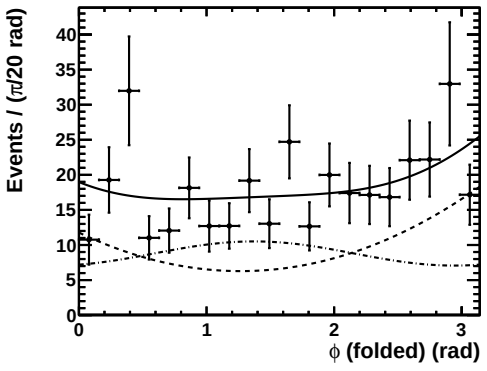
(b) $\cos\theta_L$



(c) $\cos\theta_K$

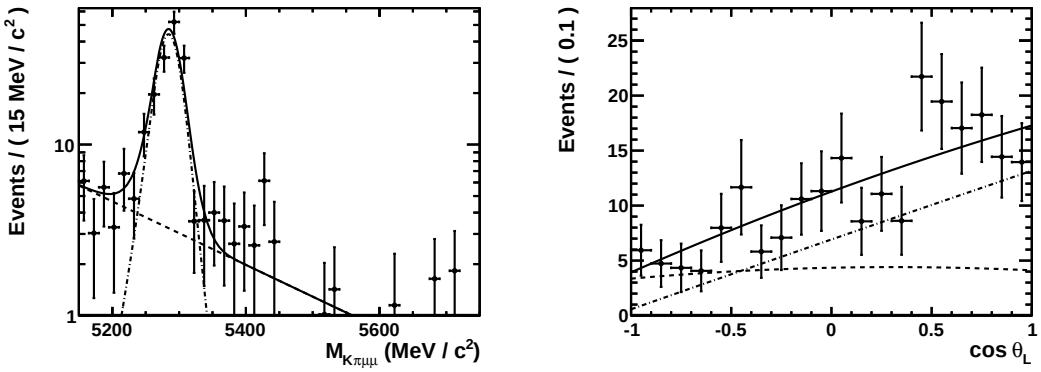


(d) ϕ for S_9



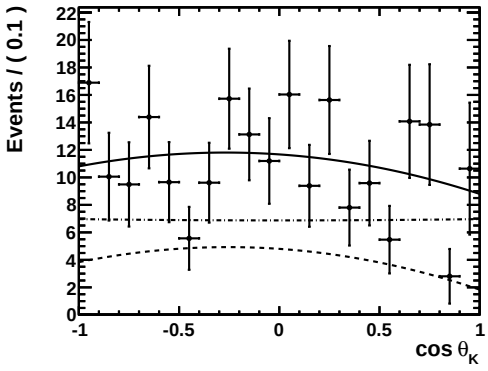
(e) ϕ for A_9

Figure 7.9: Distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$ in $\mathrm{m}_{K\pi\mu\mu}$ (a) and the decay angles $\cos\theta_L$ (b), $\cos\theta_K$ (c) and ϕ that corresponds to S_9 (d) and A_9 (e). The total fit pdf (solid black line), the $B^0 \to K^{*0}\mu^+\mu^-$ component (dot-dashed line) and the Class III background component (dashed line) are indicated.

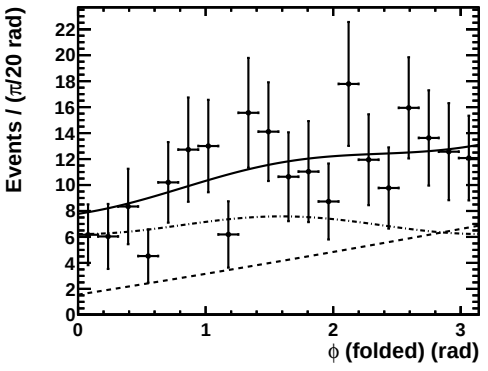


(a) $\mathrm{m}_{K\pi\mu\mu}$

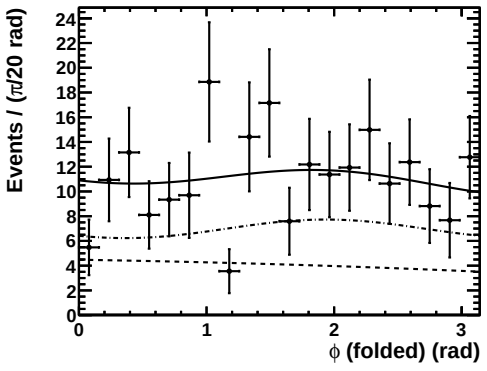
(b) $\cos\theta_L$



(c) $\cos\theta_K$

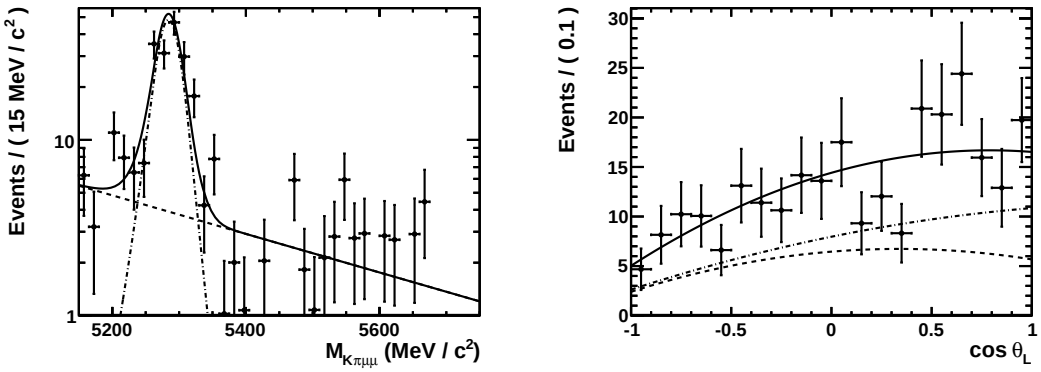


(d) ϕ for S_9



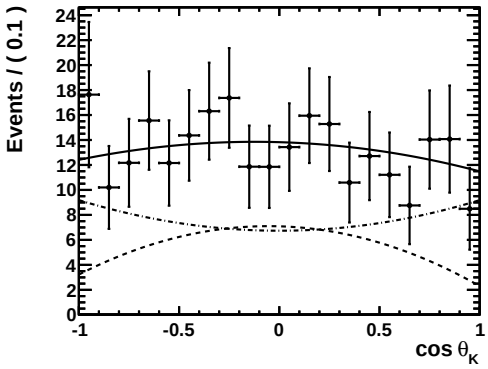
(e) ϕ for A_9

Figure 7.10: Distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $14.18 < q^2 < 16.00\,\mathrm{GeV}^2/c^4$ in $\mathrm{m}_{K\pi\mu\mu}$ (a) and the decay angles $\cos\theta_L$ (b), $\cos\theta_K$ (c) and ϕ that corresponds to S_9 (d) and A_9 (e). The total fit pdf (solid black line), the $B^0 \to K^{*0}\mu^+\mu^-$ component (dot-dashed line) and the Class III background component (dashed line) are indicated.

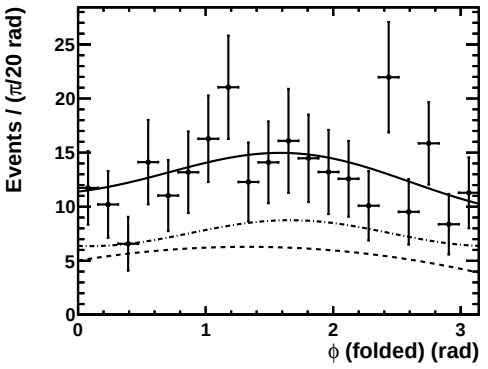


(a) $\mathrm{m}_{K\pi\mu\mu}$

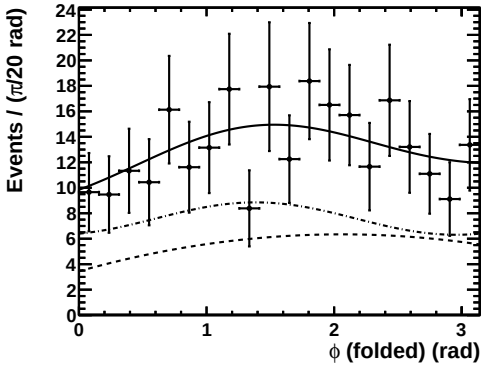
(b) $\cos\theta_L$



(c) $\cos\theta_K$

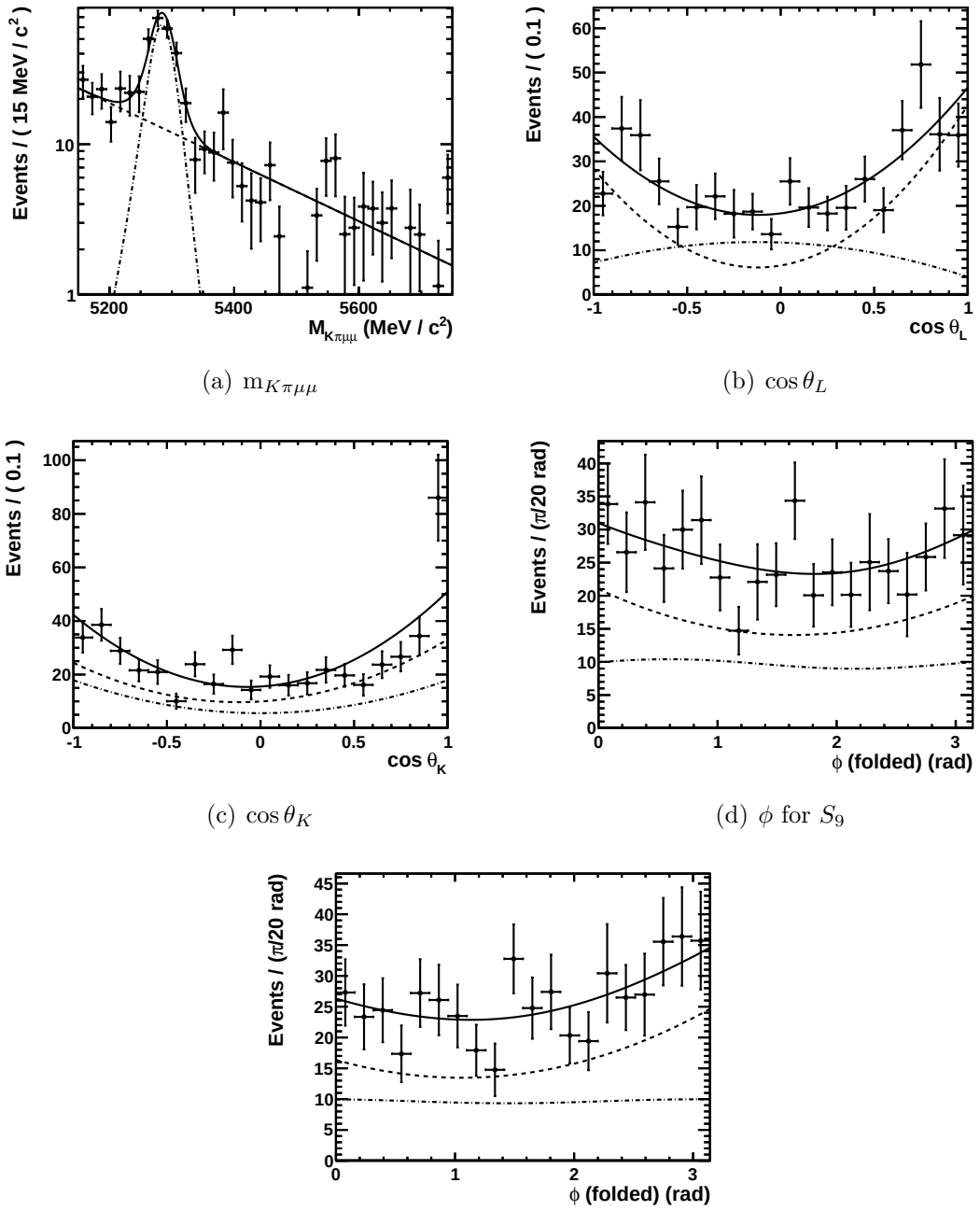


(d) ϕ for S_9



(e) ϕ for A_9

Figure 7.11: Distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $16.00 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ in $\mathrm{m}_{K\pi\mu\mu}$ (a) and the decay angles $\cos\theta_L$ (b), $\cos\theta_K$ (c) and ϕ that corresponds to S_9 (d) and A_9 (e). The total fit pdf (solid black line), the $B^0 \to K^{*0}\mu^+\mu^-$ component (dot-dashed line) and the Class III background component (dashed line) are indicated.



(a) $m_{K\pi\mu\mu}$

(b) $\cos\theta_L$

(c) $\cos\theta_K$

(d) ϕ for S_9

(e) ϕ for A_9

Figure 7.12: Distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data with $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$ in $m_{K\pi\mu\mu}$ (a) and the decay angles $\cos\theta_L$ (b), $\cos\theta_K$ (c) and ϕ that corresponds to S_9 (d) and A_9 (e). The total fit pdf (solid black line), the $B^0 \to K^{*0}\mu^+\mu^-$ component (dot-dashed line) and the Class III background component (dashed line) are indicated.

7.3.1 Statistical uncertainties on the angular observables

The FC technique is used to calculate the confidence interval on the measurement of each $B^0 \to K^{*0}\mu^+\mu^-$ angular observable. A simple correction, such as that performed in Sec. 6, cannot be used here as the observables are correlated. The confidence interval is obtained on an observable Λ by calculating the p-value of the data as a function the observable. The p-value of the data is calculated as a function of Λ by trialling different values of the observable. The confidence interval is bounded by the values Λ_{68}, which are the values of Λ at which the data rejects the hypothesis $\Lambda = \Lambda_{68}$ at 68% confidence level. For a given value of the observable Λ_i the p-value is computed using the ratio of log-likelihood

$$R_{\mathcal{L}} = \log\left[\frac{\mathcal{L}(\vec{x}|\Lambda_i, \hat{\vec{\lambda}})}{\mathcal{L}(\vec{x}|\hat{\Lambda}, \hat{\vec{\lambda}})}\right], \tag{7.3.1}$$

where $\vec{x}$ are the variables of the fit ($\mathrm{m}_{K\pi\mu\mu}$, $\cos\theta_L$, $\cos\theta_K$ and ϕ) and $\vec{\lambda}$ are the remaining (nuisance) parameters in the fit. Each log-likelihood in the ratio, $\mathcal{L}(\vec{x}|\Lambda_i, \hat{\vec{\lambda}})$ and $\mathcal{L}(\vec{x}|\hat{\Lambda}, \hat{\vec{\lambda}})$, is obtained independently from a fit to a given data set.

The ratio $R_{\mathcal{L}}^{\mathrm{data}}$ is calculated by fitting candidates in the data. Each value of Λ_i defines a fit model f_i. The maximum likelihood estimate of Λ, as obtained from the fit in which Λ is a free parameter, is denoted Λ_0^{d}. The value of $R_{\mathcal{L}}^{\mathrm{data}}$ is zero if $\Lambda_i = \Lambda_0^{\mathrm{d}}$, and is positive otherwise.

An ensemble of J simulated data sets are generated according to each fit model f_i following the procedure detailed in Sec. 7.2.1.1. Each data set, labelled j, contains the same number of events as there are $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data in the relevant analysis q^2 bin. The ratio $R_{\mathcal{L}}^{\mathrm{sim,j}}$ is calculated by performing the two independent fits, $\mathcal{L}(\vec{x}|\Lambda_i, \hat{\vec{\lambda}})$ and $\mathcal{L}(\vec{x}|\hat{\Lambda}, \hat{\vec{\lambda}})$, on each of the simulated data sets. The fraction of the J data sets with $R_{\mathcal{L}}^{\mathrm{sim,j}} > R_{\mathcal{L}}^{\mathrm{d}}$ is the p-value of the data given the hypothesis $\Lambda = \Lambda_i$.

By definition the p-value for the data is unity when $\Lambda_i = \Lambda_0^{\mathrm{d}}$. As Λ_i is moved away from Λ_0^{d} the p-value reduces, because the data becomes less consistent with the hypothesis $\Lambda = \Lambda_i$. When the distribution of the p-value crosses 0.32, that Λ_i value equals Λ_{68} (by definition) and therefore corresponds to the 68% C.L. (1σ) bound on Λ. A two-sided confidence interval is obtained when there are two crossings (two values of Λ_{68}). If there are less than two crossings, then the bound is set to the maximum or minimum physically valid value of Λ, as appropriate.

In the present analysis, two hundred values of Λ_i are trialled for each observable, with the exception of F_{L}, for which one hundred values are trialled due to the smaller range. This ensures that the uncertainty on the confidence intervals due to the number of trialled Λ_i is both small, and the same, for each observable. Eight hundred simulated data sets

q^2 range (GeV^2/c^4)		A_{FB}		F_{L}		S_3		S_9		A_9	
$0.10 < q^2 < 2.00$	FC	-0.22	0.01	0.18	0.37	-0.15	0.04	-0.12	0.06	0.03	0.21
$0.10 < q^2 < 2.00$	$\mathcal{L}_{\mathrm{p}}$	-0.20	-0.02	0.20	0.34	-0.16	0.06	-0.14	0.08	0.02	0.22
$2.00 < q^2 < 4.30$	FC	-0.20	-0.02	0.72	0.91	-0.12	0.03	-0.12	0.10	-0.01	0.15
$2.00 < q^2 < 4.30$	$\mathcal{L}_{\mathrm{p}}$	-0.20	-0.02	0.72	0.91	-0.13	0.05	-0.13	0.11	-0.03	0.16
$4.30 < q^2 < 8.68$	FC	0.13	0.22	0.58	0.68	0.00	0.12	-0.05	0.08	-0.17	-0.05
$4.30 < q^2 < 8.68$	$\mathcal{L}_{\mathrm{p}}$	0.13	0.22	0.58	0.68	-0.01	0.13	-0.05	0.08	-0.17	-0.05
$10.09 < q^2 < 12.86$	FC	0.23	0.34	0.41	0.55	-0.25	-0.10	-0.15	0.07	-0.02	0.19
$10.09 < q^2 < 12.86$	$\mathcal{L}_{\mathrm{p}}$	0.22	0.34	0.41	0.54	-0.24	-0.10	-0.15	0.07	-0.02	0.18
$14.18 < q^2 < 16.00$	FC	0.41	0.54	0.27	0.41	-0.19	0.01	-0.09	0.09	-0.17	0.03
$14.18 < q^2 < 16.00$	$\mathcal{L}_{\mathrm{p}}$	0.41	0.53	0.28	0.41	-0.18	0.01	-0.09	0.09	-0.16	0.03
$16.00 < q^2 < 19.00$	FC	0.21	0.35	0.34	0.48	-0.24	-0.06	-0.14	0.07	-0.03	0.17
$16.00 < q^2 < 19.00$	$\mathcal{L}_{\mathrm{p}}$	0.21	0.35	0.34	0.48	-0.25	-0.05	-0.13	0.08	-0.04	0.17
$1.00 < q^2 < 6.00$	FC	-0.14	-0.02	0.55	0.69	-0.04	0.10	-0.01	0.15	-0.09	0.08
$1.00 < q^2 < 6.00$	$\mathcal{L}_{\mathrm{p}}$	-0.14	-0.02	0.56	0.68	-0.05	0.11	-0.02	0.16	-0.09	0.08

Table 7.5: Confidence intervals on the measurements of the FFD angular observables which are obtained using the Feldman-Cousins technique (FC) and using the profile likelihood ($\mathcal{L}_{\mathrm{p}}$).

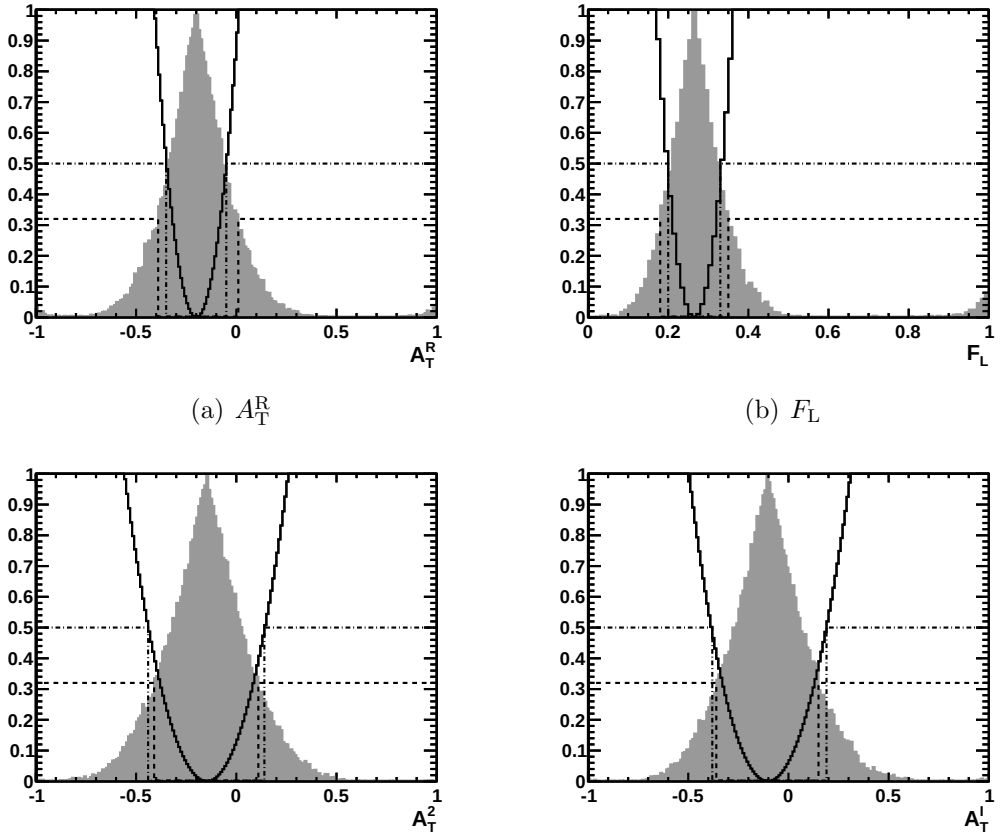
are generated at each Λ_i so that the statistical uncertainty on the determination of each p-value is small. The confidence intervals obtained using the FC technique, and those obtained using the profile likelihood $\mathcal{L}_{\mathrm{p}}$, are summarised in Tabs. 7.5 and 7.6.

The FC technique obtains larger confidence intervals for F_{L} and $A_{\mathrm{T}}^{\mathrm{R}}$, and smaller intervals for A_{T}^{2} and $A_{\mathrm{T}}^{\mathrm{I}}$, than those obtained using the profile likelihood. This is consistent with the expectation from Tab. 7.1, where the profile likelihood confidence intervals were too small in the former case and too large in the latter. The determination of the confidence intervals from the profile likelihood and the FC technique for the FFI observables in the $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$ q^2 bin are shown in Fig. 7.13.

7.3.2 Systematic uncertainties on the angular observables

Many of the systematic uncertainties associated with the measurement of the angular observables are in common with those of the differential branching fraction measurement detailed in Sec. 6.3.2. The additional sources of systematic uncertainty, and those that are treated differently for the angular fit, are described in Secs. 7.3.2.1 to 7.3.2.4.

If the maximum likelihood of the fit is outside of the physical and mathematical boundaries, the same fit results are obtained despite having applied a systematic variation. All of the systematic uncertainties of the angular fit are therefore extracted using



(a) $A_{\mathrm{T}}^{\mathrm{R}}$ (b) F_{L}

(c) A_{T}^{2} (d) $A_{\mathrm{T}}^{\mathrm{I}}$

Figure 7.13: Confidence intervals obtained for the FFI angular observables in the $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$ q^2 bin using the FC technique and using the profile likelihood. The profile likelihood (black solid histogram), the distribution of p-values produced by the FC technique (grey shaded histogram) are shown with the confidence intervals from the profile likelihood (black dot-dashed line) and from the FC technique (black dashed line) indicated.

q^2 range (GeV^2/c^4)		$A_{\mathrm{T}}^{\mathrm{R}}$		A_{T}^{2}		$A_{\mathrm{T}}^{\mathrm{I}}$	
$0.10 < q^2 < 2.00$	FC	-0.39	0.02	-0.41	0.12	-0.36	0.16
$0.10 < q^2 < 2.00$	$\mathcal{L}_{\mathrm{p}}$	-0.35	-0.04	-0.44	0.15	-0.38	0.20
$2.00 < q^2 < 4.30$	FC	-1.00	-0.18	-1.00	0.24	-1.00	0.75
$2.00 < q^2 < 4.30$	$\mathcal{L}_{\mathrm{p}}$	-1.00	-0.14	-1.00	0.46	-1.00	1.00
$4.30 < q^2 < 8.68$	FC	0.47	0.76	0.01	0.64	-0.27	0.39
$4.30 < q^2 < 8.68$	$\mathcal{L}_{\mathrm{p}}$	0.46	0.77	-0.02	0.67	-0.30	0.40
$10.09 < q^2 < 12.86$	FC	0.59	0.83	-1.00	-0.41	-0.54	0.22
$10.09 < q^2 < 12.86$	$\mathcal{L}_{\mathrm{p}}$	0.58	0.84	-0.90	-0.39	-0.56	0.23
$14.18 < q^2 < 16.00$	FC	0.84	1.00	-0.55	0.01	-0.29	0.30
$14.18 < q^2 < 16.00$	$\mathcal{L}_{\mathrm{p}}$	0.83	1.00	-0.56	0.03	-0.29	0.29
$16.00 < q^2 < 19.00$	FC	0.48	0.76	-0.80	-0.19	-0.44	0.27
$16.00 < q^2 < 19.00$	$\mathcal{L}_{\mathrm{p}}$	0.47	0.77	-0.84	-0.18	-0.44	0.26
$1.00 < q^2 < 6.00$	FC	-0.51	-0.08	-0.22	0.52	-0.04	0.76
$1.00 < q^2 < 6.00$	$\mathcal{L}_{\mathrm{p}}$	-0.51	-0.07	-0.27	0.56	-0.09	0.82

Table 7.6: Confidence intervals on the measurements of the FFI angular observables which are obtained using the Feldman-Cousins technique (FC) and using the profile likelihood ($\mathcal{L}_{\mathrm{p}}$).

simulated events as outlined below.

An ensemble of ten thousand simulated data sets is generated using the fit model following the procedure described in Sec. 7.2.1.1. Each data set in the ensemble contains the same number of events as there are candidates in the data. The value of each angular observable, the signal fraction f_{sig}, and the exponent p_0, are taken from the angular fits to $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data. Each data set in the ensemble is fitted to obtain a distribution of values for each angular observable. This process is repeated for each systematic variation. The difference in the mean of the nominal and fluctuated distributions is taken as the systematic uncertainty.

The standard error of each distribution is calculated to ensure the statistical fluctuations on the mean, due to the limited number of generated data sets, is small. The standard error is typically $\sim 1 \times 10^{-3}$.

7.3.2.1 Background angular fit model

Class III background candidates in some regions of the angular space receive large AC weights. The extent to which the measurement of the angular observables is affected by such candidates is dependant on the background angular model. To conservatively explore the impact of this uncertainty, the simulated data sets are fitted using 1$^{\mathrm{st}}$ and 3$^{\mathrm{rd}}$

order polynominal shapes, rather than the nominal 2^{nd} order polynominal.

7.3.2.2 Kaon misidentification

The statistical uncertainty on ω_{ID} is about $\sim$10%. Fluctuations of this quantity will modify the dilution of the CP-odd angular observables and change the fit results, and it is therefore considered as a source of systematic uncertainty. To explore the impact of a statistical fluctuation in the value of ω_{ID}, the simulated data sets are fitted having varied the level of ω_{ID} by $\pm$10%.

7.3.2.3 Class I and Class II backgrounds

Contributions from Class I and Class II background events are not incorporated into the angular fit model. The contribution of Class I and Class II background events is $\sim$6% of the signal yield in the $\text{m}_{K\pi\mu\mu}$ mass sidebands and $\sim$4% of the signal yield in the signal window. Since there is no evidence for the Class I and Class II contribution affecting the angular distribution in the $\text{m}_{K\pi\mu\mu}$ mass sidebands, as shown in Fig. 7.4, this contribution will not affect the fit and, therefore, is not considered to be a source of systematic uncertainty.

Since the signal window is about five times smaller than the mass sidebands, the $\sim$4% contribution of Class I and Class II background in the signal window constitutes a much larger fraction of the background in this region than the $\sim$6% contribution does in the $\text{m}_{K\pi\mu\mu}$ mass sidebands. As it is unknown whether this larger fraction of Class I and Class II backgrounds affects the angular distribution of the background, the contribution from Class I and Class II backgrounds in the signal $\text{m}_{K\pi\mu\mu}$ mass window is considered as a source of systematic uncertainty. To conservatively explore the impact of this uncertainty, an additional component is incorporated into the fit model to mimic the $\sim$4% contribution from Class I and Class II backgrounds. The additional component uses M_{B^0} to define a $\text{m}_{K\pi\mu\mu}$ mass distribution. The angular distribution of the Class I and Class II background events is modelled by the signal angular pdf with $F_{\text{L}} = 0.5$. The systematic uncertainty is explored by assuming large values for the remaining angular observables. The observables A_{FB} (A_{T}^{R}) is assumed to be $\pm$80%, and the remaining observables are assumed to be $\pm$20%, of their maximum value given $F_{\text{L}} = 0.5$. These values are chosen to produce a large effect on the angular measurement without resulting in the events being generated with observables that are outside of the mathematically allowed region.

7.3.2.4 Multiple candidates

Multiple candidates in a single event are considered as a source of systematic uncertainty as in Sec. 6.3.2.11. Such candidates have a different angular distribution to both the $B^0 \to K^{*0}\mu^+\mu^-$ and the Class III background components. This uncertainty cannot be explored separately, as the simulated data sets used to explore the systematic uncertainties are not passed through the simulation of the detector response and reconstruction algorithms. The level of multiple candidates in the data is $\sim 1\%$, and is dominated by the Class I background events as described in Sec. 4.3.5. This systematic uncertainty is therefore covered by the Class I and Class II systematic uncertainty described above.

7.3.2.5 Summary of the systematic uncertainties

The systematic uncertainties associated with the systematic variations described above, and for those described in Sec. 6.3.2, are summarised for $A_{\rm FB}$ in Tab. 7.7. The systematic uncertainties are summarised for the remaining observables in App. H. The largest systematic uncertainty is on the Class I and Class II backgrounds. Other large systematic uncertainties are on the S-wave and the AC weights. The remaining systematic uncertainties are relatively small. As in most cases multiple systematic variations are associated with a single source, the uncertainties are aggregated by selecting the largest uncertainty associated with each source. This aggregation is performed separately for each analysis q^2 bin and each observable. It is also performed separately for positive and negative systematic shifts so that asymmetric systematic uncertainties are obtained. This differs from the measurement of the differential branching fraction, as in this case asymmetric statistical uncertainties are obtained from the FC technique. The aggregated uncertainties are summed in quadrature to obtain the total asymmetric systematic uncertainty. The aggregated and total systematic uncertainties are summarised in Tabs. 7.8 and 7.9. The systematic uncertainties are $\lesssim 50\%$ of the statistical uncertainty in all of the angular observables and all of the analysis q^2 bins.

7.3.3 Testing the p-value of the Standard Model

The p-value of the data given the SM hypothesis is calculated in each analysis q^2 bin using a DT_{ptp} with $\bar{x} = \{\cos\theta_L,\ \cos\theta_K,\ \phi\}$ and $\bar{\sigma} = 0.05$. The results are shown in Tab. 7.10. The $10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$ analysis q^2 bin is omitted due to the lack of a SM theoretical prediction. The p-values are in the range 0.08 to 0.79, indicating that the data is consistent with the SM hypothesis.

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	-0.109	-0.108	0.171	0.273	0.454	0.276	-0.088
Standard Error	0.002	0.002	0.001	0.001	0.001	0.001	0.001
AC $\cos\theta_L$ Up	-0.003	-0.002	0.004	0.004	0.003	0.004	-0.000
AC $\cos\theta_L$ Down	-0.001	0.002	-0.004	-0.007	-0.006	-0.006	0.004
AC $\cos\theta_K$ Up	-0.001	0.004	-0.006	-0.008	-0.012	-0.005	0.005
AC $\cos\theta_K$ Down	-0.006	-0.003	0.007	0.007	0.010	0.007	0.001
AC $\cos\theta_L$ Up $\cos\theta_K$ Up	-0.003	-0.002	-0.003	-0.005	-0.005	-0.005	0.004
AC $\cos\theta_L$ Up $\cos\theta_K$ Down	-0.005	-0.005	0.010	0.010	0.015	0.010	-0.002
AC $\cos\theta_L$ Down $\cos\theta_K$ Up	-0.004	0.003	-0.010	-0.012	-0.018	-0.010	0.007
AC $\cos\theta_L$ Down $\cos\theta_K$ Down	-0.002	-0.001	0.002	0.001	0.002	0.003	0.002
AC Non-factorisable Up	-0.004	-0.003	-0.000	-0.000	-0.002	0.001	0.003
AC Non-factorisable Down	-0.002	-0.001	-0.001	0.001	-0.001	-0.000	0.003
AC q^2 binning +1	-0.002	-0.001	0.004	-0.001	-0.002	-0.002	0.003
AC q^2 binning -1	-0.007	0.005	-0.003	0.003	0.005	0.004	0.005
Trigger efficiency Up	-0.006	0.000	0.000	0.001	-0.001	-0.001	0.002
Trigger efficiency Down	-0.000	-0.003	-0.000	-0.001	-0.003	-0.000	0.002
DLL binnning +1	-0.003	0.001	-0.001	0.001	0.000	-0.001	0.002
DLL binning -1	-0.004	-0.002	0.000	0.002	0.001	-0.000	0.003
IsMuon efficiency Up	-0.002	0.000	-0.001	-0.001	-0.003	-0.001	0.005
IsMuon efficiency Down	-0.003	-0.004	0.001	-0.002	-0.001	-0.001	0.001
Tracking efficiency Up	-0.002	-0.002	0.000	0.002	-0.001	0.001	0.002
Tracking efficiency Down	-0.006	-0.003	0.000	-0.001	-0.001	0.001	0.002
B^0 p_T re-weighting	-0.000	-0.003	0.000	0.000	-0.001	0.004	0.005
No B^0 p re-weighting	-0.004	-0.000	0.002	0.001	-0.000	0.002	0.002
IP Correction	0.002	0.002	-0.001	0.001	0.005	0.001	-0.002
Mass resolution Up	-0.001	0.000	-0.001	-0.000	-0.002	-0.000	0.003
Mass resolution Down	-0.002	-0.002	-0.001	0.001	-0.000	0.001	0.003
S-wave component	0.005	0.005	-0.007	-0.011	-0.018	-0.015	0.005
Order 1 background model	-0.005	-0.001	0.001	0.000	-0.001	0.001	0.003
Order 3 background model	-0.005	0.002	-0.000	-0.000	-0.002	-0.001	0.002
Kaon misidentification Up	-0.005	-0.001	0.001	0.000	0.001	0.002	0.002
Kaon misidentification Down	0.003	0.001	-0.002	-0.001	-0.004	-0.003	0.004
Class II and III Up	0.015	0.016	0.005	0.001	-0.008	0.001	0.018
Class II and III Down	-0.012	-0.005	-0.020	-0.021	-0.030	-0.024	-0.008

Table 7.7: Impact on the measured value of A_{FB} associated with each systematic variation.

q^2 range (GeV^2/c^4)	A_{FB}	F_{L}	S_3	S_9	A_9	$A_{\mathrm{T}}^{\mathrm{R}}$	A_{T}^2	$A_{\mathrm{T}}^{\mathrm{I}}$
$0.10 < q^2 < 2.00$	0.02	0.04	0.02	0.01	0.01	0.03	0.02	0.03
$2.00 < q^2 < 4.30$	0.02	0.02	0.01	0.01	0.00	0.06	0.05	0.04
$4.30 < q^2 < 8.68$	0.01	0.03	0.01	0.00	0.01	0.01	0.02	0.03
$10.09 < q^2 < 12.86$	0.01	0.03	0.02	0.01	0.00	0.01	0.05	0.03
$14.18 < q^2 < 16.00$	0.02	0.03	0.01	0.00	0.01	0.01	0.05	0.01
$16.00 < q^2 < 19.00$	0.01	0.03	0.02	0.01	0.01	0.01	0.06	0.03
$1.00 < q^2 < 6.00$	0.02	0.03	0.01	0.00	0.00	0.04	0.03	0.01

Table 7.8: Total positive systematic uncertainties on the measurement of each angular observable in each analysis q^2 bin.

q^2 range (GeV^2/c^4)	A_{FB}	F_{L}	S_3	S_9	A_9	$A_{\mathrm{T}}^{\mathrm{R}}$	A_{T}^2	$A_{\mathrm{T}}^{\mathrm{I}}$
$0.10 < q^2 < 2.00$	0.02	0.03	0.00	0.00	0.01	0.01	0.01	0.00
$2.00 < q^2 < 4.30$	0.01	0.03	0.01	0.00	0.01	0.04	0.03	0.04
$4.30 < q^2 < 8.68$	0.02	0.03	0.01	0.01	0.01	0.08	0.04	0.01
$10.09 < q^2 < 12.86$	0.03	0.02	0.01	0.00	0.01	0.05	0.02	0.01
$14.18 < q^2 < 16.00$	0.04	0.03	0.01	0.01	0.00	0.04	0.01	0.01
$16.00 < q^2 < 19.00$	0.03	0.03	0.01	0.01	0.00	0.05	0.04	0.01
$1.00 < q^2 < 6.00$	0.01	0.03	0.00	0.01	0.01	0.01	0.05	0.07

Table 7.9: Total negative systematic uncertainties on the measurement of each angular observable in each analysis q^2 bin.

q^2 range (GeV^2/c^4)	p-value
$0.10 < q^2 < 2.00$	0.17
$2.00 < q^2 < 4.30$	0.79
$4.30 < q^2 < 8.68$	0.49
$14.18 < q^2 < 16.00$	0.20
$16.00 < q^2 < 19.00$	0.47
$1.00 < q^2 < 6.00$	0.08

Table 7.10: The p-value of the data given the SM hypothesis in each analysis q^2 bin.

q^2 range (GeV^2/c^4)	A_{FB}	F_{L}	S_3	S_9	A_9
$0.10 < q^2 < 2.00$	$-0.11^{+0.12+0.02}_{-0.11-0.02}$	$0.26^{+0.11+0.04}_{-0.08-0.03}$	$-0.05^{+0.09+0.02}_{-0.10-0.00}$	$-0.04^{+0.10+0.01}_{-0.08-0.00}$	$0.12^{+0.09+0.01}_{-0.09-0.01}$
$2.00 < q^2 < 4.30$	$-0.11^{+0.09+0.02}_{-0.09-0.01}$	$0.82^{+0.09+0.02}_{-0.10-0.03}$	$-0.06^{+0.09+0.01}_{-0.06-0.01}$	$-0.02^{+0.12+0.01}_{-0.10-0.00}$	$0.08^{+0.07+0.00}_{-0.09-0.01}$
$4.30 < q^2 < 8.68$	$0.17^{+0.05+0.01}_{-0.04-0.02}$	$0.63^{+0.05+0.03}_{-0.05-0.03}$	$0.06^{+0.06+0.01}_{-0.06-0.01}$	$0.01^{+0.07+0.00}_{-0.06-0.01}$	$-0.12^{+0.07+0.01}_{-0.05-0.01}$
$10.09 < q^2 < 12.86$	$0.28^{+0.06+0.01}_{-0.05-0.03}$	$0.47^{+0.08+0.03}_{-0.06-0.02}$	$-0.20^{+0.10+0.02}_{-0.05-0.01}$	$-0.04^{+0.11+0.01}_{-0.11-0.00}$	$0.08^{+0.11+0.00}_{-0.10-0.01}$
$14.18 < q^2 < 16.00$	$0.47^{+0.07+0.02}_{-0.06-0.04}$	$0.34^{+0.07+0.03}_{-0.07-0.03}$	$-0.10^{+0.11+0.01}_{-0.09-0.01}$	$-0.00^{+0.09+0.00}_{-0.09-0.01}$	$-0.07^{+0.10+0.01}_{-0.10-0.00}$
$16.00 < q^2 < 19.00$	$0.28^{+0.07+0.01}_{-0.07-0.03}$	$0.41^{+0.07+0.03}_{-0.07-0.03}$	$-0.16^{+0.10+0.02}_{-0.08-0.01}$	$-0.03^{+0.10+0.01}_{-0.11-0.01}$	$0.06^{+0.11+0.01}_{-0.09-0.00}$
$1.00 < q^2 < 6.00$	$-0.08^{+0.06+0.02}_{-0.06-0.01}$	$0.62^{+0.07+0.03}_{-0.07-0.03}$	$0.03^{+0.07+0.01}_{-0.07-0.00}$	$0.07^{+0.08+0.00}_{-0.08-0.01}$	$-0.01^{+0.09+0.00}_{-0.08-0.01}$

Table 7.11: Results of the measurement of the FFD angular observables, where the first uncertainty is statistical and the second is systematic, in each analysis q^2 bin.

q^2 range (GeV^2/c^4)	$A_{\mathrm{T}}^{\mathrm{R}}$	A_{T}^{2}	$A_{\mathrm{T}}^{\mathrm{I}}$
$0.10 < q^2 < 2.00$	$-0.20^{+0.22+0.03}_{-0.19-0.01}$	$-0.15^{+0.27+0.02}_{-0.26-0.01}$	$-0.10^{+0.26+0.03}_{-0.26-0.00}$
$2.00 < q^2 < 4.30$	$-0.82^{+0.64+0.06}_{-0.18-0.04}$	$-0.61^{+0.85+0.05}_{-0.39-0.03}$	$-0.22^{+0.97+0.04}_{-0.78-0.04}$
$4.30 < q^2 < 8.68$	$0.61^{+0.15+0.01}_{-0.14-0.08}$	$0.32^{+0.32+0.02}_{-0.31-0.04}$	$0.05^{+0.34+0.03}_{-0.32-0.01}$
$10.09 < q^2 < 12.86$	$0.72^{+0.11+0.01}_{-0.13-0.05}$	$-0.77^{+0.36+0.05}_{-0.23-0.02}$	$-0.17^{+0.39+0.03}_{-0.37-0.01}$
$14.18 < q^2 < 16.00$	$0.94^{+0.06+0.01}_{-0.10-0.04}$	$-0.30^{+0.31+0.05}_{-0.25-0.01}$	$-0.01^{+0.31+0.01}_{-0.28-0.01}$
$16.00 < q^2 < 19.00$	$0.62^{+0.14+0.01}_{-0.14-0.05}$	$-0.52^{+0.33+0.06}_{-0.28-0.04}$	$-0.10^{+0.37+0.03}_{-0.34-0.01}$
$1.00 < q^2 < 6.00$	$-0.29^{+0.21+0.04}_{-0.22-0.01}$	$0.15^{+0.37+0.03}_{-0.37-0.05}$	$0.36^{+0.40+0.01}_{-0.40-0.07}$

Table 7.12: Results of the measurement of the FFI angular observables, where the first uncertainty is statistical and the second is systematic, in each analysis q^2 bin.

7.3.4 Summary of the angular fit results

The results of the angular fit including the asymmetric statistical and systematic uncertainties are summarised for the FFD and FFI observables in Tabs. 7.11 and 7.12 respectively. These results, the SM prediction, and the SM prediction binned in the analysis q^2 bins, are plotted in Figs. 7.14 to 7.21. No theoretical prediction is shown between $8.68 < q^2 < 10.09\,\mathrm{GeV}^2/c^4$ since the $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to \psi(2S) K^{*0}$ decays dominate the decay rate in this region of q^2. Similarly, no measurement is performed between $8.68 < q^2 < 10.09\,\mathrm{GeV}^2/c^4$ and $12.86 < q^2 < 14.18\,\mathrm{GeV}^2/c^4$ as these regions of q^2 are removed to reject contributions of the $B^0 \to J/\psi\, K^{*0}$ and $B^0 \to \psi(2S) K^{*0}$ decays. No prediction is available for the $A_{\mathrm{T}}^{\mathrm{I}}$ observable for all values of q^2. No prediction is available for the S_9 and A_9 observables, however these observables are small ($\sim 10^{-3}$) in the SM for all values of q^2. As demonstrated by the p-values obtained in the previous section, the results are consistent with the SM predictions.

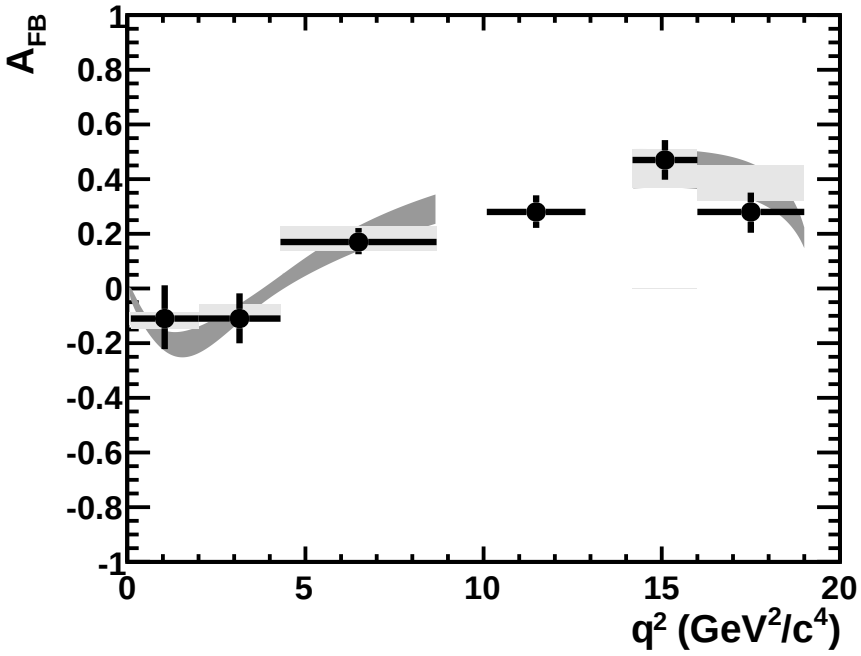


Figure 7.14: Results of the measurement of $A_{\rm FB}$ including statistical and systematic uncertainties (black points). The SM prediction curve (dark grey) and the SM prediction binned in the analysis q^2 bins (light grey) are indicated.

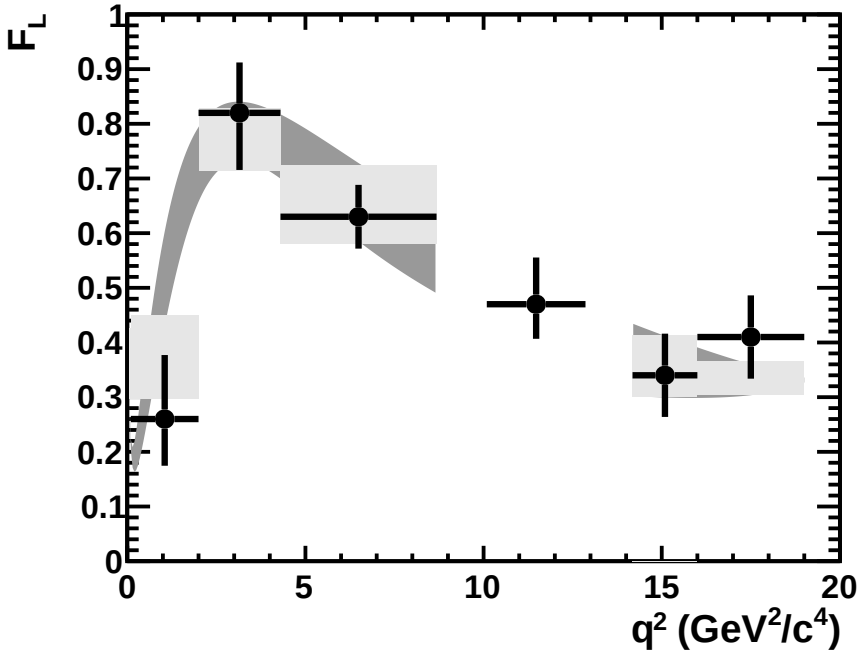


Figure 7.15: Results of the measurement of $F_{\rm L}$ including statistical and systematic uncertainties (black points). The SM prediction curve (dark grey) and the SM prediction binned in the analysis q^2 bins (light grey) are indicated.

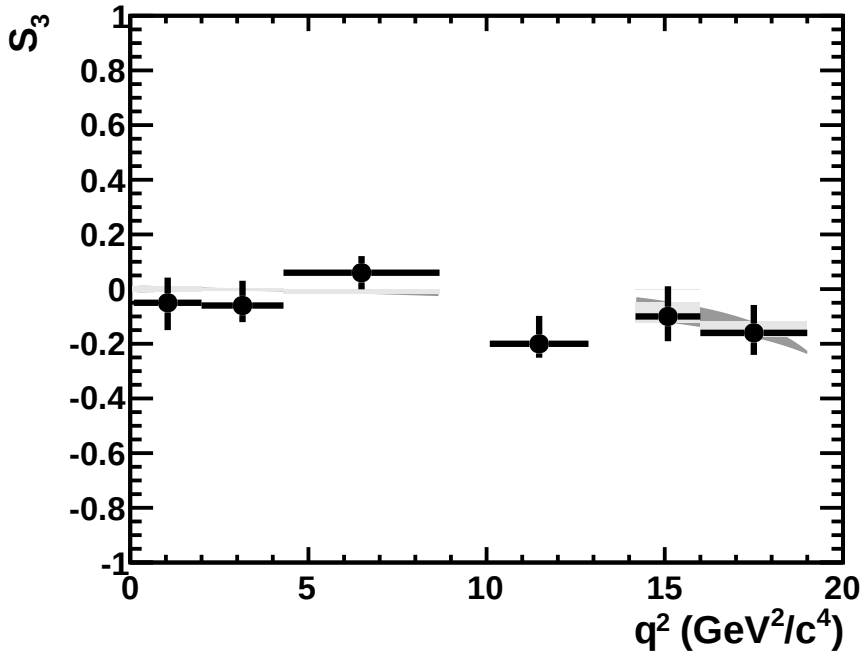


Figure 7.16: Results of the measurement of S_3 including statistical and systematic uncertainties (black points). The SM prediction curve (dark grey) and the SM prediction binned in the analysis q^2 bins (light grey) are indicated.

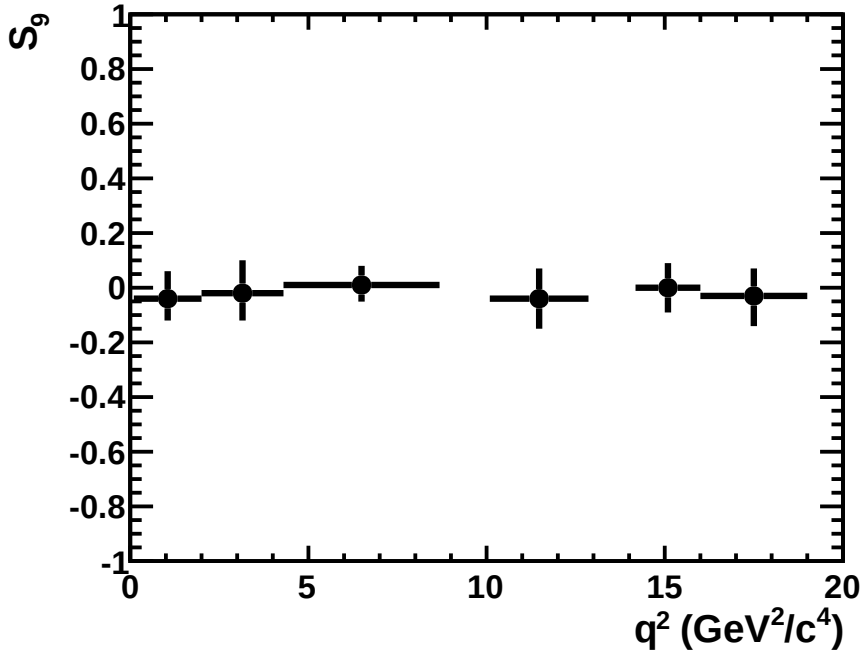


Figure 7.17: Results of the measurement of S_9 including statistical and systematic uncertainties (black points). The SM prediction is approximately zero (not drawn).

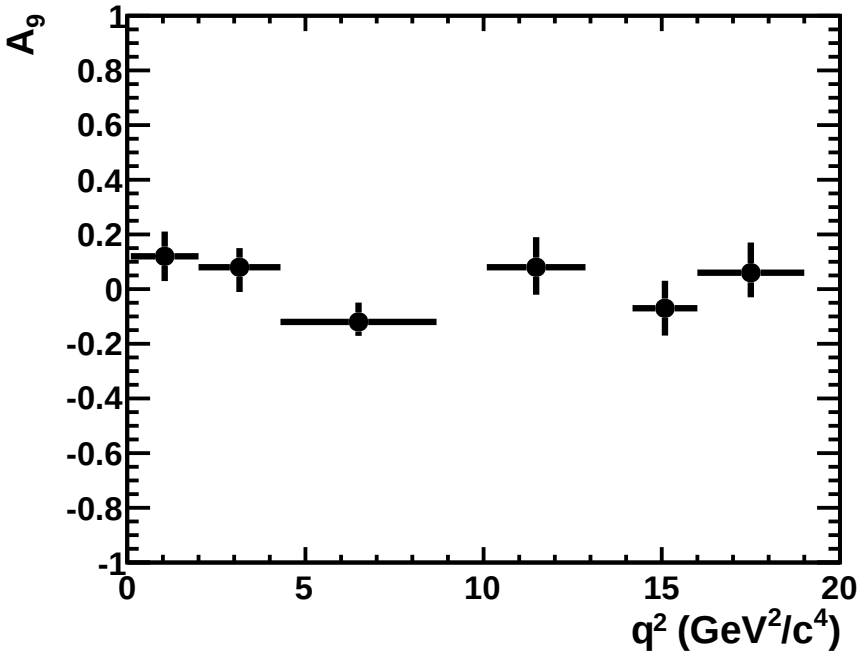


Figure 7.18: Results of the measurement of A_9 including statistical and systematic uncertainties (black points). The SM prediction is approximately zero (not drawn).

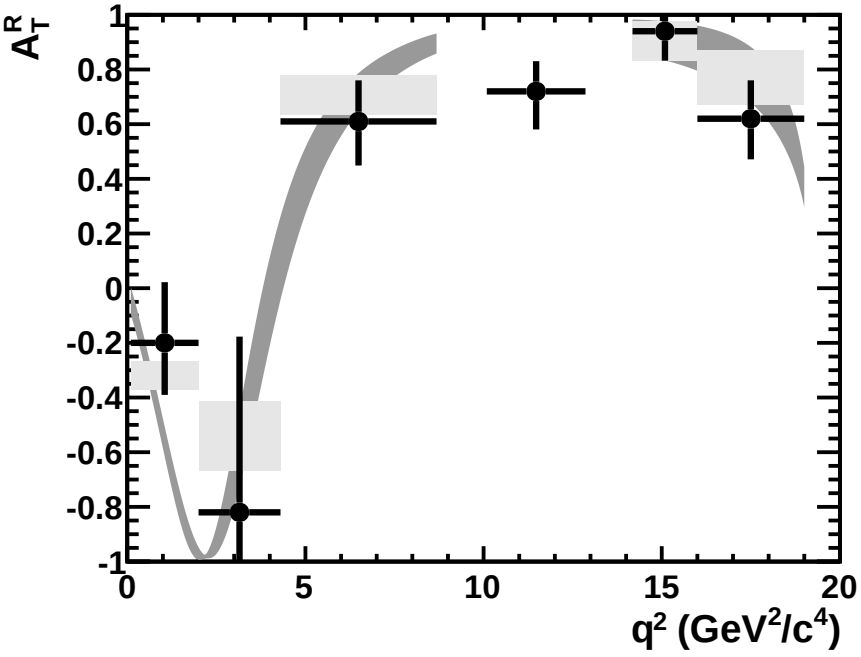


Figure 7.19: Results of the measurement of $A_{\mathrm{T}}^{\mathrm{R}}$ including statistical and systematic uncertainties (black points).The SM prediction curve (dark grey) and the SM prediction binned in the analysis q^2 bins (light grey) are indicated.

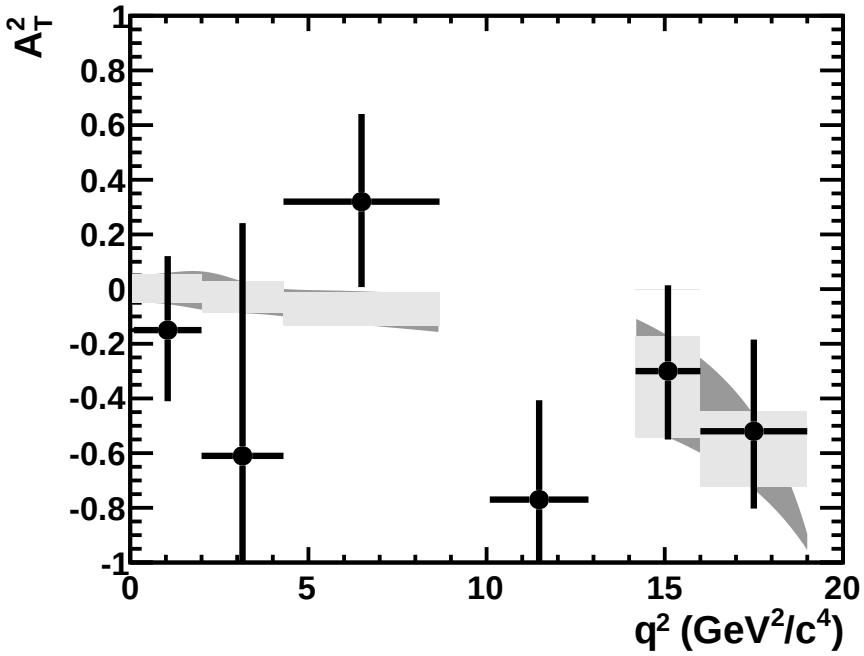


Figure 7.20: Results of the measurement of A_{T}^2 including statistical and systematic uncertainties (black points). The SM prediction curve (dark grey) and the SM prediction binned in the analysis q^2 bins (light grey) are indicated.

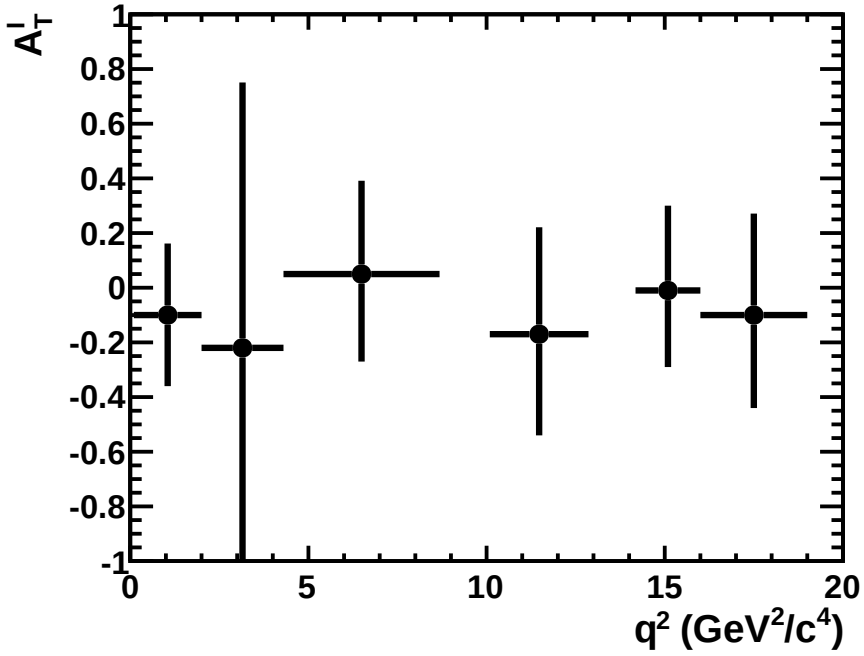


Figure 7.21: Results of the measurement of $A_{\mathrm{T}}^{\mathrm{I}}$ including statistical and systematic uncertainties (black points). The SM prediction is not known.

Chapter 8

Comparison of measurements and the theoretical implications

To put the results of the present analysis into a wider context, the most recent measurements of the $B^0 \to K^{*0}\mu^+\mu^-$ observables are summarised in Sec. 8.1. The theoretical implications of the measurements are then discussed in Sec. 8.2.

8.1 Comparison of the measurements with those of other collaborations

The analysis presented in this thesis is based on the analysis of the LHCb collaboration published in Ref. [89]. The two analyses use the same $1\,\text{fb}^{-1}$ of integrated luminosity. The author produced the selection for the LHCb collaborations published analysis, as well as developing and performing the angular fit. There is no significant difference in the angular fit between the LHCb analysis and that presented in this thesis. The selection was re-optimised for this thesis to retain more particles with low p_{T} and to use a considerably simpler multivariate classifier. Consequently, in the present analysis the acceptance effects are smaller, and the analysis more robust, than in the $1\,\text{fb}^{-1}$ analysis published in Ref. [89]. The present analysis measures each $B^0 \to K^{*0}\mu^+\mu^-$ observable with near identical precision to the analysis of Ref. [89]. The measurements of the observables A_{FB} and F_{L} in the two analyses are considerably more precise than, and consistent with, the previously published results of the LHCb collaboration presented in Ref. [42]

Each $B^0 \to K^{*0}\mu^+\mu^-$ observable that is measured by at least two collaborations is indicated in Tab. 8.1. The relevant collaborations, and the region of q^2 in which each collaboration measures the observables, are also indicated. The measurements of $B^0 \to K^{*0}\mu^+\mu^-$ observables in the present analysis are compared to previous measurements in Figs. 8.1

Collaboration	Observables					q^2 range (GeV^2/c^4)	Ref.
	$d\mathcal{B}/dq^2$	A_{FB}	F_{L}	A_{T}^2	A_9		
Present analysis	✓	✓	✓	✓	✓	$0.10 < q^2 < 19.00$	
LHCb	✓	✓	✓	✓	✓	$0.10 < q^2 < 19.00$	[89]
CDF	✓	✓	✓	✓	✓	$0.00 < q^2 < 19.30$	[40, 43]
CMS	✓	✓	✓			$1.00 < q^2 < 19.00$	[90]
Belle	✓	✓	✓			$0.00 < q^2 < (m_B - m_{K^{*0}})^2$	[36]
BaBar	✓	✓	✓			$0.10 < q^2 < (m_B - m_{K^{*0}})^2$	[37, 91]
ATLAS		✓	✓			$2.00 < q^2 < 19.00$	[92]

Table 8.1: Experimentally measured $B^0 \to K^{*0}\mu^+\mu^-$ observables at the present time, and the corresponding collaborations.

to 8.4. The markers are displaced from the centre of the q^2 bins so that the different measurements can be distinguished.

The measurements of $d\mathcal{B}/dq^2$, A_{FB} and F_{L} made with the LHCb data are more precise than all other previous measurements. The A_{T}^2 and A_9 observables are measured with considerably better precision than the only previous measurement, which was made by the CDF collaboration. All of the measurements are consistent with each other and with the prediction of the SM. The implication of the consistency of the measurements and the SM predictions is discussed in the next section.

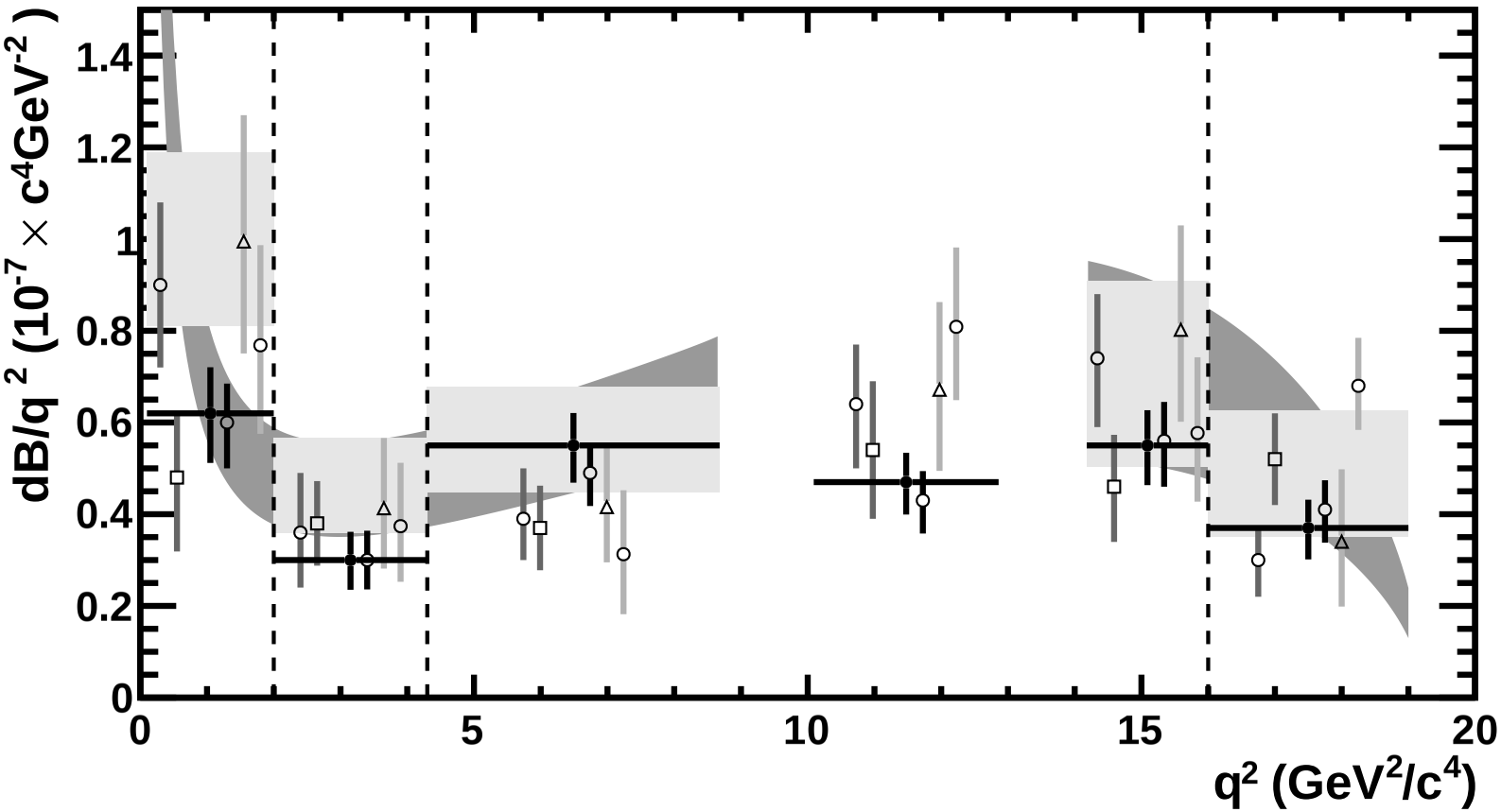


Figure 8.1: Comparison of measurements of the $d\mathcal{B}/dq^2$ by: the CDF (dark grey open circles) and CMS (dark grey open squares) collaborations; the present analysis (black circles) and the LHCb collaboration (black open circles); and the BaBar (light grey open triangles) and Belle (light grey circles) collaborations. The SM prediction curve (dark grey band) and the SM prediction binned in the analysis q^2 bins (light grey area) are indicated. In each q^2 bin the markers are displaced q^2 for clarity, and dashed lines are overlaid to distinguish between adjacent analysis q^2 bins.

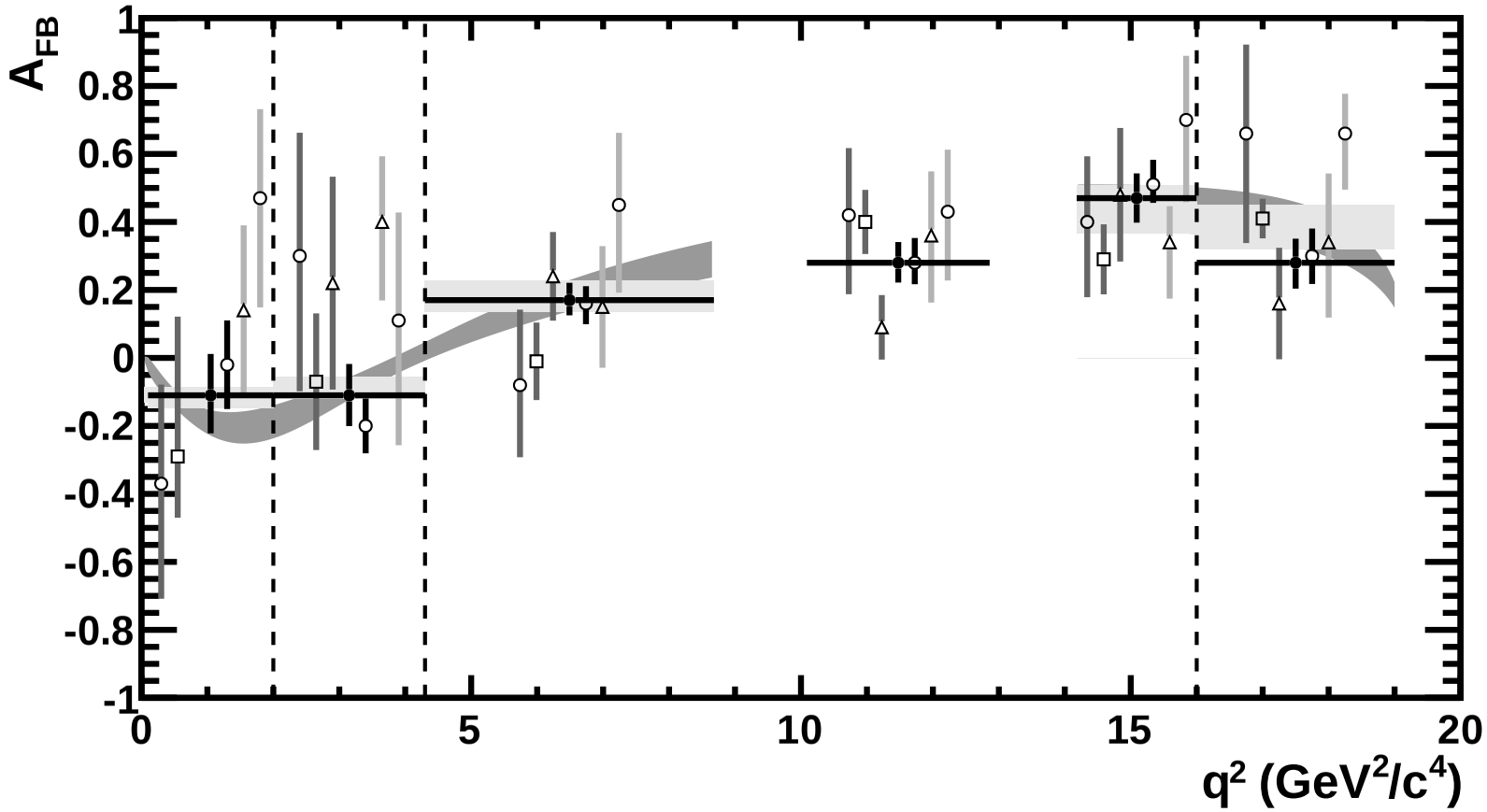


Figure 8.2: Comparison of measurements of $A_{\rm FB}$ by: the CDF (dark grey open circles), CMS (dark grey open squares) and ATLAS (dark grey triangles) collaborations; the present analysis (black circles) and the LHCb collaboration (black open circles); and the BaBar (light grey open triangles) and Belle (light grey circles) collaborations. The SM prediction curve (dark grey band) and the SM prediction binned in the analysis q^2 bins (light grey area) are indicated. In each q^2 bin the markers are displaced in q^2 for clarity, and dashed lines are overlaid to distinguish between adjacent analysis q^2 bins. There is no measurement by the ATLAS collaboration in the $0.10 < q^2 < 2.00\,{\rm GeV}^2/c^4$ analysis q^2 bin.

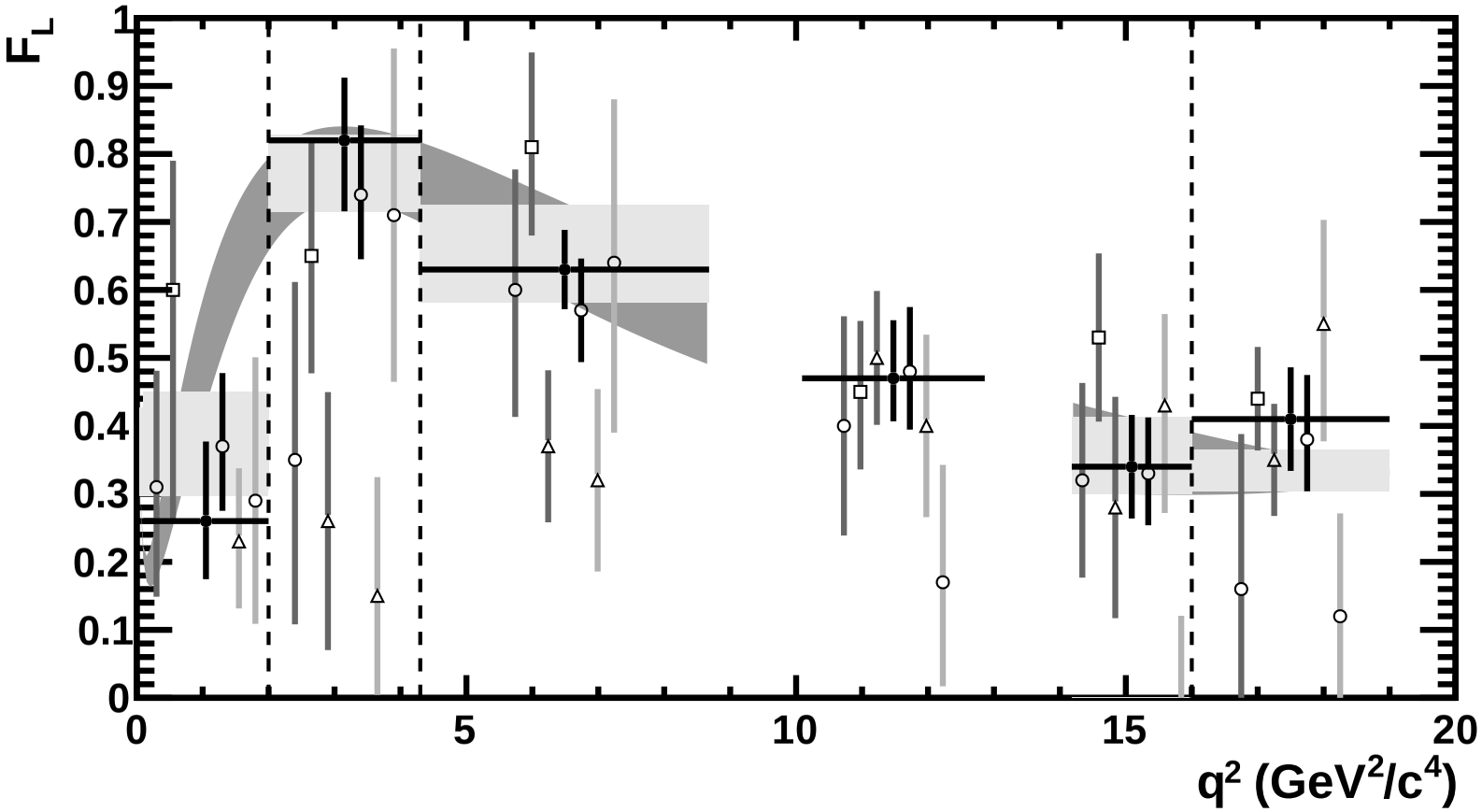


Figure 8.3: Comparison of measurements of $F_{\rm L}$ by: the CDF (dark grey open circles), CMS (dark grey open squares) and ATLAS (dark grey triangles) collaborations; the present analysis (black circles) and the LHCb collaboration (black open circles); and the BaBar (light grey open triangles) and Belle (light grey circles) collaborations. The SM prediction curve (dark grey band) and the SM prediction binned in the analysis q^2 bins (light grey area) are indicated. In each q^2 bin the markers are displaced in q^2 for clarity, and dashed lines are overlaid to distinguish between adjacent analysis q^2 bins. There is no measurement by the ATLAS collaboration in the $0.10 < q^2 < 2.00\,{\rm GeV}^2/c^4$ analysis q^2 bin.

Obs.	low q^2	high q^2
$\mathcal{B}$	$\mathcal{C}_{7,9,10}$, $\mathcal{C}'_{7,9,10}$	$\mathcal{C}_{9,10}$, $\mathcal{C}'_{9,10}$
$F_{\rm L}$	$\mathcal{C}_{7,9}$, $\mathcal{C}'_{7,9,10}$	$\mathcal{C}'_{9,10}$
S_3	$\mathcal{C}'_{7,10}$	$\mathcal{C}'_{9,10}$
$A_{\rm FB}$	$\mathcal{C}_{7,9}$	$\mathcal{C}_{9,10}$, $\mathcal{C}'_{9,10}$
A_9	$\mathcal{C}'_{7,9}$	$\mathcal{C}'_{9,10}$

Table 8.2: The $B^0 \to K^{*0}\mu^+\mu^-$ observables and the Wilson coefficients to which each is sensitive in the low- and high-q^2 regions. Adapted from Ref. [38].

8.2 Theoretical implications

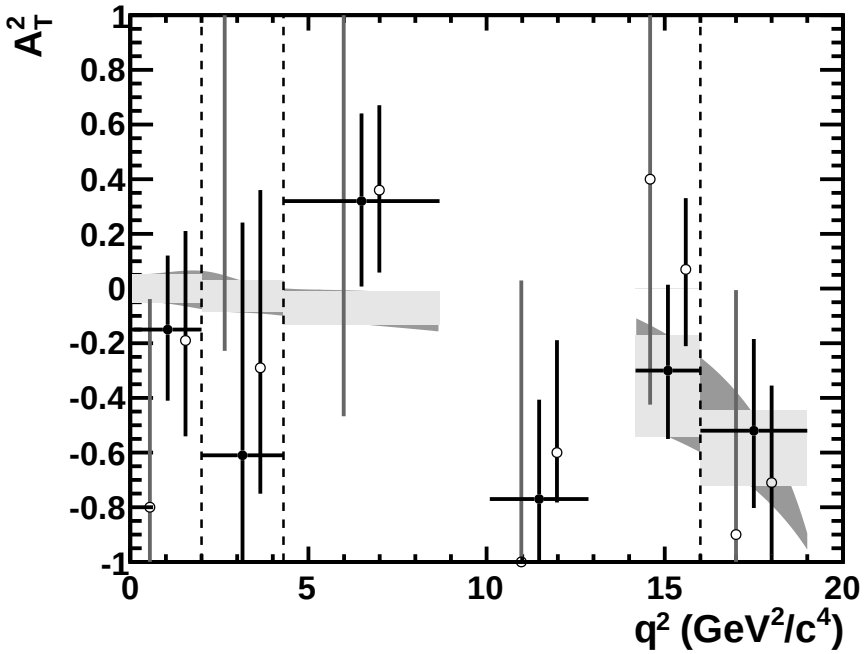
In Sec. 2 the relationships between the angular observables and the Wilson Coefficients $\mathcal{C}_{7,9,10}$ and $\mathcal{C}'_{7,9,10}$ were discussed. The measurement of the $B^0 \to K^{*0}\mu^+\mu^-$ angular observables constrain these Wilson coefficients and therefore set model-independent constraints on NP models. In this section the improvement of the constraints on $\mathcal{C}_{7,9,10}$ and $\mathcal{C}'_{7,9,10}$ as a result of the $1\,\text{fb}^{-1}$ measurement made by the LHCb collaboration [89] are discussed, based on the analysis of Ref. [38].

Ref. [38] considers NP contributions to the six (complex) Wilson coefficients

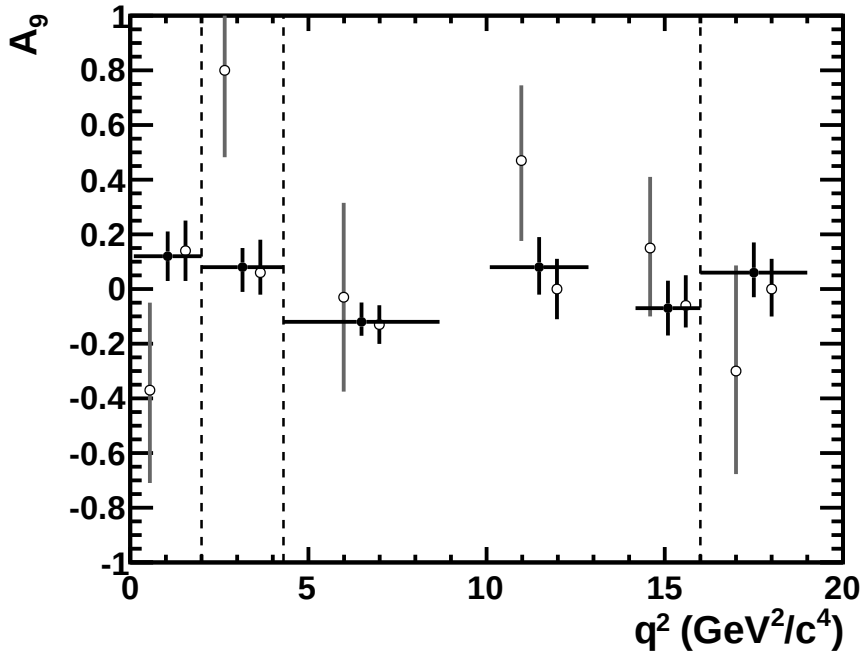
$$\mathcal{C}_7,\ \mathcal{C}_9,\ \mathcal{C}_{10},\ \mathcal{C}'_7,\ \mathcal{C}'_9,\ \mathcal{C}'_{10}. \tag{8.2.1}$$

The sensitivity of each angular observable to NP contributions varies as a function of q^2. At low q^2 the operator $\mathcal{O}_7$ is important but at high q^2 this operator is suppressed and the operators $\mathcal{O}_{9,10}$ become important. Measurements of the $B^0 \to K^{*0}\mu^+\mu^-$ observable in these q^2 regions therefore give complementary constraints on $\mathcal{C}_{7,9,10}$ and $\mathcal{C}'_{7,9,10}$. A list of the observables considered in Ref. [38], and the Wilson coefficients to which each is sensitive, is given in Tab. 8.2.

Measurements of B meson decays other than $B^0 \to K^{*0}\mu^+\mu^-$ also constrain $\mathcal{C}_{7,9,10}$ and $\mathcal{C}'_{7,9,10}$. The experimental inputs to the calculation of the constraints are: the combination of measurements of $\mathcal{B}(B \to X_s\gamma)$, $A_{\rm CP}(b \to s\gamma)$ and $S_{K^*\gamma}$ by the Heavy Flavour Averaging Group (HFAG) [80]; the measurements of $\mathcal{B}(B \to X_s\ell^+\ell^-)$ by the BaBar and Belle collaborations [93, 94]; the measurements of $\mathcal{B}(B \to K^*\ell^+\ell^-)$, $A_{\rm FB}$ and $F_{\rm L}$ by the LHCb, Belle and CDF collaborations [36, 40, 43, 89]; the measurements of S_3 by the LHCb and CDF collaborations [40, 89]; the measurement of A_9 by the CDF collaboration [40]; and the measurement of (or limits set on) $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ by the LHCb, CMS, ATLAS and CDF collaborations [95–98]. The most precise measurements of the $B \to K^*\ell^+\ell^-$ observables are made by the LHCb collaboration.



(a) $A_{\rm T}^2$



(b) A_9

Figure 8.4: Comparison of measurements of $A_{\rm T}^2$ (a) and A_9 (b) by: the CDF (dark grey open circles) and LHCb (black open circles) collaborations; and the present analysis (black circles). The SM prediction curve (dark grey band) and the SM prediction binned in the analysis q^2 bins (light grey area) are indicated. In each q^2 bin the markers are displaced in q^2 for clarity, and dashed lines are overlaid to distinguish between adjacent analysis q^2 bins.

In Ref. [38] the constraints on the Wilson coefficients are calculated using a χ^2 function

$$\chi^2 = \sum_i \frac{\left(\mathrm{Obs}_i^{\mathrm{E}} - \mathrm{Obs}_i^{\mathrm{T}}\right)^2}{\left(\sigma_i^{\mathrm{E}}\right)^2 + \left(\sigma_i^{\mathrm{T}}\right)^2}, \tag{8.2.2}$$

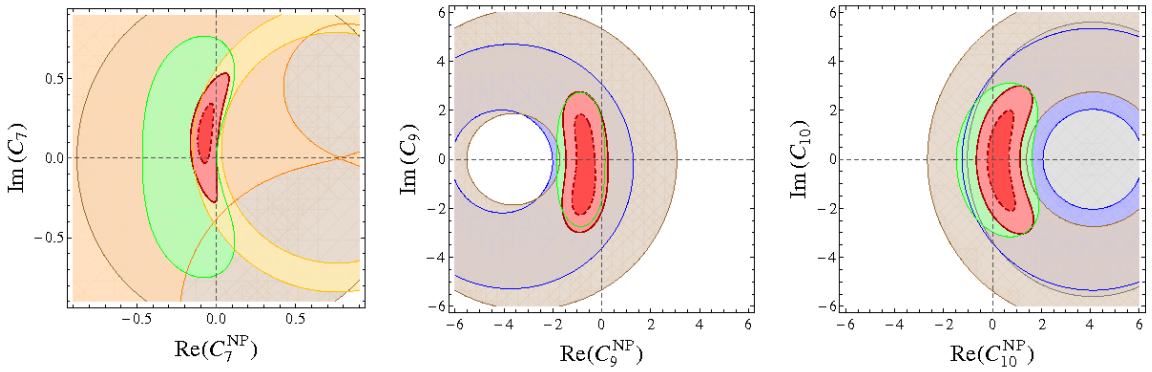
where i is an index over all the observables under consideration, $\mathrm{Obs}_i^{\mathrm{E}}$ is the measured value of observable i, σ_i^{E} is the uncertainty on the measurement, $\mathrm{Obs}_i^{\mathrm{T}}$ is the predicted value of observable i in the SM and σ_i^{T} is the uncertainty on the prediction. The constraints are presented in 2D projections of the 12D space defined by the six (complex) Wilson coefficients. Each of the 12 parameters are allowed to vary simultaneously when calculating the constraints, ensuring that any correlations are taken into account.

The 2σ constraints on the real and imaginary parts of the left-handed Wilson coefficients $\mathcal{C}_{7,9,10}$ are shown in Fig. 8.5(a). The axes are scaled such that the SM prediction is at the origin, so that $\mathrm{Re}(\mathcal{C}_{7,9,10}^{\mathrm{NP}})$ is shown on the x-axes. In the $(\mathrm{Re}(\mathcal{C}_7^{\mathrm{NP}}), \mathrm{Im}(\mathcal{C}_7))$ plane, the measurements of $B^0 \to K^{*0}\mu^+\mu^-$, particularly at low-q^2, constrain the value of $\mathrm{Re}(\mathcal{C}_7^{\mathrm{NP}})$ to be close to the SM prediction. This excludes non-SM values of this quantity that are otherwise consistent with $\mathcal{B}(B \to X_s\gamma)$ constraints and are favoured by the measurement of A_{FB} by the Belle collaboration [36]. The measurements of $B^0 \to K^{*0}\mu^+\mu^-$, particularly at high-q^2, are the dominant constraint on the $(\mathrm{Re}(\mathcal{C}_{9,10}^{\mathrm{NP}}), \mathrm{Im}(\mathcal{C}_{9,10}))$ planes. The measurements exclude large regions of the available phase-space and constrain the $\mathcal{C}_9$ and $\mathcal{C}_{10}$ coefficients to be close to the SM prediction. In summary, the fit is consistent with the SM prediction of $\mathcal{C}_{7,9,10}$ at the 2σ level.

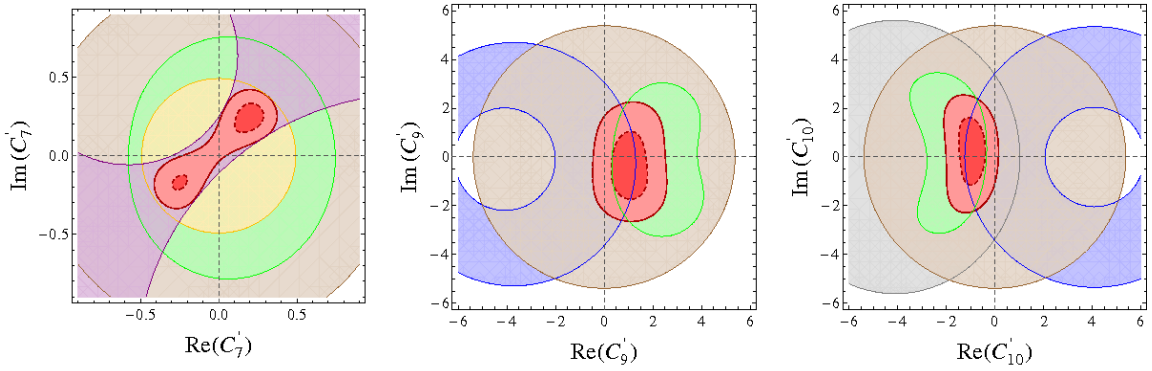
The 2σ constraints on the real and imaginary parts of the right-handed Wilson coefficients $\mathcal{C}'_{7,9,10}$ are shown in Fig. 8.5(b). The axes are scaled such that the SM prediction is at the origin. The measurements of $B^0 \to K^{*0}\mu^+\mu^-$ are the dominant constraints on the $(\mathrm{Re}(\mathcal{C}'_{9,10}), \mathrm{Im}(\mathcal{C}'_{9,10}))$ planes, and the favoured region deviates from the SM prediction at greater than 2σ. However, the fit is consistent with the SM prediction at the 2σ level.

The 2σ constraints on the real parts of the $(\mathcal{C}_{7,9,10}, \mathcal{C}'_{7,9,10})$ Wilson coefficient planes are shown in Fig. 8.5(c). The measurements of $B^0 \to K^{*0}\mu^+\mu^-$ dominate the constraints on the $(\mathrm{Re}(\mathcal{C}_{9,10}), \mathrm{Re}(\mathcal{C}'_{9,10}))$ planes, which again deviate from the SM prediction at a level greater than 2σ. However, the fit is again consistent with the SM prediction at 2σ. The measurements of $B^0 \to K^{*0}\mu^+\mu^-$ also constrain the $(\mathrm{Re}(\mathcal{C}'_7), \mathrm{Im}(\mathcal{C}'_7))$ and $(\mathrm{Re}(\mathcal{C}_7), \mathrm{Re}(\mathcal{C}'_7))$ planes. The sensitivity to $\mathcal{C}_7$ and $\mathcal{C}'_7$ increases as $q^2 \to 0$, hence the dominant constraints on these planes are obtained through $\mathcal{B}(B \to X_s\gamma)$ and $S_{K^*\gamma}$

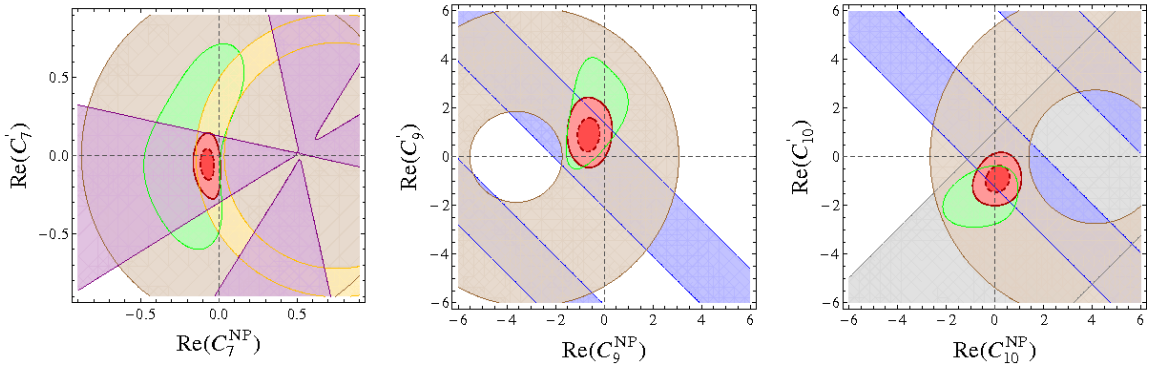
The analysis of Ref. [38] also calculates the effect of a measurement of A_9 by the LHCb collaboration. This has a significant effect on the $(\mathrm{Re}(\mathcal{C}'_{10}), \mathrm{Im}(\mathcal{C}'_{10}))$ plane, as shown in Fig. 8.6. The analysis was performed before a measurement of A_9 was made by LHCb, using the expected precision. The precision of the LHCb measurement of A_9 is



(a) Real and imaginary parts of the left-handed Wilson coefficients



(b) Real and imaginary parts of the right-handed Wilson coefficients



(c) The real parts of the $(\mathcal{C}_{7,9,10}, \mathcal{C}'_{7,9,10})$ plane

Figure 8.5: The 2σ constraints on Wilson Coefficients from measurements of $B \to X_s \ell^+ \ell^-$ (brown), $\mathcal{B}(B \to X_s \gamma)$ (yellow), $A_{\rm CP}(b \to s\gamma)$ (orange), $S_{K^*\gamma}$ (purple), $B \to K^* \mu^+ \mu^-$ (green), $B \to K \mu^+ \mu^-$ (blue), $\mathcal{B}(B_s^0 \to \mu^+ \mu^-)$ (grey), and the combined 1 and 2σ constraints (pink and red respectively). The axes are such that the SM prediction is at the origin. Taken with permission from Ref. [38].

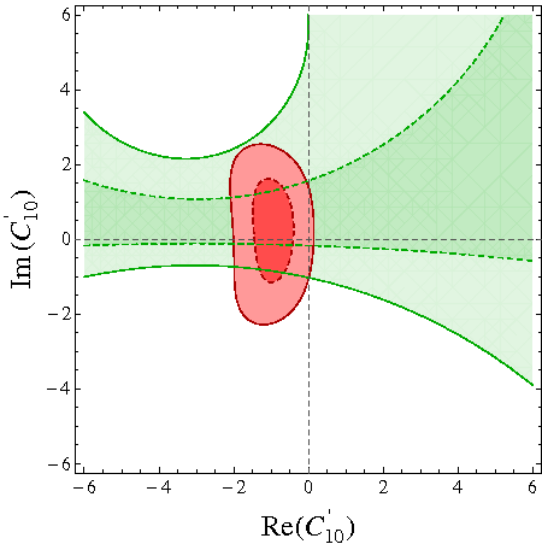


Figure 8.6: Combined 1 and 2σ constraints (pink and red respectively) and the expected 1 and 2σ constraints from a measurement of A_9 by the LHCb collaboration (light and dark green areas respectively) are indicated. The axes are such that the SM prediction is at the origin. Taken with permission from Ref. [38].

comparable to that which was assumed in Ref. [38] above, and is comparable to that of the present analysis.

The constraints on the Wilson coefficients can be interpreted as constraints on the mass scale of NP particles or the coupling strength of the corresponding operators

$$\vec{\mathcal{O}} = \{\mathcal{O}_{7,9,10}, \mathcal{O}'_{7,9,10}\} \tag{8.2.3}$$

by expressing the effective Hamiltonian as

$$\mathcal{H}_{\text{eff}} = -\sum_{i=7,9,10} \left(\frac{c_i}{\Lambda^2}\mathcal{O}_i + \frac{c'_i}{\Lambda^2}\mathcal{O}'_i \right) + \text{h.c.} \ . \tag{8.2.4}$$

In this expression c_i is related to $\mathcal{C}_i$ through

$$\mathcal{C}_i = (1\,\text{TeV})^2 \frac{4G_\text{F}}{\sqrt{2}} V_{tb}V_{ts}^* \frac{e^2}{16\pi^2} c_i. \tag{8.2.5}$$

In Ref. [38], constraints are set on the NP scale Λ by setting the coupling strength $|c_i| = 1$ and trialling values of $c_i/|c_i| = +1, -1, +i, -i$. These constitute maximal coupling strengths for these flavour-changing operators, which are relevant in the case that NP allows generic flavour-changing processes. Constraints are also set on the coupling strength assuming a NP scale of 1 TeV, which is the mass scale at which NP is expected to emerge. Only one Wilson coefficient is varied at a time when calculating the constraints.

Operator	Λ [TeV] for $\lvert c_i\rvert = 1$				$\lvert c_i\rvert$ for $\Lambda = 1$ TeV			
	$+1$	-1	$+i$	$-i$	$+1$	-1	$+i$	$-i$
$\mathcal{O}_7$	69	270	43	38	$1.6 \cdot 10^{-4}$	$9.7 \cdot 10^{-6}$	$4.2 \cdot 10^{-4}$	$5.3 \cdot 10^{-4}$
$\mathcal{O}'_7$	46	70	78	47	$3.6 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$3.5 \cdot 10^{-4}$
$\mathcal{O}_9$	29	64	21	22	$1.2 \cdot 10^{-3}$	$2.4 \cdot 10^{-4}$	$2.2 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$
$\mathcal{O}'_9$	51	22	21	23	$3.8 \cdot 10^{-4}$	$2.1 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$
$\mathcal{O}_{10}$	43	33	23	23	$5.4 \cdot 10^{-4}$	$9.2 \cdot 10^{-4}$	$1.9 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$
$\mathcal{O}'_{10}$	25	89	24	23	$1.7 \cdot 10^{-3}$	$1.3 \cdot 10^{-4}$	$1.7 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$

Table 8.3: Constraints on the NP scale Λ, in units of TeV, with $c_i = +1, -1, +i, -i$, or constraints on $|c_i|$ with $\Lambda = 1\,\text{TeV}$. Modified from Ref. [38].

The constraints on Λ and $|c_i|$ are presented in Tab. 8.3. The scale of NP for each operator is constrained to $\gtrsim 20\,\text{TeV}$. The coupling strength of each operator at 1 TeV is constrained to be $\lesssim 10^{-3}$. These results imply that either the mass scale of NP is greater than $\sim 20\,\text{TeV}$, or that NP flavour-changing couplings are highly suppressed.

These constraints raise the minimum mass scale of NP particles contributing in $\Delta_F = 1$ flavour-changing operators to be $\gtrsim 20\,\text{TeV}$. This minimum scale is now considerably higher than the $\sim 1\,\text{TeV}$ scale at which many models predict NP particles to appear, therefore requiring the suppression of flavour-changing couplings of the particles. Such a suppression is consistent with the Minimal Flavour Violation (MFV) hypothesis.

Chapter 9

Conclusions

This thesis presented the world's most precise measurement of the $B^0 \to K^{*0}\mu^+\mu^-$ observables. The measurements place the dominant constraints on each of the relevant Wilson coefficients, thereby setting model independent constraints on NP models. In the analysis of Ref. [38], the measurements by the LHCb collaboration presented in Ref. [89] are interpreted as model independent constraints on NP models via the Wilson coefficients. Similarly, assuming the flavour couplings of NP particles are of order unity, the analysis presented in this thesis constrains the mass scale of NP particles to be $\gtrsim 20\,\text{TeV}$. This mass scale is much higher than the $\sim$ TeV scale at which NP particles are expected. If NP particles are at the $\sim$ TeV mass scale, their couplings must be suppressed to the level of $\sim 10^{-3}$. Such a suppression would be consistent with the MFV hypothesis.

Bibliography

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

[2] A. H. G. Peter, *Dark matter: a brief review*, arXiv:1201.3942.

[3] K. Garrett and G. Duda, *Dark matter: a primer*, Adv. Astron. **2011** (2011) 968283, arXiv:1006.2483.

[4] Planck Collaboration, P. A. R. Ade *et al.*, *Planck 2013 results. XVI. Cosmological parameters*, arXiv:1303.5076.

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

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

[7] ATLAS Collaboration, G. Aad *et al.*, *The ATLAS experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008), no. 08 S08003.

[8] CMS Collaboration, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, Journal of Instrumentation **3** (2008), no. 08 S08004.

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

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

[11] L. Evans and P. Bryant, *LHC machine*, Journal of Instrumentation **3** (2008), no. 08 S08001.

[12] LHCb collaboration, A. A. Alves Jr *et al.*, *The LHCb detector at the LHC*, Journal of Instrumentation **3** (2008), no. 08 S08005.

[13] C. Burgess and G. Moore, *The Standard Model: A Primer*, Cambridge University Press, 2006.

[14] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*, Cambridge University Press, 1996.

[15] I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics Volume II: QCD and the Electroweak Theory*, Taylor and Francis Group, 2004.

[16] G. Isidori, Y. Nir and G. Perez, *Flavor physics constraints for physics beyond the Standard Model*, Ann. Rev. Nucl. Part. Sci. **60** (2010) 355, arXiv:1002.0900.

[17] UTfit Collaboration, D. Derkach, *Unitarity triangle fitter results for CKM Angles*, 2013. arXiv:1301.3300.

[18] M. Ciuchini *et al.*, *2000 CKM triangle analysis: a critical review with updated experimental inputs and theoretical parameters*, JHEP **0107** (2001) 013, arXiv:hep-ph/0012308.

[19] G. D'Ambrosio, G. F. Giudice, G. Isidori and A. Strumia, *Minimal flavor violation: an effective field theory approach*, Nucl. Phys. **B645** (2002) 155, arXiv:hep-ph/0207036.

[20] W. Altmannshofer, P. Ball, A. Bharucha A. J. Buras, D. M. Straub and M. Wick, *Symmetries and asymmetries of $B \to K^* \mu^+ \mu^-$ decays in the Standard Model and beyond*, JHEP **0901** (2009) 019, arXiv:0811.1214.

[21] C. Bobeth, A. J. and Buras, F. Kruger and J. Urban, *QCD corrections to $\bar{B} \to X_{d,s} \nu \bar{\nu}$, $\bar{B}_{d,s} \to \ell^+ \ell^-$, $K \to \pi \nu \bar{\nu}$ and $K_L \to \mu^+ \mu^-$ in the MSSM*, Nucl. Phys. **B630** (2002) 87, arXiv:hep-ph/0112305.

[22] LHCb Collaboration, B. Adeva *et al.*, *Roadmap for selected key measurements of LHCb*, arXiv:0912.4179.

[23] W. Reece and U. Egede, *Exploiting angular correlations in the rare decay $B^0 \to K^{*0} \mu^+ \mu^-$ at LHCb*, PhD thesis, London, Imperial College London, London, 2010, Presented on 14 May 2010.

[24] M. Beylich, G. Buchalla and T. Feldmann, *Theory of $B \to K^{(*)} \ell^+ \ell^-$ decays at high q^2: OPE and quark-hadron duality*, Eur. Phys. J. **C71** (2011) 1635, arXiv:1101.5118.

[25] M. Beneke, G. Buchalla, M. Neubert and C. T. Sachrajda, *QCD factorization for exclusive, nonleptonic B meson decays: general arguments and the case of heavy light final states*, Nucl. Phys. **B591** (2000) 313, arXiv:hep-ph/0006124.

[26] C. W. Bauer*et al.*, *An effective field theory for collinear and soft gluons: heavy to light decays*, Phys. Rev. **D63** (2001) 114020, arXiv:hep-ph/0011336.

[27] 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.

[28] M. Artuso, D. M. Asner, P. Ball, E. Baracchini, G. Bell *et al.*, *B, D and K decays*, Eur. Phys. J. **C57** (2008) 309, arXiv:0801.1833.

[29] T. Hurth and M. Nakao, *Radiative and Electroweak Penguin Decays of B Mesons*, Ann. Rev. Nucl. Part. Sci. **60** (2010) 645, arXiv:1005.1224.

[30] M. Beneke, T. Feldmann and D. Seidel, *Systematic approach to exclusive $B \to V\ell^+\ell^-$, $V\gamma$ decays*, Nucl. Phys. **B612** (2001) 25, arXiv:hep-ph/0106067.

[31] S. Jager and J. M. Camalich, *On $B \to V\ell\ell$ at small dilepton invariant mass, power corrections, and new physics*, JHEP **1305** (2013) 043, arXiv:1212.2263.

[32] J. Matias, F. Mescia, M. Ramon and J. Virto, *Complete anatomy of $\bar{B}^0 \to \bar{K}^{*0}(\to K\pi)\ell^+\ell^-$ and its angular distribution*, JHEP **1204** (2012) 104, arXiv:1202.4266.

[33] Z. Liu *et al.*, *A lattice calculation of $B \to K^{(*)}$ form factors*, arXiv:1101.2726.

[34] A. Ali, E. Lunghi, C. Greub and G. Hiller, *Improved model-independent analysis of semileptonic and radiative rare B decays*, Phys. Rev. D **66** (2002) 034002.

[35] 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.

[36] BELLE Collaboration, J.-T. Wei *et al.*, *Measurement of the differential branching fraction and forward-backward asymmetry for $B \to K^{(*)}\ell^+\ell^-$*, Phys. Rev. Lett. **103** (2009) 171801, arXiv:0904.0770.

[37] BaBar Collaboration, J. L. Ritchie, *Angular analysis of $B \to K^*\ell^+\ell^-$ in BABAR*, in *Proceedings of CKM 2012, the 7th International Workshop on the CKM Unitarity Triangle*, 2013. arXiv:1301.1700.

[38] W. Altmannshofer and D. M. Straub, *Cornering New Physics in $b \to s$ transitions*, JHEP **1208** (2012) 121, arXiv:1206.0273.

[39] D. Becirevic and E. Schneider, *On transverse asymmetries in $B \to K^* \ell^+ \ell^-$*, Nucl. Phys. **B854** (2012) 321, arXiv:1106.3283.

[40] CDF Collaboration, T. Aaltonen *et al.*, *Measurements of the angular distributions in the decays $B \to K^{(*)} \mu^+ \mu^-$ at CDF*, Phys. Rev. Lett. **108** (2012) 081807, arXiv:1108.0695.

[41] BABAR Collaboration, B. Aubert *et al.*, *Ambiguity-free measurement of* $\cos(2\beta)$*: time-integrated and time-dependent angular analyses of $B \to J/\psi K \pi$*, Phys. Rev. **D71** (2005) 032005, arXiv:hep-ex/0411016.

[42] LHCb Collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0} \mu^+ \mu^-$*, Phys. Rev. Lett. **108** (2012) 181806, arXiv:1112.3515.

[43] CDF Collaboration, T. Aaltonen *et al.*, *Observation of the baryonic flavor-changing neutral current decay $\Lambda_b \to \Lambda \mu^+ \mu^-$*, Phys. Rev. Lett. **107** (2011) 201802, arXiv:1107.3753.

[44] BaBar Collaboration, B. Aubert *et al.*, *Angular distributions in the decays $B \to K^* \ell^+ \ell^-$*, Phys. Rev. **D79** (2009) 031102, arXiv:0804.4412.

[45] V. Coco, *Luminosity: experimental prospects and challenges, LHCb*, in *LHC Lumi Days 2012, CERN, Geneva, Switzerland*, 2012.

[46] P. M. Bjørnstad, *Performance of the LHCb Vertex Locator*, Journal of Instrumentation **6** (2011), no. 12 C12024.

[47] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, Journal of Instrumentation **8** (2013), no. 04 P04022, arXiv:1211.3055.

[48] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, Journal of Instrumentation **8** (2013) 2022P, arXiv:1211.1346.

[49] G. Lanfranchi *et al.*, *The muon identification procedure of the LHCb experiment for the first data*, Tech. Rep. LHCb-PUB-2009-013. CERN-LHCb-PUB-2009-013, CERN, Geneva, 2009.

[50] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, arXiv:1306.0249.

[51] A. Jaeger *et al.*, *Measurement of the track finding efficiency*, Tech. Rep. LHCb-PUB-2011-025. CERN-LHCb-PUB-2011-025, CERN, Geneva, 2012.

[52] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, European Physical Journal C **73** (2013) 2431, arXiv:1211.6759.

[53] V. V. Gligorov, C. Thomas and M. Williams, *The HLT inclusive B triggers*, Tech. Rep. LHCb-PUB-2011-016. CERN-LHCb-PUB-2011-016., CERN, Geneva, Sep, 2011.

[54] R. Aaij and J. Albrecht, *Muon triggers in the High Level Trigger of LHCb*, Tech. Rep. LHCb-PUB-2011-017. CERN-LHCb-PUB-2011-017, CERN, Geneva, 2011.

[55] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the ASME Journal of Basic Engineering (1960), no. 82 (Series D) 35.

[56] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444.

[57] S. Kullback and R. A. Leibler, *On information and sufficiency*, The Annals of Mathematical Statistics **22** (1951), no. 1 pp. 79.

[58] M. Needham, *Clone track identification using the Kullback-Liebler distance*, Tech. Rep. LHCb-2008-002. CERN-LHCb-2008-002. LPHE-2008-002, CERN, Geneva, 2008.

[59] V. V. Gligorov, *A single track HLT1 trigger*, Tech. Rep. LHCb-PUB-2011-003. CERN-LHCb-PUB-2011-003. LHCb-INT-2010-053, CERN, Geneva, 2011.

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

[61] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

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

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

[64] 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.

[65] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: a simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[66] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[67] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B^0_s \to J/\psi \bar{K}^{*0}$ branching fraction and angular amplitudes*, Phys. Rev. D **86** (2012) 071102., arXiv:1208.0738.

[68] LHCb collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, arXiv:1301.5286.

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

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

[71] 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.

[72] A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039.

[73] M. Pivk and F. R. Le Diberder, *Plots: a statistical tool to unfold data distributions*, Nuclear Instruments and Methods in Physics Research A **555** (2005) 356, arXiv:physics/0402083.

[74] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the CP asymmetry in $B^0 \to K^{*0}\mu^+\mu^-$ decays*, Phys. Rev. Lett. **110** (2013) 031801, arXiv:1210.4492.

[75] A. Sarti, S. Furcas, G. Lanfranchi and M. Palutan, *Calibration strategy and efficiency measurement of the muon identification procedure at LHCb*, Tech. Rep. LHCb-PUB-2010-002. CERN-LHCb-PUB-2010-002, CERN, Geneva, 2010.

[76] 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.

[77] Belle Collaboration, K. Abe *et al.*, *Measurements of branching fractions and decay amplitudes in $B \to J/\psi K^*$ decays*, Phys. Lett. **B538** (2002) 11, arXiv:hep-ex/0205021.

[78] LHCb Collaboration, R. Aaij *et al.*, *Measurement of J/ψ production in pp collisions at $\sqrt{s} = 7\,\text{TeV}$*, Eur. Phys. J. **C71** (2011) 1645, arXiv:1103.0423.

[79] LHCb Collaboration, R. Aaij *et al.*, *Absolute luminosity measurements with the LHCb detector at the LHC*, JINST **7** (2012) P01010, arXiv:1110.2866.

[80] Heavy Flavor Averaging Group, D. Asner *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties*, arXiv:1010.1589.

[81] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[82] LHCb Collaboration, R. Aaij *et al.*, *Measurement of b-hadron masses*, Phys. Lett. **B708** (2012) 241, arXiv:1112.4896.

[83] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the ratio of branching fractions for $B^0_s \to \phi\mu^+\mu^-$ and $B^0_s \to J/\psi\,\phi$*, 2012. LHCb-CONF-2012-003.

[84] BABAR Collaboration, B. Aubert *et al.*, *Measurement of the $B \to J/\psi\, K^*(892)$ decay amplitudes*, Phys. Rev. Lett. **87** (2001) 241801, arXiv:hep-ex/0107049.

[85] G. J. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Phys. Rev. D **57** (1998) 3873, arXiv:physics/9711021.

[86] G. Zech and B. Aslan, *New test for the multivariate two-sample problem based on the concept of minimum energy*, J. Stat. Compu. Simul. **75 (2)** (2005) 109, arXiv:math/0309164.

[87] M. Williams, *How good are your fits? Unbinned multivariate goodness-of-fit tests in high energy physics*, JINST **5** (2010) P09004, arXiv:1006.3019.

[88] BABAR Collaboration, B. Aubert *et al.*, *Measurement of decay amplitudes of $B \to J/\psi\, K^*$, $\psi(2S)K^*$, and $\chi_{c1}K^*$ with an angular analysis*, Phys. Rev. **D76** (2007) 031102, arXiv:0704.0522.

[89] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$*, arXiv:1304.6325.

[90] CMS Collaboration, S. Chatrchyan *et al.*, *Angular analysis and branching ratio measurement of the decay $B^0 \to K^{*0}\mu^+\mu^-$*, Tech. Rep. CMS-PAS-BPH-11-009, CERN, Geneva, 2013.

[91] 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.

[92] ATLAS Collaboration, G. Aad *et al.*, *Angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ with the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2013-038, CERN, Geneva, 2013.

[93] BaBar Collaboration, B. Aubert *et al.*, *Measurement of the $B \to X_s\ell^+\ell^-$ branching fraction with a sum over exclusive modes*, Phys. Rev. Lett. **93** (2004) 081802, arXiv:hep-ex/0404006.

[94] Belle Collaboration, M. Iwasaki *et al.*, *Improved measurement of the electroweak penguin process $B \to X_s\ell^+\ell^-$*, Phys. Rev. **D72** (2005) 092005, arXiv:hep-ex/0503044.

[95] LHCb Collaboration, R. Aaij *et al.*, *Strong constraints on the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$*, Phys. Rev. Lett. **108** (2012) 231801, arXiv:1203.4493.

[96] CMS Collaboration, S. Chatrchyan *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays*, JHEP **1204** (2012) 033, arXiv:1203.3976.

[97] ATLAS Collaboration, G. Aad *et al.*, *Search for the decay $Bs \to \mu^+\mu^-$ with the ATLAS detector*, Phys. Lett. **B713** (2012) 387, arXiv:1204.0735.

[98] CDF Collaboration, T. Aaltonen *et al.*, *Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with CDF II*, Phys. Rev. Lett. **107** (2011) 191801, arXiv:1107.2304.

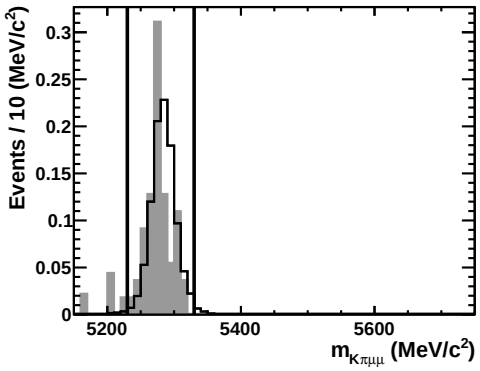
Glossary

SM	Standard Model	15
CP	charge-parity	15
NP	New Physics	16
FCNC	Flavour-Changing Neutral Current	16
LHC	Large Hadron Collider	16
HQET	Heavy Quark Effective Theory	28
QCDF	QCD Factorisation	28
SCET	Soft-Collinear Effective Theory	28
FFI	form-factor independent	30
FFD	form-factor dependent	30
MFV	Minimal Flavour Violation	25
PV	primary vertex	43
SV	secondary vertex	44
IP	impact parameter	44
PID	particle identification	44
VELO	Vertex Locator	45
TT	Tracker Turicensis	45
RICH	Ring Imaging Cherenkov	45
IT	Inner Tracker	48
OT	Outer Tracker	48
ST	Silicon Tracker	48
HPD	Hybrid Photon Detector	50
DLL	difference in log-likelihood	50
ECAL	Electromagnetic Calorimeter	50
HCAL	Hadronic Calorimeter	50
SPD	Scintillator Pad Detector	50
PS	Pre-Shower	51

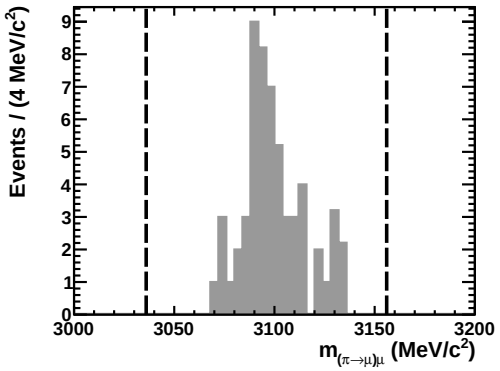
MWPC Multi-Wire Proportion Chamber 51
IsMuon stringent muon identification criteria 51
IsMuonLoose less stringent muon identification criteria 52
pdf probability density function 55
L0 First Level Trigger 57
HLT High Level Trigger 57
HLT1 first stage of the HLT 57
HLT2 second stage of the HLT 57
KL Kullback-Leibler distance 60
DOCA distance of closest approach 61
MVA multivariate analysis 64
BDT Boosted Decision Tree 76
DT Decision Tree 76
AC acceptance correction 97
CB Crystal Ball 118
FC Feldman-Cousins 133
HFAG Heavy Flavour Averaging Group 172

Appendix A

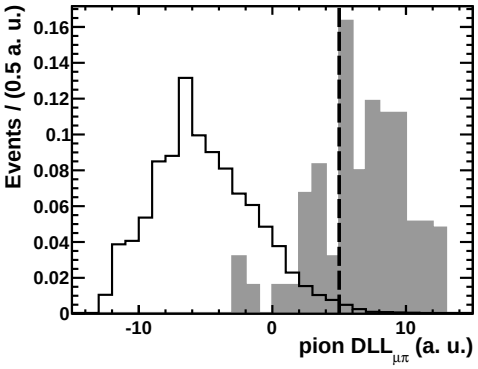
Supplementary plots regarding the selection



(a) $\mathrm{m}_{K\pi\mu\mu}$

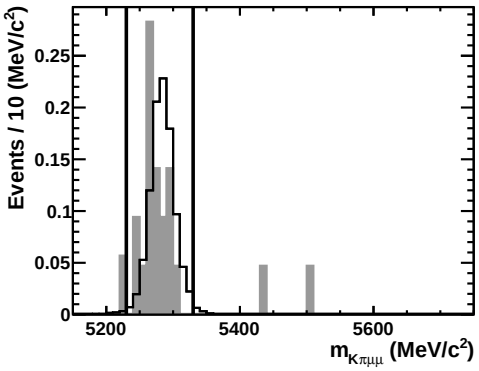


(b) $\mathrm{m}_{(\pi\to\mu)\mu}$

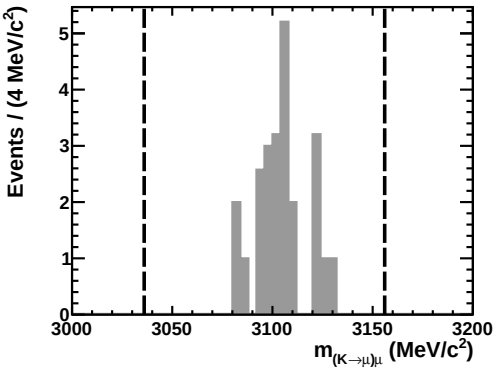


(c) pion $\mathrm{DLL}_{\mu\pi}$

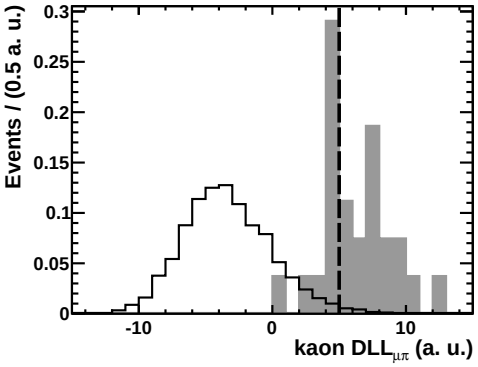
Figure A.1: Distributions of invariant masses and DLL values used to veto $B^0 \to J/\psi\, K^{*0}$ decays where the pion is misidentified as a muon and the muon is misidentified as a pion, for $B^0 \to K^{*0}\mu^+\mu^-$ (black solid histogram) and misidentified $B^0 \to J/\psi\, K^{*0}$ (grey shaded histogram) simulated events. Solid black lines indicate selected regions; dashed black lines indicate veto requirements. The $\mathrm{m}_{K\pi\mu\mu}$ (a), $\mathrm{m}_{\mu\mu}$ (b) and pion $\mathrm{DLL}_{\mu\pi}$ (c) distributions are shown.



(a) $\mathrm{m}_{K\pi\mu\mu}$

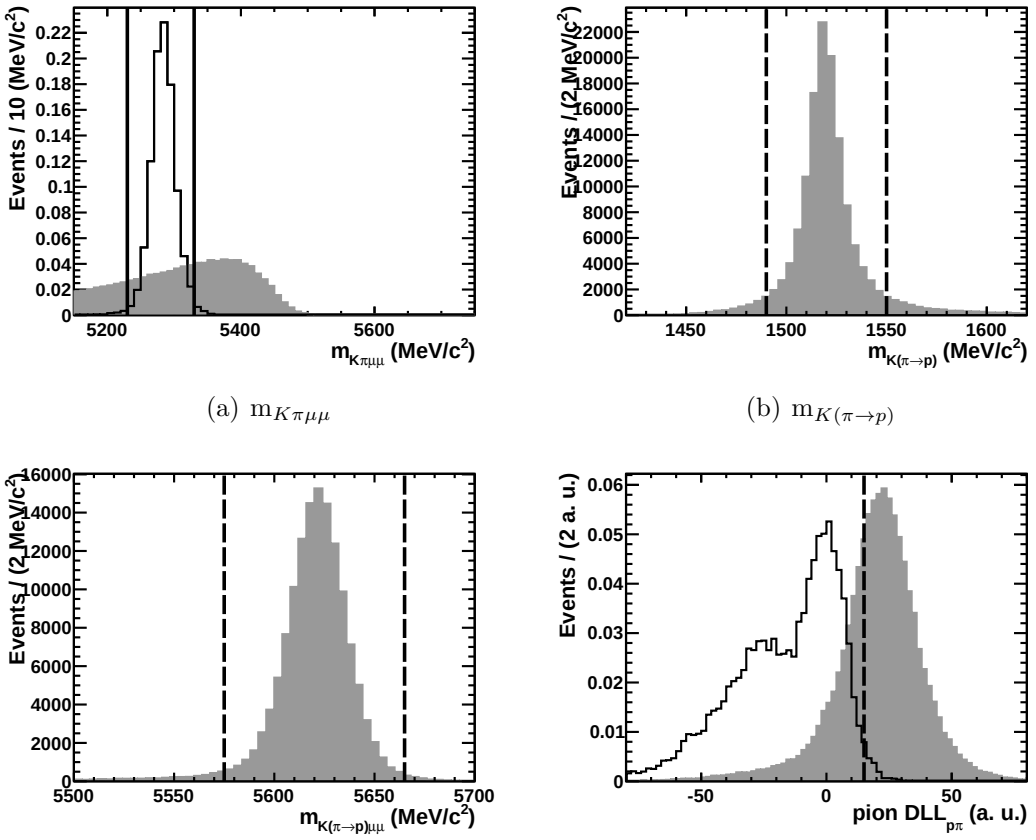


(b) $\mathrm{m}_{(K\to\mu)\mu}$



(c) kaon $\mathrm{DLL}_{\mu\pi}$

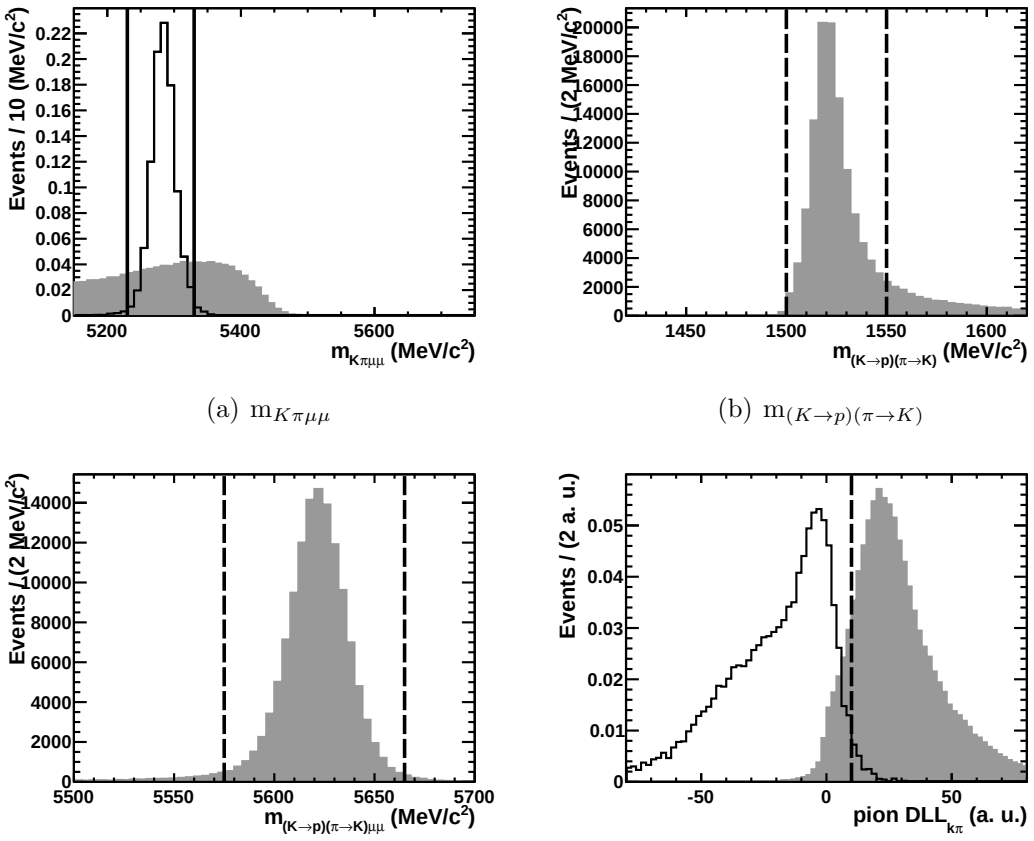
Figure A.2: Distributions of invariant masses and DLL values used to veto $B^0 \to J/\psi\, K^{*0}$ decays where the kaon is misidentified as a muon and the muon is misidentified as a kaon, for $B^0 \to K^{*0}\mu^+\mu^-$ (black solid histogram) and misidentified $B^0 \to J/\psi\, K^{*0}$ (grey shaded histogram) simulated events. Solid black lines indicate selected regions; dashed black lines indicate veto requirements. The $\mathrm{m}_{K\pi\mu\mu}$ (a), $\mathrm{m}_{\mu\mu}$ (b) and kaon $\mathrm{DLL}_{\mu\pi}$ (c) distributions are shown.



(a) $\mathrm{m}_{K\pi\mu\mu}$ (b) $\mathrm{m}_{K(\pi\to p)}$

(c) $\mathrm{m}_{K(\pi\to p)\mu\mu}$ (d) pion $\mathrm{DLL}_{p\pi}$

Figure A.3: Distributions of invariant masses and DLL values used to veto $\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$ decays where the proton is misidentified as a pion, for $B^0 \to K^{*0}\mu^+\mu^-$ (black solid histogram) and $\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$ (grey shaded histogram) simulated events. Solid black lines indicate selected regions; dashed black lines indicate veto requirements. The $\mathrm{m}_{K\pi\mu\mu}$ (a), m_{pK} (b), $\mathrm{m}_{pK\mu\mu}$ (c) and pion $\mathrm{DLL}_{p\pi}$ (d) distributions are shown.



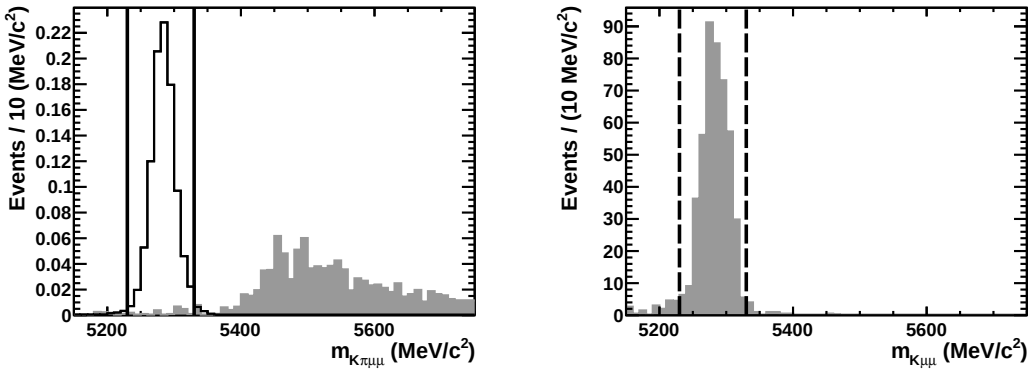
(a) $\mathrm{m}_{K\pi\mu\mu}$

(b) $\mathrm{m}_{(K\to p)(\pi\to K)}$

(c) $\mathrm{m}_{(K\to p)(\pi\to K)\mu\mu}$

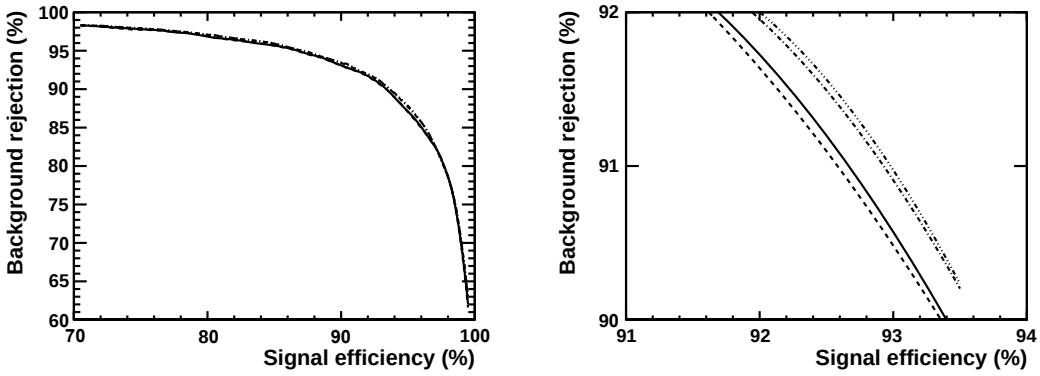
(d) pion $\mathrm{DLL}_{K\pi}$

Figure A.4: Distributions of invariant masses and DLL values used to veto $\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$ decays where the proton is misidentified as a kaon and the kaon is misidentified as a pion, for $B^0 \to K^{*0}\mu^+\mu^-$ (black solid histogram) and $\Lambda_b^0 \to \Lambda^{*0}(1520)\mu^+\mu^-$ (grey shaded histogram) simulated events. Solid black lines indicate selected regions; dashed black lines indicate veto requirements. The $\mathrm{m}_{K\pi\mu\mu}$ (a), m_{pK} (b), $\mathrm{m}_{pK\mu\mu}$ (c) and pion $\mathrm{DLL}_{K\pi}$ (d) distributions are shown.



(a) $m_{K\pi\mu\mu}$ (b) $m_{K\mu\mu}$

Figure A.5: Distributions of invariant masses used to veto $B^+ \to K^+\mu^+\mu^-$ decays for $B^0 \to K^{*0}\mu^+\mu^-$ (black solid histogram) and mis-reconstructed $B^+ \to K^+\mu^+\mu^-$ (grey shaded histogram) simulated events. Solid black lines indicate selected regions; dashed black lines indicate veto requirements. The $m_{K\pi\mu\mu}$ (a) and $m_{K\mu\mu}$ (b) distributions are shown.



(a) Full BDT performance curve (b) Zoomed BDT performance curve

Figure A.6: The signal efficiency and background rejection (1 - efficiency) as a function of the requirement imposed on the BDT response, for BDTs produced using: only the initial variable set (solid line); the initial variable set with kaon and pion $DLL_{\mu\pi}$ added (dashed line); the initial variable set with μ^+ and μ^- $DLL_{\mu\pi}$ added (dot-dashed line); the initial variable set with $DLL_{\mu\pi}$ added to all final state particles (triple-dot-dashed line). No significant improvement of the BDT performance is visible. The full range (a) and a zoomed view where the separation of the individual curves is visible (b) are shown.

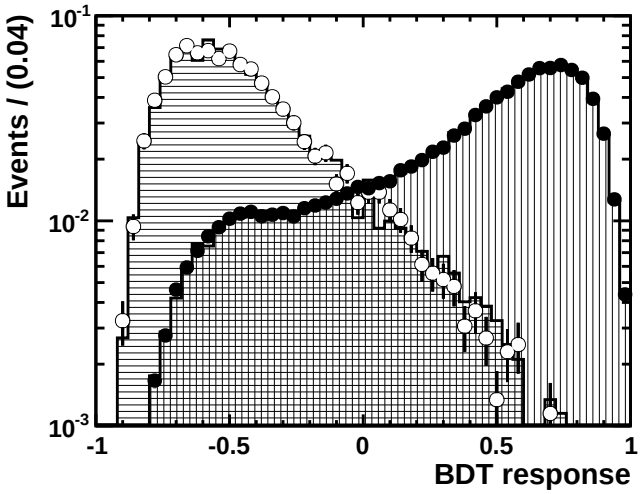
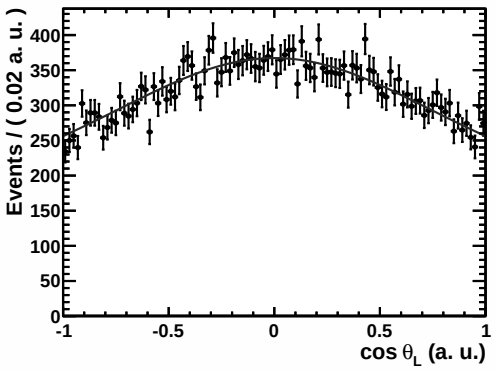


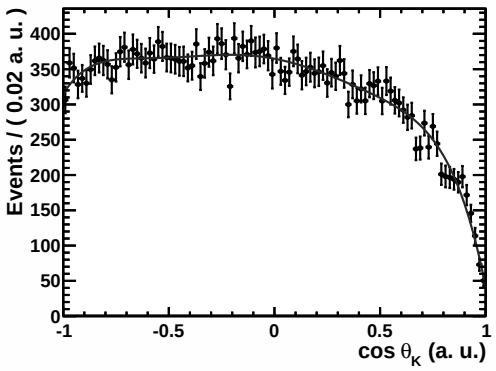
Figure A.7: Comparison of BDT B response on the signal input (blue circles), background input (red open circles), signal independent (blue vertical striped histogram) and background independent (red horizontally striped histogram) samples for BDT B. Distributions are normalised to equal area.

Appendix B

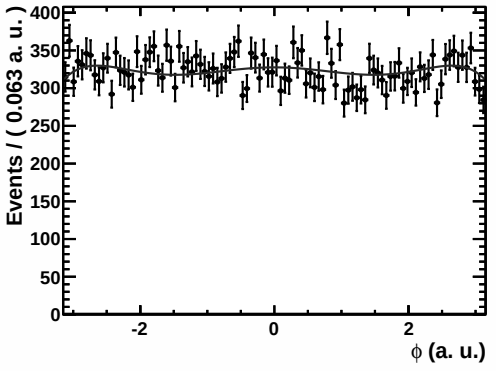
Supplementary plots regarding the acceptance effects



(a) $\cos\theta_L$

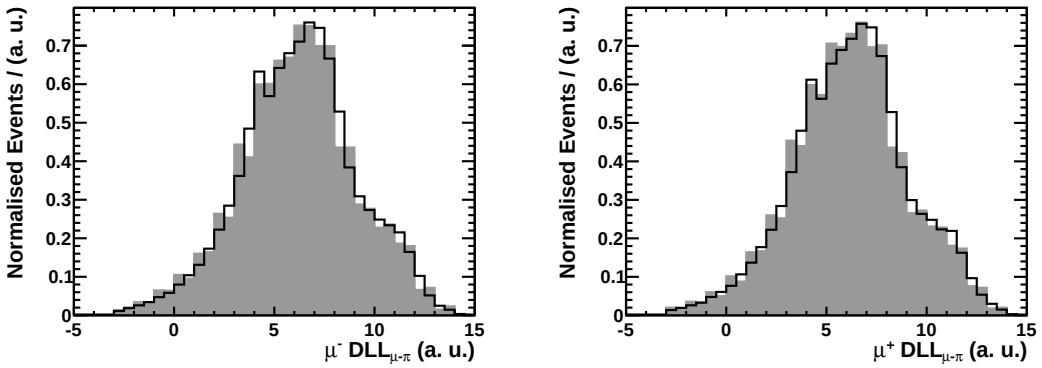


(b) $\cos\theta_K$



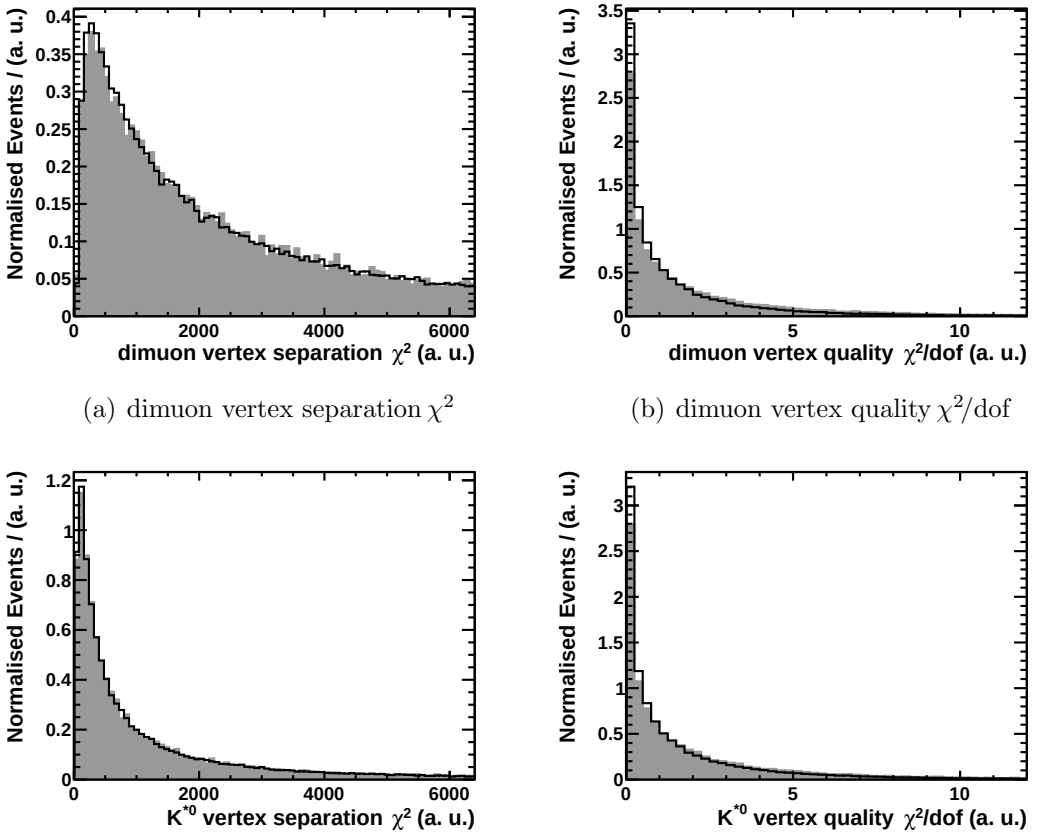
(c) ϕ

Figure B.1: Angular distribution of simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $3.40 < q^2 < 3.60\,\mathrm{GeV}^2/c^4$. The fitted acceptance correction pdf is overlaid.



(a) μ^- DLL$_{\mu\pi}$ (b) μ^+ DLL$_{\mu\pi}$

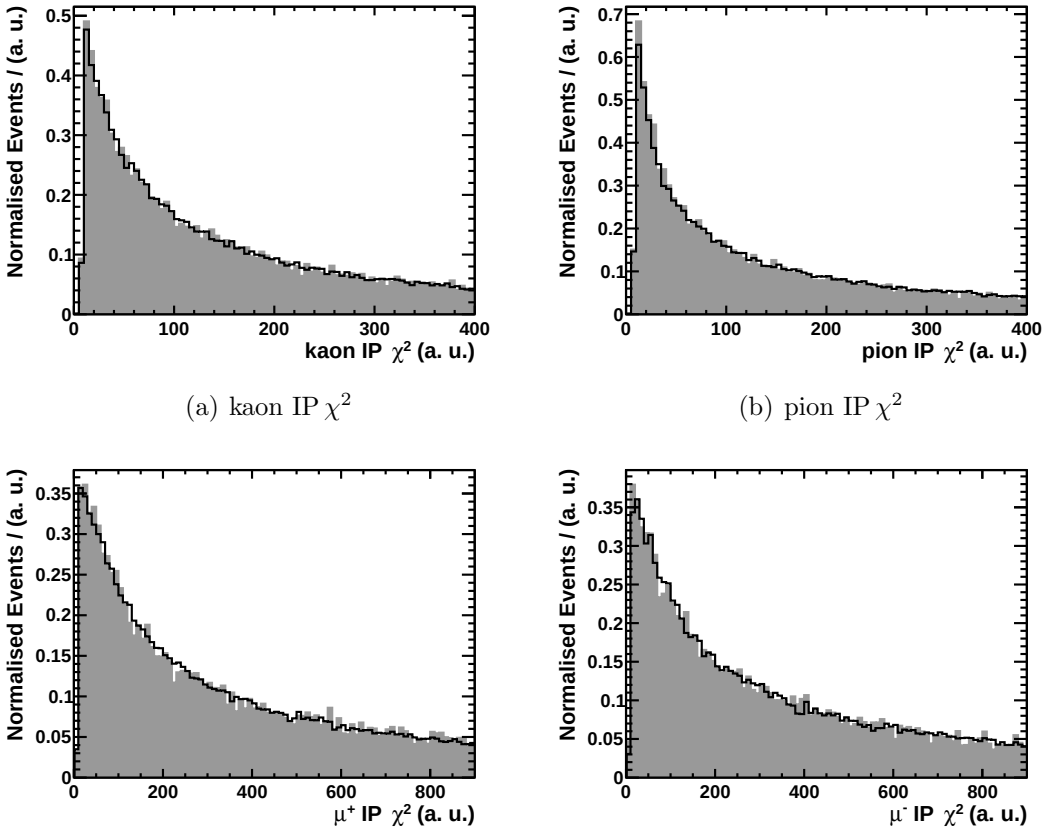
Figure B.2: Comparison of the DLL$_{\mu\pi}$ of the muons in $B^0 \to J/\psi\, K^{*0}$ candidates from the data (black solid histogram) and simulated $B^0 \to J/\psi\, K^{*0}$ events (grey shaded histogram).



(a) dimuon vertex separation χ^2 (b) dimuon vertex quality χ^2/dof

(c) K^{*0} vertex separation χ^2 (d) K^{*0} vertex quality χ^2/dof

Figure B.3: Comparison of the dimuon and K^{*0} properties used in the pre-selection in $B^0 \to J/\psi\, K^{*0}$ candidates from the data (black solid histogram) and simulated $B^0 \to J/\psi\, K^{*0}$ events (grey shaded histogram).



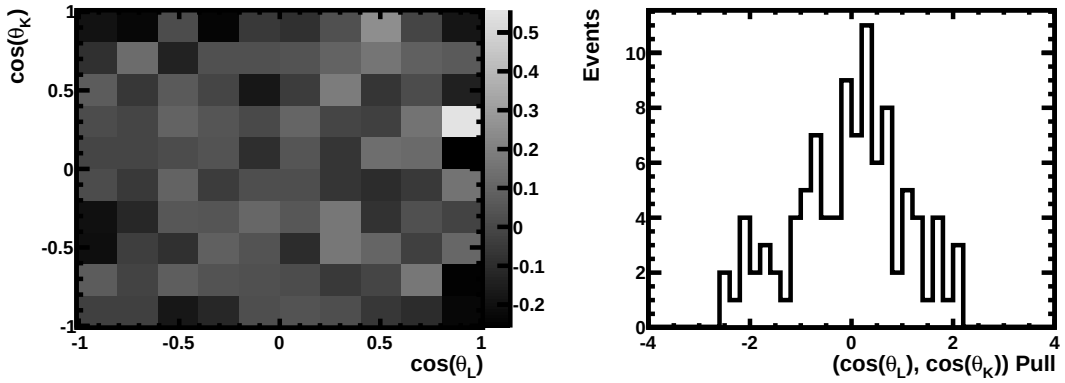
(a) kaon IP χ^2

(b) pion IP χ^2

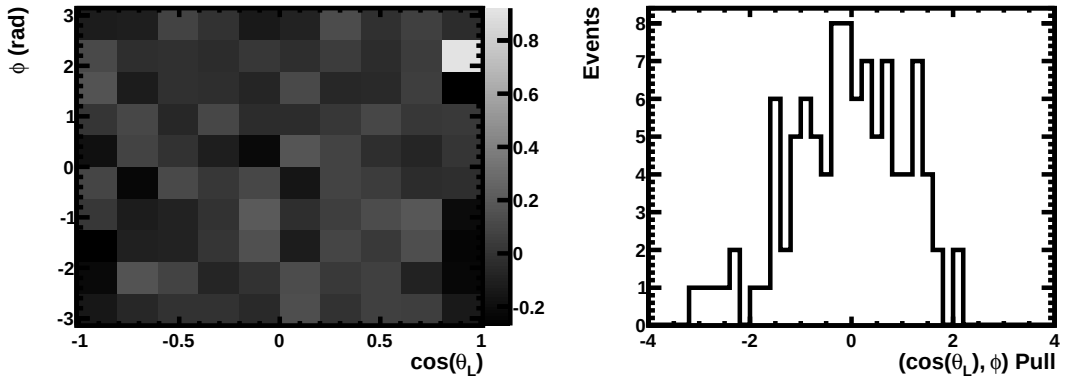
(c) μ^+ IP χ^2

(d) μ^- IP χ^2

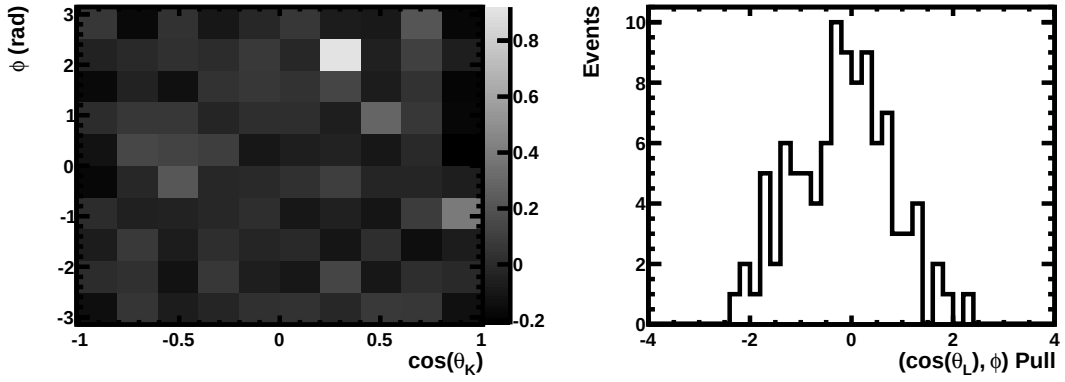
Figure B.4: Comparison of the IP χ^2 of the final state particles in $B^0 \to J/\psi\, K^{*0}$ candidates from the data (black solid histogram) and simulated $B^0 \to J/\psi\, K^{*0}$ events (grey shaded histogram).



(a) ($\cos\theta_L$, $\cos\theta_K$) and associated pull distribution

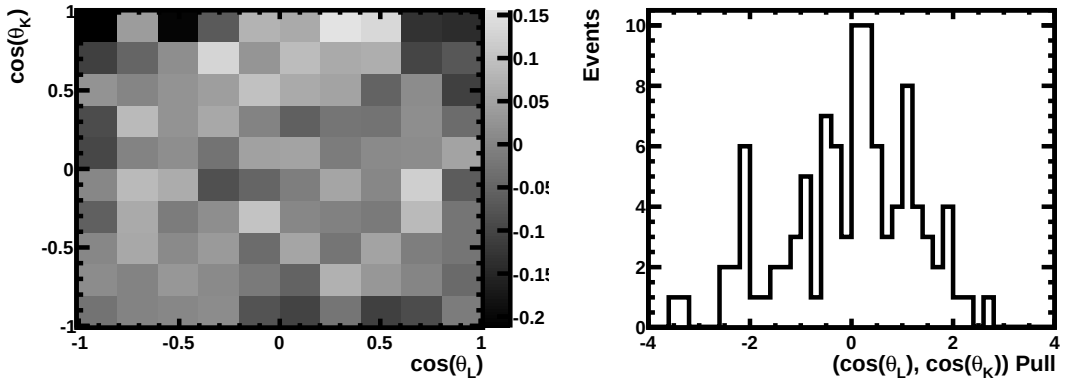


(b) ($\cos\theta_L$, ϕ) and associated pull distribution

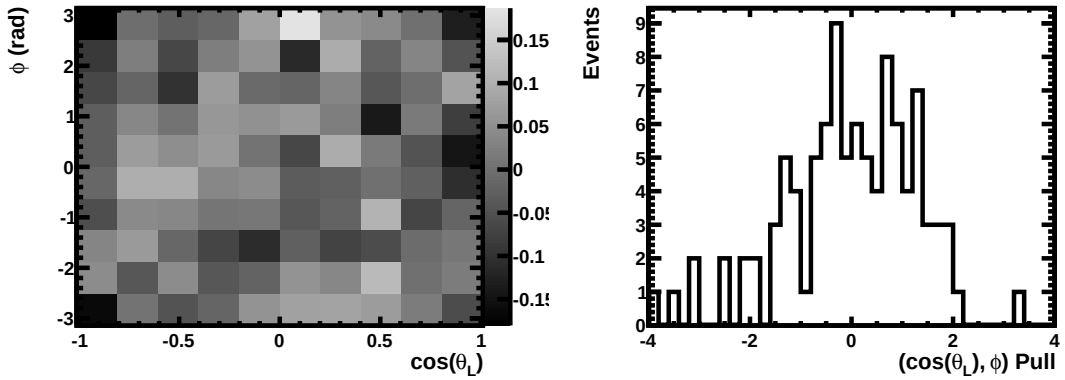


(c) ($\cos\theta_K$, ϕ) and associated pull distribution

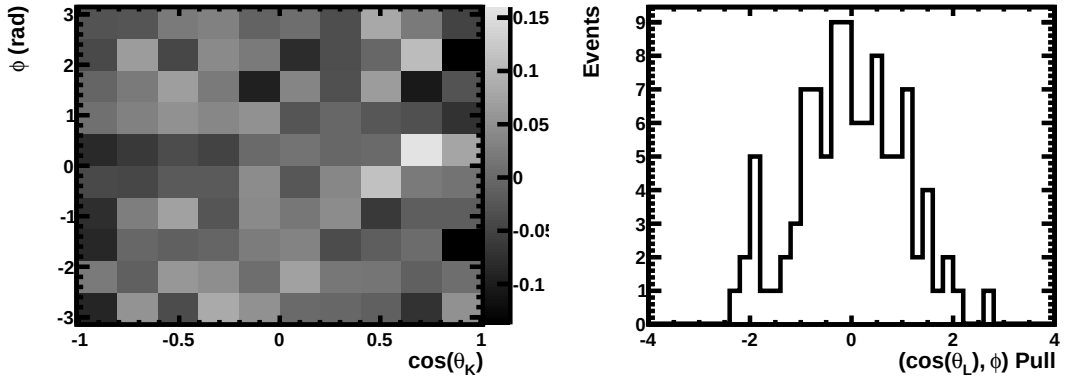
Figure B.5: The three 2D projections of acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $0.50 < q^2 < 0.60\,\mathrm{GeV}^2/c^4$ (left). The pull distribution, formed by calculating the pull in each bin of the corresponding 2D projection (right).



(a) ($\cos\theta_L$, $\cos\theta_K$) and associated pull distribution

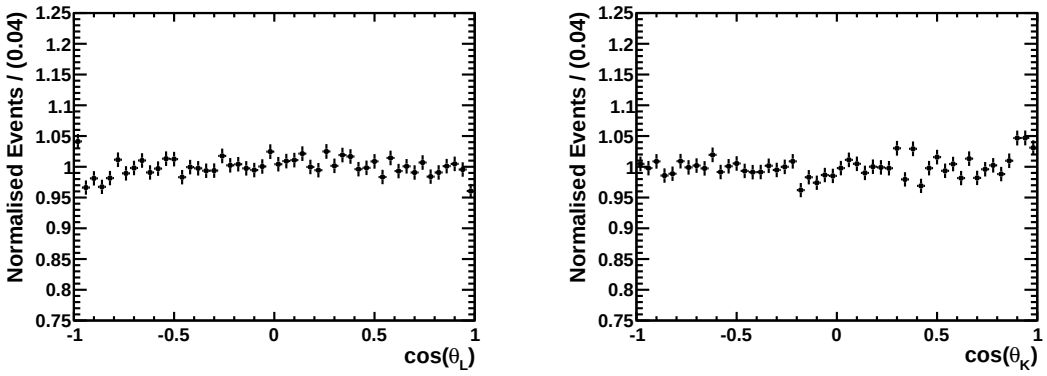


(b) ($\cos\theta_L$, ϕ) and associated pull distribution

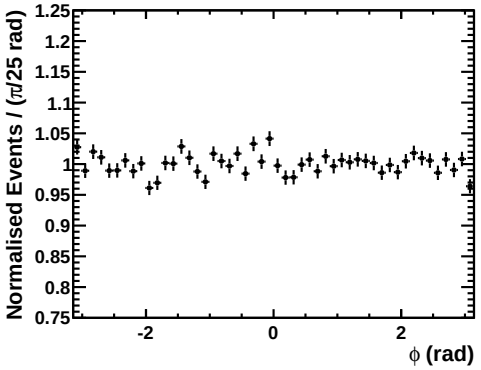


(c) ($\cos\theta_K$, ϕ) and associated pull distribution

Figure B.6: The three 2D projections of acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $12.50 < q^2 < 13.00\,\mathrm{GeV}^2/c^4$ (left). The pull distribution, formed by calculating the pull in each bin of the corresponding 2D projection (right).

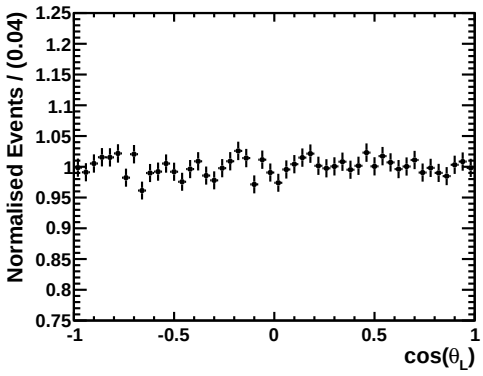


(a) $\cos\theta_L$ (b) $\cos\theta_K$

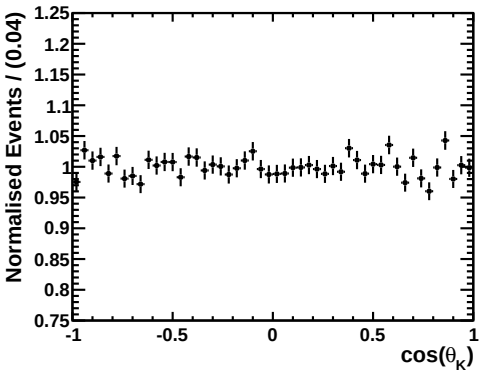


(c) ϕ

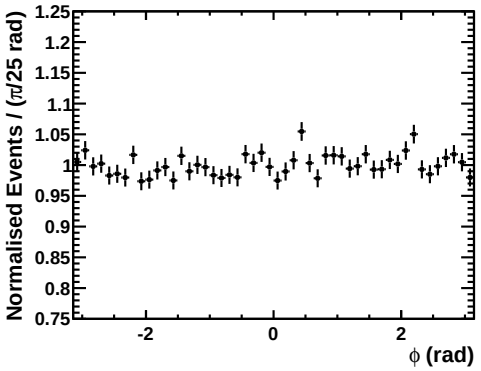
Figure B.7: Distribution of acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$ in the decay angles $\cos\theta_L$ (a), $\cos\theta_K$ (b) and ϕ (c).



(a) $\cos\theta_L$

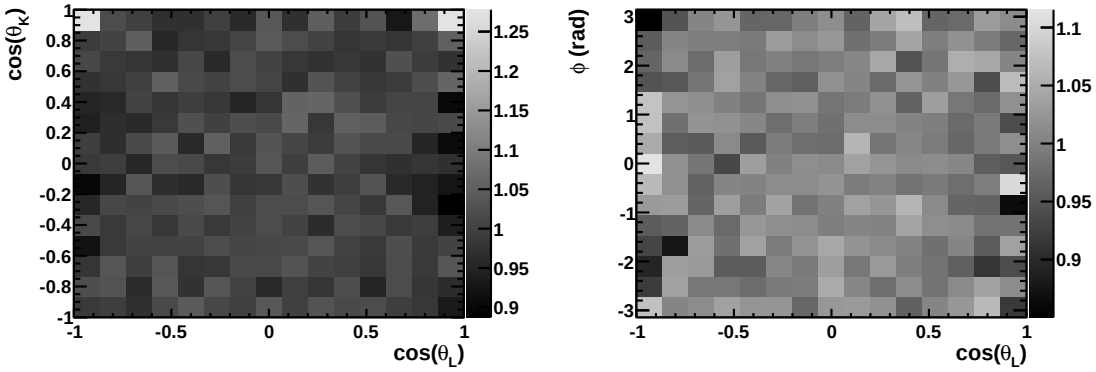


(b) $\cos\theta_K$



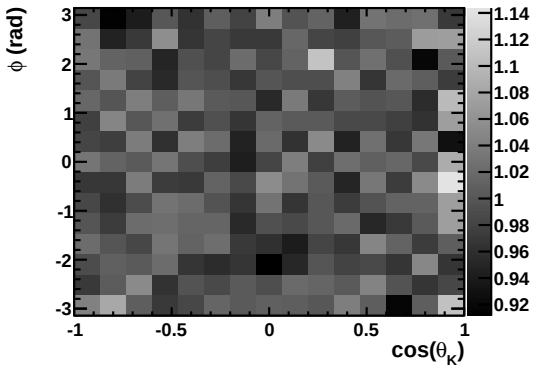
(c) ϕ

Figure B.8: Distribution of acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $14.18 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$ in the decay angles $\cos\theta_L$ (a), $\cos\theta_K$ (b) and ϕ (c).



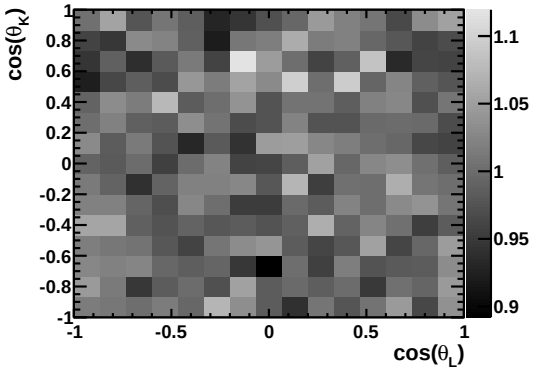
(a) $(\cos\theta_L, \cos\theta_K)$ distribution of events

(b) $(\cos\theta_L, \phi)$ distribution of events

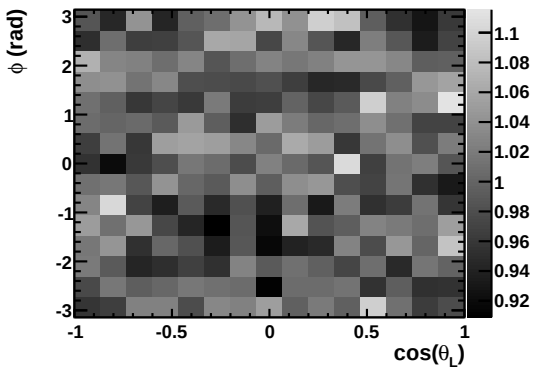


(c) $(\cos\theta_K, \phi)$ distribution of events

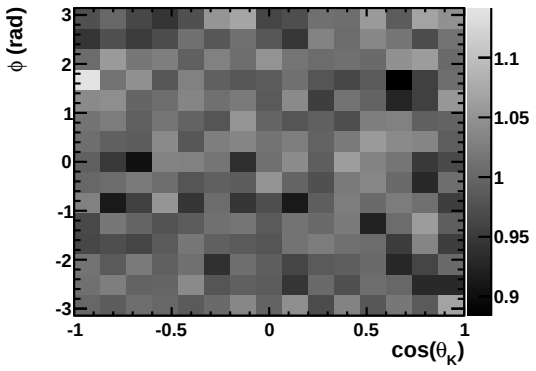
Figure B.9: 2D projection of acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$, for the angle-pairs $(\cos\theta_L,\ \cos\theta_K)$ (a), $(\cos\theta_L,\ \phi)$ (b) and $(\cos\theta_K,\ \phi)$ (c).



(a) $(\cos\theta_L, \cos\theta_K)$ distribution of events



(b) $(\cos\theta_L, \phi)$ distribution of events



(c) $(\cos\theta_K, \phi)$ distribution of events

Figure B.10: 2D projection of acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $14.18 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$, for the angle-pairs $(\cos\theta_L,\ \cos\theta_K)$ (a), $(\cos\theta_L,\ \phi)$ (b) and $(\cos\theta_K,\ \phi)$ (c).

Appendix C

Calculation of the pulls of the acceptance correction

This deviation from a uniform distribution of events indicates that the acceptance correction does not properly account for the acceptance effects present in the decay angles. The impact of these inadequacies on the overall validity of the acceptance correction is quantified by comparing the distribution of simulated events to the uniform distribution expectated for a perfect acceptance correction in regions of the angular space. This comparison is made by calculating the pull in each region of the angular space. Unfortunately, considerably more events are simulated at low q^2 than at high q^2. If the pulls are calculated over all q^2 the effects at low q^2 would overwhelm any effect at high q^2. To overcome this, the pulls are calculated in each $q^2_{\rm AC}$ bin separately. The pulls are subsequently combined to examine the validity of the acceptance correction over all q^2.

The pull is related to the discrepancy between the distribution of simulated events and the distribution expected if the acceptance correction is perfect in a given region (bin) of the angular space. The discrepancy is calculated as

$$\Delta_n = \left(\frac{n}{\bar{n}} - 1\right), \tag{C.0.1}$$

where n is the sum-of-weights of the acceptance-corrected events in a given bin and $\bar{n}$ is the sum-of-weights expected in that bin if the events are uniformly distributed in the angular space. The $\bar{n}$ quantity is calculated as the sum-of-weights of the simulated events divided by the ratio of the size of the bin and the size of the angular space. The pull is then calculated as

$$\text{pull} = \frac{\Delta_n}{\sigma_{\Delta_n}}, \tag{C.0.2}$$

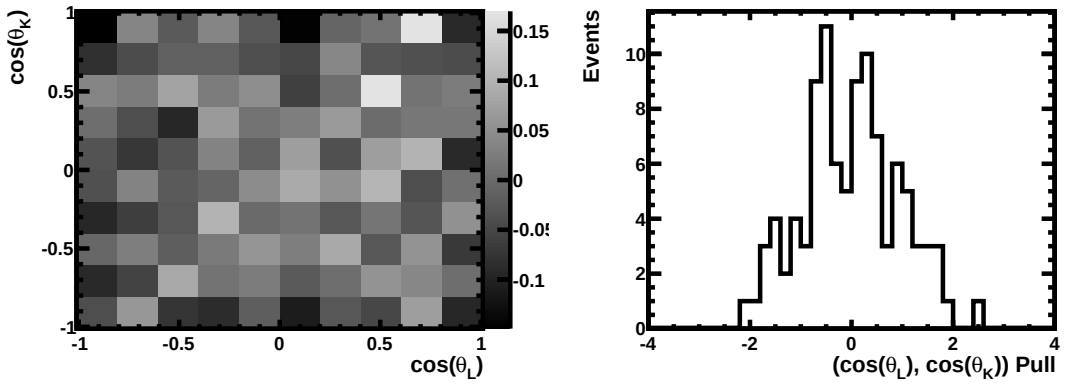
where σ_{Δ_n} is the uncertainty on Δ_n. If the values of Δ_n are due only to statistical fluctuations of n then an ensemble of pulls, each calculated in a distinct region of phase-

space, will be distributed as a Gaussian around zero with a width of unity. Any deviation from this is indicative of an imperfect acceptance correction.

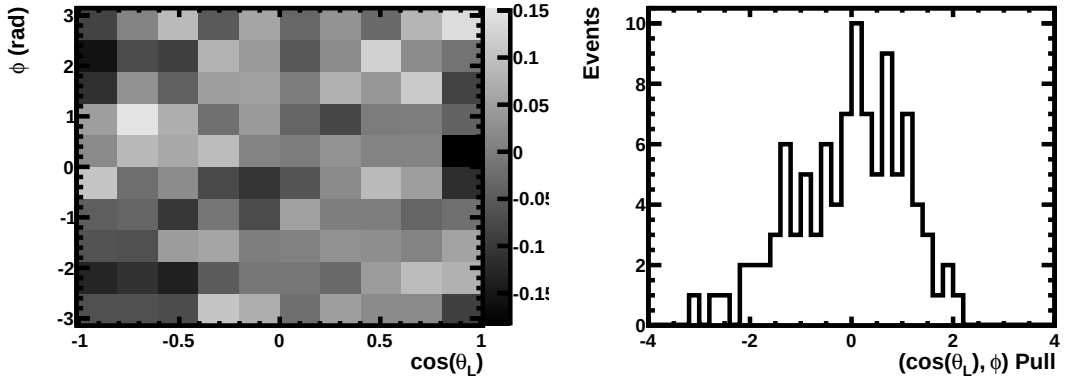
A 3D comparison is not possible due to the limited number of simulated events in each q^2_{AC} bin. Instead, each of the 2D projections of the angular-space ($\cos\theta_L$, $\cos\theta_K$), ($\cos\theta_L$, ϕ) and ($\cos\theta_K$, ϕ) are separated into 100 bins. The distribution of Δ_n in each 2D projection, and the associated pull distribution, is shown for simulated events with $3.40 < q^2 < 3.60\,\mathrm{GeV}^2/c^4$ in Fig. C.1. The same plots for simulated events with $0.50 < q^2 < 0.60\,\mathrm{GeV}^2/c^4$ and with $12.50 < q^2 < 13.00\,\mathrm{GeV}^2/c^4$ are shown in App. B.

To examine the acceptance correction over $0.10 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$, the pull distributions of each q^2_{AC} bin are combined. The results are shown in Fig. C.2. Each of the three distributions is fitted with a Gaussian pdf, which in each case returns $\bar{x} \approx -0.05$ and $\sigma \approx 1.05$. The shift of the mean to negative values is caused by an imperfect acceptance correction resulting in a small number of events with too-large weights. These events cause $\bar{n}$ to be larger than it would be with a perfect acceptance correction. This results in Δ_n, and therefore the distribution of pulls, being shifted to values less than zero in the majority of the bins.

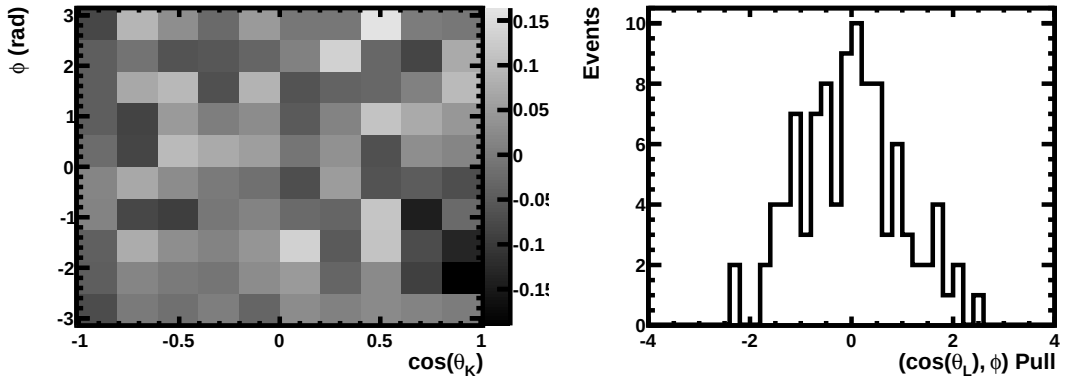
The pull distributions for each projection are combined into the total pull distribution, which is shown in Fig. C.3. A Gaussian fit is indicated, which returns $\bar{x} \approx -0.05$ and $\sigma \approx 1.02$. This indicates a non-uniform distribution at the level of 5%, which is considered as a source of systematic uncertainty in Sec. 6.3.2.



(a) ($\cos\theta_L$, $\cos\theta_K$) and associated pull distribution

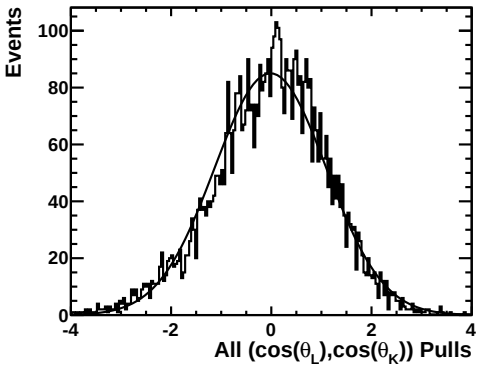


(b) ($\cos\theta_L$, ϕ) and associated pull distribution

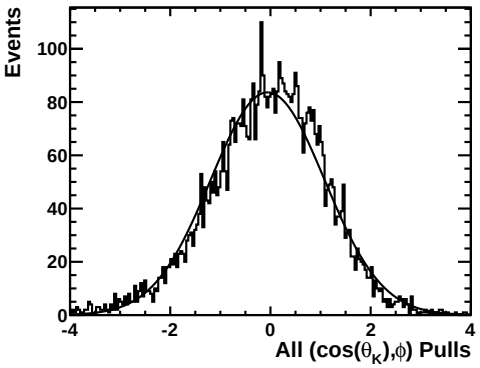


(c) ($\cos\theta_K$, ϕ) and associated pull distribution

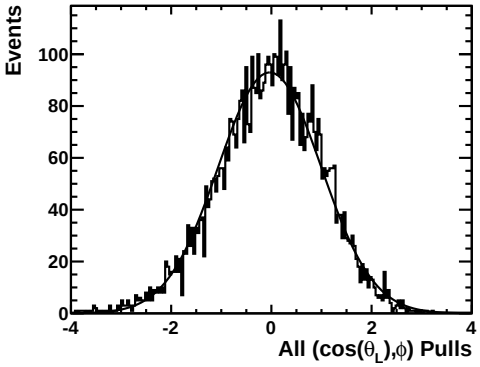
Figure C.1: Binned distribution of Δ_n in 2D projections of the angular space, using acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events with $3.40 < q^2 < 3.60\,\mathrm{GeV}^2/c^4$ (left). The pull distribution, formed by calculating the pull in each bin of the corresponding 2D projection (right). These are shown for the 2D angular projections ($\cos\theta_L$, $\cos\theta_K$) (a); ($\cos\theta_L$, ϕ) (b); and ($\cos\theta_K$, ϕ) (c).



(a) ($\cos\theta_L$, $\cos\theta_K$) total pull distribution



(b) ($\cos\theta_L$, ϕ) total pull distribution



(c) ($\cos\theta_K$, ϕ) total pull distribution

Figure C.2: Total pull distributions of acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events, for the angle-pairs ($\cos\theta_L$, $\cos\theta_K$) (a), ($\cos\theta_L$, ϕ) (b) and ($\cos\theta_K$, ϕ) (c).

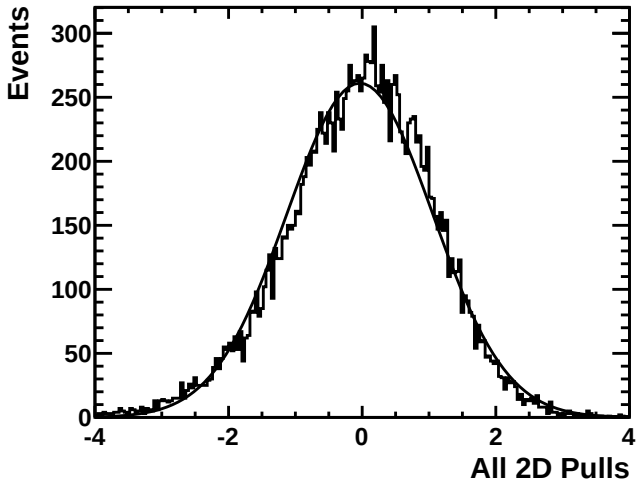


Figure C.3: Total pull distribution of acceptance-corrected simulated $B^0 \to K^{*0}\mu^+\mu^-$ events. A Gaussian fit is overlaid.

Appendix D

Differential Branching Fraction Plots (Linear Scale)

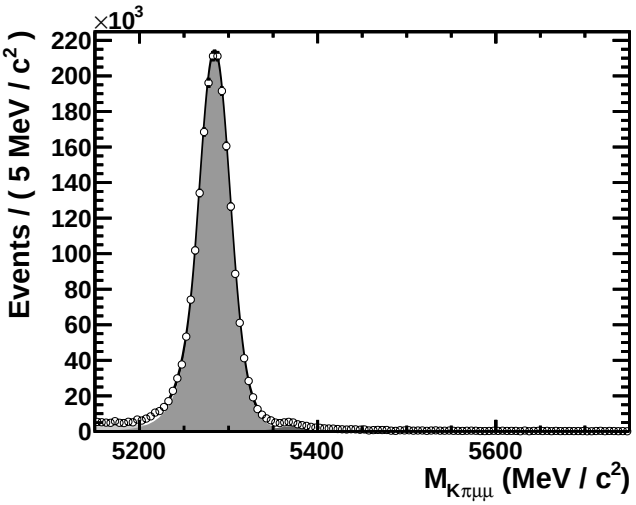
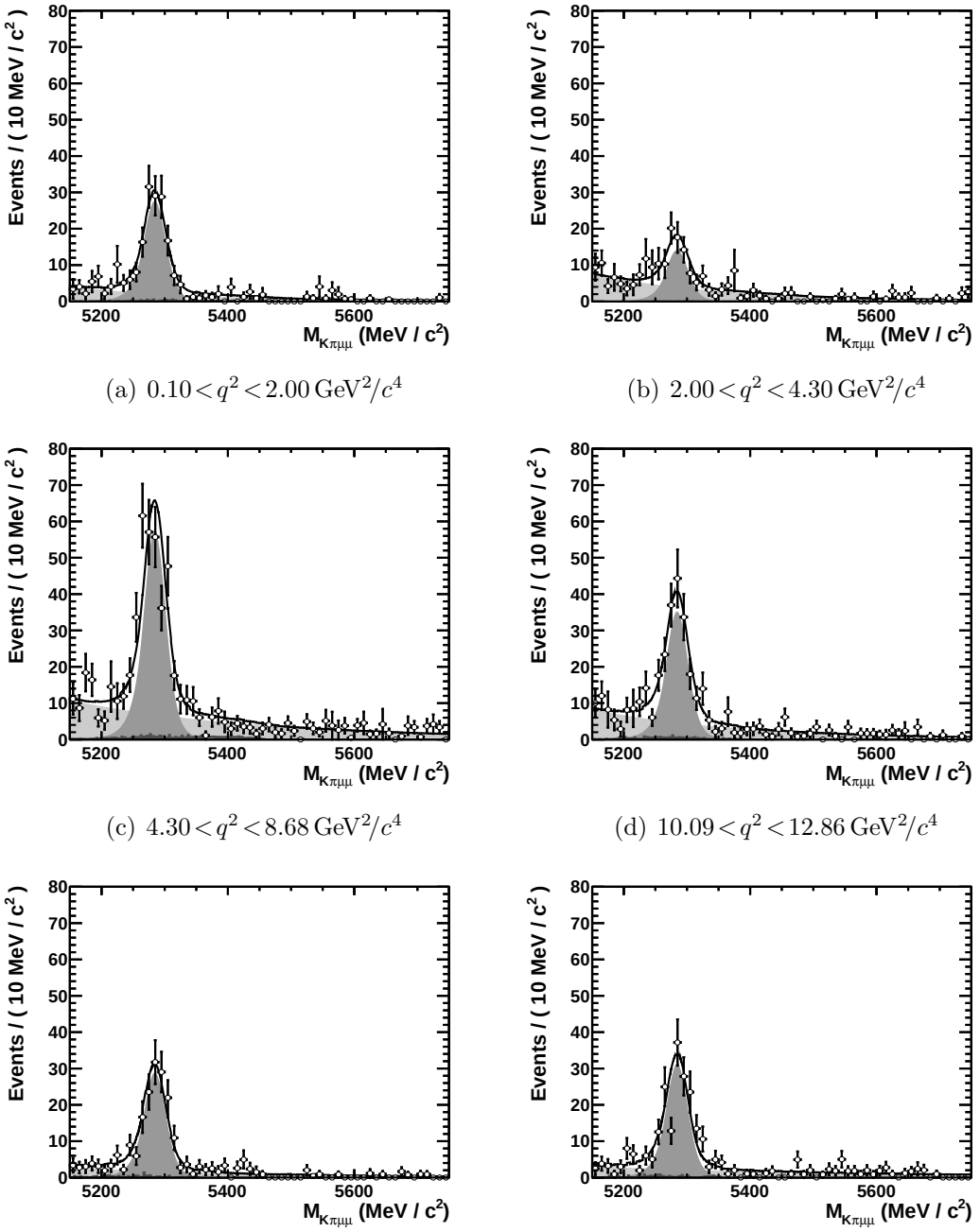


Figure D.1: The $m_{K\pi\mu\mu}$ mass distribution of acceptance-corrected $B^0 \to J/\psi\, K^{*0}$ candidates from the data (black open points) in the mass range $5150 < m_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. A fit (black solid line) to the data is overlaid, with the combinatoric (light grey area), $B^0 \to J/\psi\, K^{*0}$ (grey area) and $\overline{B}{}^0_s \to J/\psi\, K^{*0}$ (dark grey area) components indicated.



(a) $0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$

(b) $2.00 < q^2 < 4.30\,\mathrm{GeV}^2/c^4$

(c) $4.30 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$

(d) $10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$

(e) $14.18 < q^2 < 16.00\,\mathrm{GeV}^2/c^4$

(f) $16.00 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$

Figure D.2: The $\mathrm{m}_{K\pi\mu\mu}$ distribution of acceptance-corrected $B^0 \to K^{*0}\mu^+\mu^-$ candidates from the data (open points) in the mass range $5150 < \mathrm{m}_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$ in each of the analysis q^2 bins. A fit (black solid line) to the data is overlaid, with the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ (white area); Class III background (light grey area); $B^0 \to K^{*0}\mu^+\mu^-$ (grey area); and Class I and Class II background (dark grey histogram) components indicated. The fit model is defined in Sec. 6.3.

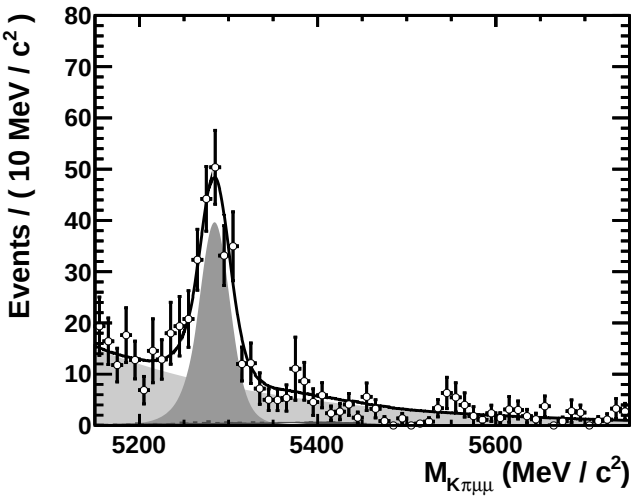


Figure D.3: The $m_{K\pi\mu\mu}$ distribution of acceptance-corrected $B^0 \to K^{*0}\mu^+\mu^-$ candidates from the data with $1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$ (black open points) in the mass range $5150 < m_{K\pi\mu\mu} < 5750\,\mathrm{MeV}/c^2$. A fit (black line) to the data is overlaid, with the $\overline{B}{}^0_s \to K^{*0}\mu^+\mu^-$ (white area); Class III background (light grey area); $B^0 \to K^{*0}\mu^+\mu^-$ (grey area); and Class I and Class II background (dark grey histogram) components indicated. The fit model is defined in Sec. 6.3.

Appendix E

Impact of physical boundaries on the angular fit

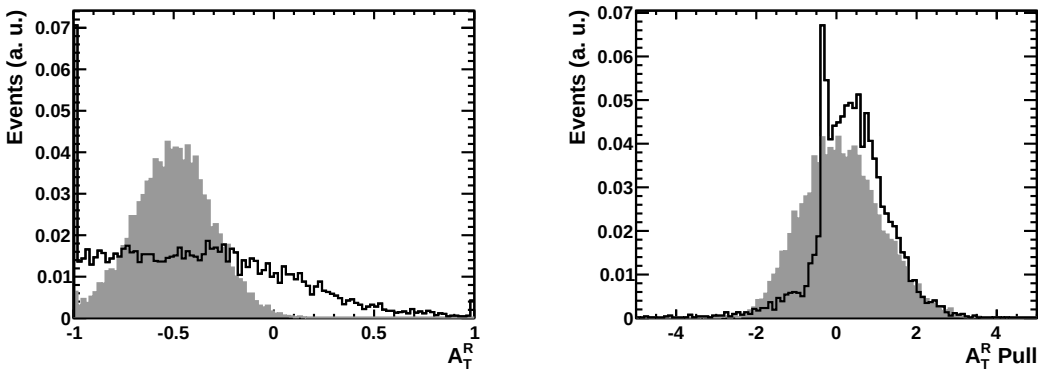
To examine the impact of the physical boundaries on the measurement of the angular observables, the fitted distribution of $A_{\mathrm{T}}^{\mathrm{R}}$ is shown for an ensemble of simulated data sets, each of which correspond to $1\,\mathrm{fb}^{-1}$ of integrated luminosity, as the black histogram in Fig. E.1(a). Constraining the fit as described above truncates the distribution of fitted results, causing the spike of events at $A_{\mathrm{T}}^{\mathrm{R}} = -1$.

The pull of a fitted value of an observable is defined as

$$\text{pull} = \frac{1}{\sigma_\lambda}\left(\lambda - \bar{\lambda}\right) \tag{E.0.1}$$

where λ is the fitted value of an observable, σ_λ is the uncertainty on λ, and $\bar{\lambda}$ is the value of λ used to generate the simulated events. In the absence of bias, the pulls are distributed as a Gaussian centred about zero with a width equal to unity. Any deviation of the mean away from zero is indicative of a bias in the fit. The black histogram in Fig. E.1(b) is not Gaussian distributed and is biased to positive values. This is due to the truncation of events as a result of the physical boundaries, see in Fig. E.1(a).

The number of measurements that are truncated in this way is related to the precision of the $A_{\mathrm{T}}^{\mathrm{R}}$ measurement and therefore to the size of the simulated data sets. The fitted distribution of $A_{\mathrm{T}}^{\mathrm{R}}$, and the associated distribution of pulls, for an ensemble of data sets each of which correspond to $10\,\mathrm{fb}^{-1}$ of integrated luminosity are shown as the grey shaded histogram in Fig. E.1(a) and Fig. E.1(b) respectively. The increased number of candidates improves the measurement precision and reduces the spread of the fitted $A_{\mathrm{T}}^{\mathrm{R}}$ values. This reduces the truncation at the physical boundary, making the distribution of pulls Gaussian distributed and unbiased.



(a) $A_{\mathrm{T}}^{\mathrm{R}}$ (b) $A_{\mathrm{T}}^{\mathrm{R}}$ pull distribution

Figure E.1: Distribution of fitted $A_{\mathrm{T}}^{\mathrm{R}}$ values (a) and the associated pull distribution (b) for two simulated data sets that correspond to $1\,\mathrm{fb}^{-1}$ (black solid histogram) and $10\,\mathrm{fb}^{-1}$ (grey shaded histogram) of integrated luminosity.

Appendix F

The point-to-point dissimilarity test

The DT_{ptp} calculates a p-value for the goodness-of-fit of a model f_0 to a true model f. In this case f is the pdf that defines the distribution of candidates in the data. As f is unknown, it is not possible to compare the two models directly. The DT_{ptp} therefore compares events generated from f_0 to candidates in the data, using the test statistic T.

The events contained in two data sets a and b are compared in regions of a multivariate space $\vec{x}$ using a weighting function Ψ. The weighting function Ψ is defined as

$$\Psi(|\vec{x}_i - \vec{x}_j|) = e^{-|\vec{x}_i - \vec{x}_j|^2 / 2\sigma(\vec{x}_i)\sigma(\vec{x}_j)}, \tag{F.0.1}$$

where $|\vec{x}_i - \vec{x}_j|$ is the n-dimensional distance between event i and event j and is used to measure the dissimilarity of the two events. The $\sigma(\vec{x})$ parameter defines the region of the n-dimensional space covered by Ψ and is defined as

$$\sigma(\vec{x}) = \frac{\bar{\sigma}}{f_0(\vec{x}) \int d\vec{x}'}, \tag{F.0.2}$$

where the mean value of $f_0(\vec{x}) \int d\vec{x}'$ is unity and $\bar{\sigma}$ sets the scale of $\sigma(\vec{x})$. The Ψ function ensures that the dissimilarity of events in similar regions of $\vec{x}$ contribute more to T than that of events in disparate regions of $\vec{x}$. The parameter $\bar{\sigma}$ sets the fraction of $\vec{x}$ that is considered 'similar', which is typically on the order of a few percent [87].

In the following, two data sets are identified by a super-script a or b and contain a number of events n_a and n_b and that occupy the n-dimensional spaces $\vec{x}^a$ and $\vec{x}^b$

respectively. The test statistic T is defined as

$$\begin{aligned} T = & \frac{1}{n_a(n_a-1)} \sum_{i,j>i}^{n_a} \Psi\left(|\vec{x}_i^a - \vec{x}_j^a|\right) \\ + & \frac{1}{n_b(n_b-1)} \sum_{i,j>i}^{n_b} \Psi\left(|\vec{x}_i^b - \vec{x}_j^b|\right) \\ - & \frac{1}{n_a n_b} \sum_{i,j}^{n_a,n_b} \Psi\left(|\vec{x}_i^a - \vec{x}_j^b|\right). \end{aligned} \tag{F.0.3}$$

The first term in Eq. (F.0.3) corresponds to the dissimilarity of the events within data set a and the second to the dissimilarity of the events within data set b. The third term corresponds to the dissimilarity between the events in data set a and those in data set b.

In practice, as will be seen below, n_a is small and n_b is large. The normalisation term for data set a is therefore changed to $1/n_a^2$ as this quantity does not diverge as $n_a \to 1$. As statistical fluctuations on data set b are expected to be small, the second term in Eq. (F.0.3) is dropped. This leaves

$$T = \frac{1}{n_a^2} \sum_{i,j>i}^{n_a} \Psi\left(|\vec{x}_i^a - \vec{x}_j^a|\right) - \frac{1}{n_a n_b} \sum_{i,j}^{n_a,n_b} \Psi\left(|\vec{x}_i^a - \vec{x}_j^b|\right). \tag{F.0.4}$$

To calculate a p-value for the goodness-of-fit of f_0, the candidates in the data are fitted to obtain f_0. A simulated data set is generated using f_0 following the procedure described in Sec. 7.2.1.1. This data set contains five thousand times more events than there are candidates in the data ($n_b \gg n_a$). One thousand data sets, each of which contain the same number of events as there are candidates in the data, are then generated in the same way. Each of these small simulated data sets are compared, using Eq. (F.0.4), to the large simulated data set to obtain an ensemble of T_{sim} values, $\vec{T}_{\text{sim}}$. The distribution of $\vec{T}_{\text{sim}}$ constitutes the distribution of T when $f_0 = f$. The candidates in the data are also compared to the large simulated data set to obtain T_{data}. The p-value is the fraction of $\vec{T}_{\text{sim}}$ with $T_{\text{sim}} > T_{\text{data}}$.

Appendix G

Supplementary plots regarding the angular fit

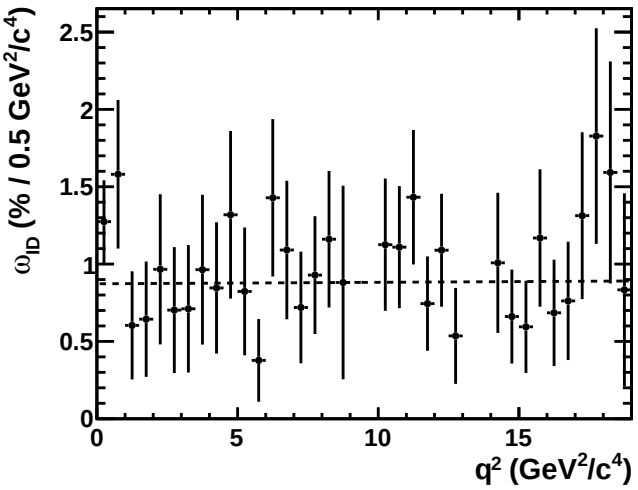


Figure G.1: Variation of ω_{ID} in q^2, with a linear fit overlaid (black dashed line).

Appendix H

Systematic uncertainties of the angular fit

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	0.266	0.781	0.625	0.462	0.330	0.402	0.612
Standard Error	0.001	0.002	0.001	0.001	0.001	0.001	0.001
AC $\cos\theta_L$ Up	-0.010	-0.007	-0.007	-0.007	-0.007	-0.008	-0.008
AC $\cos\theta_L$ Down	0.010	0.007	0.009	0.010	0.009	0.012	0.010
AC $\cos\theta_K$ Up	0.017	0.012	0.016	0.018	0.016	0.018	0.017
AC $\cos\theta_K$ Down	-0.017	-0.013	-0.020	-0.017	-0.017	-0.019	-0.021
AC $\cos\theta_L$ Up $\cos\theta_K$ Up	0.007	0.008	0.011	0.014	0.007	0.012	0.008
AC $\cos\theta_L$ Up $\cos\theta_K$ Down	-0.026	-0.021	-0.028	-0.024	-0.025	-0.027	-0.027
AC $\cos\theta_L$ Down $\cos\theta_K$ Up	0.026	0.018	0.026	0.027	0.022	0.027	0.028
AC $\cos\theta_L$ Down $\cos\theta_K$ Down	-0.007	-0.007	-0.010	-0.009	-0.010	-0.008	-0.007
AC Non-factorisable Up	0.002	0.000	-0.001	0.002	-0.001	-0.000	0.001
AC Non-factorisable Down	0.001	0.000	-0.001	0.000	-0.001	0.001	0.004
AC q^2 binning +1	0.006	-0.003	-0.009	0.000	-0.001	0.009	0.000
AC q^2 binning -1	-0.000	0.008	0.008	-0.009	-0.010	-0.005	0.011
Trigger efficiency Up	-0.001	-0.002	-0.000	-0.000	-0.000	0.001	0.001
Trigger efficiency Down	0.003	0.002	0.001	-0.002	-0.001	0.001	0.004
DLL binnning +1	-0.004	-0.000	-0.003	-0.001	-0.004	0.002	0.003
DLL binning -1	-0.009	-0.002	-0.003	-0.001	-0.001	0.002	0.000
IsMuon efficiency Up	0.011	0.003	-0.000	0.002	-0.000	-0.002	0.006
IsMuon efficiency Down	-0.013	-0.007	-0.004	0.002	0.002	0.003	-0.007
Tracking efficiency Up	0.002	0.001	-0.002	-0.000	-0.001	-0.001	-0.000
Tracking efficiency Down	-0.000	0.001	0.002	0.001	-0.000	0.001	0.002
B^0 p_T re-weighting	-0.001	-0.002	-0.002	-0.003	-0.001	-0.003	0.000
No B^0 p re-weighting	0.000	-0.001	-0.004	-0.002	-0.004	-0.003	0.001
IP Correction	0.019	-0.001	0.003	0.002	-0.006	-0.004	-0.009
Mass resolution Up	0.001	-0.002	0.000	0.003	0.001	-0.000	0.002
Mass resolution Down	0.001	0.005	-0.001	-0.001	0.001	-0.000	0.001
S-wave component	0.009	-0.019	-0.012	0.001	0.009	0.001	-0.009
Order 1 background model	-0.002	0.004	-0.002	0.001	0.000	0.002	0.003
Order 3 background model	0.001	0.002	-0.001	0.001	-0.003	-0.001	0.003
Kaon misidentification Up	-0.000	0.002	-0.002	0.000	0.002	-0.001	-0.000
Kaon misidentification Down	-0.003	0.002	-0.001	0.001	0.001	-0.001	0.001
Class II and III Up	0.011	-0.006	-0.005	0.002	0.007	0.002	-0.002
Class II and III Down	0.010	-0.011	-0.005	0.003	0.008	0.004	-0.004

Table H.1: Impact on the measured value of F_{L} associated with each systematic variation.

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	-0.056	-0.036	0.057	-0.182	-0.094	-0.149	0.027
Standard Error	0.002	0.001	0.001	0.001	0.001	0.001	0.001
AC $\cos\theta_L$ Up	0.003	-0.002	0.001	-0.000	-0.002	-0.002	-0.001
AC $\cos\theta_L$ Down	0.004	-0.002	0.000	0.002	-0.002	-0.003	0.000
AC $\cos\theta_K$ Up	0.005	-0.003	-0.001	0.006	-0.001	0.002	-0.002
AC $\cos\theta_K$ Down	0.001	-0.002	0.004	-0.004	-0.001	-0.004	0.000
AC $\cos\theta_L$ Up $\cos\theta_K$ Up	0.005	-0.001	-0.003	0.005	0.000	-0.001	-0.002
AC $\cos\theta_L$ Up $\cos\theta_K$ Down	0.001	-0.003	0.005	-0.005	-0.004	-0.005	0.001
AC $\cos\theta_L$ Down $\cos\theta_K$ Up	0.003	0.005	-0.001	0.006	0.001	-0.000	-0.002
AC $\cos\theta_L$ Down $\cos\theta_K$ Down	0.001	-0.003	0.003	-0.000	-0.003	-0.003	-0.001
AC Non-factorisable Up	0.002	-0.003	0.000	0.001	-0.002	-0.003	-0.002
AC Non-factorisable Down	0.003	0.002	-0.000	0.003	-0.001	-0.004	-0.001
AC q^2 binning +1	-0.000	-0.002	-0.000	0.000	0.005	-0.003	-0.002
AC q^2 binning -1	0.000	-0.002	-0.001	-0.007	0.005	-0.002	-0.003
Trigger efficiency Up	0.001	0.000	0.001	-0.001	-0.004	-0.002	-0.001
Trigger efficiency Down	0.003	-0.001	0.002	0.001	-0.003	-0.001	0.003
DLL binnning +1	0.003	-0.005	0.001	-0.000	-0.006	-0.002	0.000
DLL binning -1	0.003	-0.004	0.001	-0.000	-0.003	-0.002	0.000
IsMuon efficiency Up	0.005	-0.001	0.001	0.001	0.001	-0.000	0.001
IsMuon efficiency Down	0.002	-0.001	-0.001	-0.001	0.000	-0.006	-0.001
Tracking efficiency Up	0.002	-0.001	0.001	0.000	-0.002	-0.003	-0.001
Tracking efficiency Down	0.004	-0.002	0.001	0.002	-0.001	-0.001	-0.001
B^0 p_T re-weighting	0.004	-0.002	0.002	-0.001	-0.000	0.001	-0.001
No B^0 p re-weighting	0.004	-0.002	0.002	0.001	-0.001	-0.004	-0.001
IP Correction	0.003	-0.002	0.000	0.003	0.003	-0.010	0.005
Mass resolution Up	0.003	-0.001	-0.000	0.001	-0.001	-0.003	0.002
Mass resolution Down	0.001	-0.000	0.001	0.002	0.002	0.001	0.000
S-wave component	0.007	0.005	-0.005	0.014	0.007	0.014	-0.002
Order 1 background model	0.003	-0.001	0.001	0.001	-0.001	-0.002	0.001
Order 3 background model	0.001	-0.002	0.000	0.001	-0.001	-0.002	-0.000
Kaon misidentification Up	0.001	0.000	0.003	0.002	-0.001	0.000	-0.002
Kaon misidentification Down	0.005	0.001	0.000	0.001	-0.003	-0.001	0.001
Class II and III Up	0.008	0.003	0.002	0.010	0.004	0.007	-0.000
Class II and III Down	0.003	-0.002	-0.001	0.006	-0.001	0.002	-0.001

Table H.2: Impact on the measured value of S_3 associated with each systematic variation.

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	-0.037	-0.016	0.009	-0.043	-0.004	-0.030	0.065
Standard Error	0.002	0.001	0.001	0.001	0.001	0.001	0.001
AC $\cos\theta_L$ Up	0.001	0.001	-0.000	0.000	-0.004	0.001	-0.001
AC $\cos\theta_L$ Down	-0.002	0.001	-0.001	0.001	-0.003	0.001	-0.002
AC $\cos\theta_K$ Up	0.003	0.003	0.000	0.005	-0.003	0.001	-0.002
AC $\cos\theta_K$ Down	-0.002	-0.000	0.001	-0.001	-0.003	-0.002	0.003
AC $\cos\theta_L$ Up $\cos\theta_K$ Up	0.001	-0.002	-0.001	0.003	-0.004	0.004	-0.002
AC $\cos\theta_L$ Up $\cos\theta_K$ Down	0.000	-0.000	0.002	0.003	0.000	-0.000	0.002
AC $\cos\theta_L$ Down $\cos\theta_K$ Up	0.002	0.002	0.002	0.003	-0.001	-0.000	-0.003
AC $\cos\theta_L$ Down $\cos\theta_K$ Down	-0.002	0.000	0.001	0.000	-0.006	-0.001	0.002
AC Non-factorisable Up	0.001	0.002	0.000	0.002	-0.002	-0.001	-0.000
AC Non-factorisable Down	0.002	0.001	0.001	-0.000	-0.004	-0.003	0.001
AC q^2 binning +1	-0.000	-0.000	-0.002	-0.001	-0.001	-0.001	0.000
AC q^2 binning -1	-0.000	-0.000	0.001	0.001	-0.001	-0.003	-0.001
Trigger efficiency Up	-0.002	0.002	0.000	0.003	-0.003	0.001	0.002
Trigger efficiency Down	0.001	0.001	0.001	0.000	-0.002	-0.003	-0.002
DLL binnning +1	0.001	-0.002	0.001	0.001	-0.004	0.002	0.001
DLL binning -1	0.000	-0.001	0.001	-0.000	-0.002	-0.001	-0.001
IsMuon efficiency Up	0.000	0.003	0.000	0.002	-0.001	0.002	-0.003
IsMuon efficiency Down	-0.001	-0.001	-0.000	0.001	-0.005	0.000	0.002
Tracking efficiency Up	0.001	0.001	-0.001	0.000	-0.002	-0.003	-0.000
Tracking efficiency Down	-0.002	0.001	-0.001	0.001	-0.000	-0.000	-0.001
B^0 p_T re-weighting	-0.000	-0.001	0.000	0.002	-0.003	-0.001	-0.001
No B^0 p re-weighting	0.000	0.001	-0.001	0.001	-0.005	0.002	0.000
IP Correction	-0.001	0.001	-0.001	-0.000	-0.001	0.000	0.000
Mass resolution Up	-0.003	0.004	-0.002	0.001	-0.003	0.001	-0.000
Mass resolution Down	0.002	-0.003	0.000	0.001	-0.003	-0.001	0.000
S-wave component	0.004	-0.001	-0.000	0.006	0.000	0.004	-0.007
Order 1 background model	-0.000	0.001	0.002	0.002	0.000	-0.001	0.000
Order 3 background model	-0.001	0.003	-0.001	-0.001	-0.003	-0.000	0.001
Kaon misidentification Up	-0.000	0.002	-0.001	0.002	-0.004	0.002	0.001
Kaon misidentification Down	-0.002	0.002	-0.002	0.003	-0.002	-0.001	0.000
Class II and III Up	0.002	0.003	0.001	0.005	-0.001	0.005	-0.002
Class II and III Down	0.001	-0.002	-0.004	-0.000	-0.004	-0.004	-0.005

Table H.3: Impact on the measured value of S_9 associated with each systematic variation.

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	0.116	0.045	-0.113	0.064	-0.071	0.058	-0.005
Standard Error	0.002	0.001	0.001	0.001	0.001	0.001	0.001
AC $\cos\theta_L$ Up	0.001	-0.003	-0.001	0.000	-0.001	0.001	-0.004
AC $\cos\theta_L$ Down	0.002	0.000	-0.000	-0.002	-0.001	0.002	-0.003
AC $\cos\theta_K$ Up	0.000	-0.003	0.002	-0.001	-0.000	0.003	-0.005
AC $\cos\theta_K$ Down	0.003	-0.001	-0.006	0.002	-0.000	0.003	-0.004
AC $\cos\theta_L$ Up $\cos\theta_K$ Up	-0.000	-0.001	0.001	-0.003	-0.003	0.000	-0.002
AC $\cos\theta_L$ Up $\cos\theta_K$ Down	0.002	0.001	-0.008	0.002	-0.003	0.004	-0.003
AC $\cos\theta_L$ Down $\cos\theta_K$ Up	0.000	-0.004	0.004	-0.005	0.003	-0.001	-0.001
AC $\cos\theta_L$ Down $\cos\theta_K$ Down	0.004	0.000	-0.005	0.001	-0.000	0.000	-0.002
AC Non-factorisable Up	-0.001	-0.002	0.000	-0.003	0.001	0.002	-0.003
AC Non-factorisable Down	0.004	-0.001	-0.001	-0.002	0.001	0.004	-0.002
AC q^2 binning +1	-0.001	-0.003	-0.002	-0.002	0.000	0.002	-0.001
AC q^2 binning -1	0.005	0.000	0.001	-0.003	0.000	0.002	-0.003
Trigger efficiency Up	0.004	-0.003	-0.001	-0.000	-0.002	0.000	-0.004
Trigger efficiency Down	0.002	-0.001	-0.000	-0.001	-0.001	-0.000	-0.004
DLL binnning +1	0.000	-0.002	-0.002	-0.002	0.001	0.002	-0.002
DLL binning -1	-0.000	-0.001	-0.003	-0.000	0.000	0.001	-0.003
IsMuon efficiency Up	0.002	-0.001	0.001	-0.002	-0.002	0.001	-0.003
IsMuon efficiency Down	0.004	-0.001	-0.002	0.000	-0.001	0.002	-0.003
Tracking efficiency Up	0.004	-0.002	-0.002	-0.003	-0.001	-0.000	-0.001
Tracking efficiency Down	0.001	-0.003	-0.002	-0.002	0.000	0.003	-0.002
B^0 p_T re-weighting	0.002	-0.003	-0.001	-0.002	-0.001	-0.000	-0.005
No B^0 p re-weighting	0.004	-0.001	-0.002	-0.001	0.002	0.002	-0.002
IP Correction	-0.001	0.001	-0.001	-0.000	-0.001	0.001	0.000
Mass resolution Up	-0.001	0.000	-0.001	-0.002	-0.001	0.002	-0.002
Mass resolution Down	0.002	0.001	-0.002	0.002	0.000	0.004	-0.003
S-wave component	-0.010	-0.006	0.010	-0.006	0.003	-0.002	-0.002
Order 1 background model	-0.000	-0.002	-0.001	-0.001	-0.001	0.003	-0.003
Order 3 background model	-0.000	0.000	-0.000	0.000	-0.003	-0.000	-0.004
Kaon misidentification Up	0.001	0.000	-0.003	-0.003	-0.001	0.003	-0.004
Kaon misidentification Down	0.001	-0.003	-0.001	-0.000	-0.002	-0.001	-0.003
Class II and III Up	-0.001	-0.002	0.005	-0.001	0.004	0.001	0.001
Class II and III Down	-0.005	-0.004	0.000	-0.002	0.000	-0.003	-0.006

Table H.4: Impact on the measured value of A_9 associated with each systematic variation.

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	-0.206	-0.539	0.617	0.688	0.909	0.622	-0.297
Standard Error	0.003	0.007	0.002	0.002	0.001	0.002	0.004
AC $\cos\theta_L$ Up	0.009	-0.016	-0.008	0.000	-0.002	0.002	0.005
AC $\cos\theta_L$ Down	-0.001	-0.015	0.000	-0.004	0.000	0.002	0.005
AC $\cos\theta_K$ Up	0.001	-0.006	0.005	0.001	0.005	0.004	0.003
AC $\cos\theta_K$ Down	0.010	-0.014	-0.008	-0.011	-0.004	-0.008	0.003
AC $\cos\theta_L$ Up $\cos\theta_K$ Up	0.005	-0.010	-0.001	0.004	0.002	0.001	0.001
AC $\cos\theta_L$ Up $\cos\theta_K$ Down	0.004	-0.006	-0.012	-0.007	-0.003	-0.008	0.010
AC $\cos\theta_L$ Down $\cos\theta_K$ Up	0.003	-0.005	0.008	-0.001	0.000	0.003	0.000
AC $\cos\theta_L$ Down $\cos\theta_K$ Down	0.001	0.002	-0.013	-0.006	-0.003	-0.003	0.005
AC Non-factorisable Up	-0.005	-0.016	-0.005	0.001	0.001	-0.003	0.002
AC Non-factorisable Down	0.004	-0.004	-0.009	-0.005	0.003	0.001	-0.002
AC q^2 binning +1	0.009	-0.011	-0.007	-0.004	0.001	0.000	0.005
AC q^2 binning -1	0.004	-0.011	-0.001	-0.006	0.001	-0.003	0.004
Trigger efficiency Up	0.003	-0.000	-0.004	-0.004	-0.000	-0.001	-0.003
Trigger efficiency Down	0.005	-0.001	-0.000	0.001	-0.002	0.001	0.007
DLL binnning +1	0.007	-0.010	-0.002	-0.006	-0.001	0.001	0.006
DLL binning -1	-0.001	-0.002	-0.003	-0.000	-0.001	0.002	0.002
IsMuon efficiency Up	-0.005	-0.019	-0.003	-0.003	-0.000	0.003	-0.001
IsMuon efficiency Down	0.001	-0.016	-0.002	-0.005	-0.001	-0.000	0.005
Tracking efficiency Up	0.003	-0.009	-0.007	0.003	-0.000	-0.002	0.002
Tracking efficiency Down	0.002	-0.007	-0.008	-0.001	0.004	-0.003	-0.001
B^0 p_T re-weighting	0.007	0.002	-0.007	-0.002	-0.001	-0.005	-0.000
No B^0 p re-weighting	0.008	-0.017	-0.000	0.000	-0.000	-0.005	0.005
IP Correction	0.004	0.004	0.001	0.002	0.001	-0.003	0.001
Mass resolution Up	0.007	-0.015	-0.006	-0.001	0.001	-0.001	-0.000
Mass resolution Down	0.000	-0.006	-0.005	-0.004	0.001	0.002	0.008
S-wave component	0.011	0.043	-0.049	-0.027	-0.023	-0.031	0.018
Order 1 background model	0.004	-0.000	-0.003	0.001	0.003	0.002	-0.005
Order 3 background model	-0.003	-0.016	-0.003	0.001	-0.002	-0.005	-0.001
Kaon misidentification Up	0.001	-0.016	-0.002	0.000	0.004	-0.001	0.001
Kaon misidentification Down	0.009	-0.006	-0.005	-0.003	-0.000	-0.005	0.002
Class II and III Up	0.024	0.048	-0.018	-0.009	-0.012	-0.010	0.038
Class II and III Down	-0.008	0.005	-0.064	-0.038	-0.035	-0.035	-0.007

Table H.5: Impact on the measured value of A_{T}^{R} associated with each systematic variation.

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	-0.148	-0.254	0.307	-0.657	-0.287	-0.506	0.149
Standard Error	0.004	0.010	0.005	0.004	0.004	0.004	0.007
AC $\cos\theta_L$ Up	0.001	0.001	-0.003	0.005	0.005	0.006	-0.012
AC $\cos\theta_L$ Down	0.006	0.006	0.002	-0.008	-0.001	0.004	0.006
AC $\cos\theta_K$ Up	0.003	0.004	-0.003	-0.002	0.006	-0.005	-0.006
AC $\cos\theta_K$ Down	0.002	-0.001	-0.005	0.002	0.006	0.000	-0.015
AC $\cos\theta_L$ Up $\cos\theta_K$ Up	0.008	0.012	-0.008	-0.002	-0.001	-0.006	-0.010
AC $\cos\theta_L$ Up $\cos\theta_K$ Down	0.000	0.015	-0.003	0.001	0.018	0.010	-0.015
AC $\cos\theta_L$ Down $\cos\theta_K$ Up	0.003	-0.020	-0.004	-0.016	0.002	-0.007	-0.005
AC $\cos\theta_L$ Down $\cos\theta_K$ Down	0.000	-0.008	0.004	-0.003	0.001	0.001	-0.011
AC Non-factorisable Up	-0.000	0.008	-0.000	-0.001	0.001	0.004	-0.015
AC Non-factorisable Down	0.006	0.009	0.010	-0.008	0.009	0.008	-0.015
AC q^2 binning +1	-0.010	-0.008	-0.017	-0.000	0.010	-0.016	-0.018
AC q^2 binning -1	-0.007	-0.027	-0.010	-0.008	0.028	-0.007	-0.026
Trigger efficiency Up	0.005	-0.006	-0.003	-0.003	0.004	-0.002	-0.004
Trigger efficiency Down	0.004	0.010	0.003	-0.004	0.003	0.005	0.007
DLL binnning +1	0.003	0.004	0.005	0.001	0.003	0.002	0.005
DLL binning -1	-0.001	-0.001	-0.003	-0.002	-0.001	0.007	0.002
IsMuon efficiency Up	0.003	0.011	0.008	-0.001	0.006	0.002	-0.001
IsMuon efficiency Down	-0.002	-0.001	-0.000	-0.002	-0.009	-0.002	-0.023
Tracking efficiency Up	-0.005	0.003	-0.004	0.003	0.006	-0.001	-0.021
Tracking efficiency Down	0.006	0.008	0.005	-0.001	0.001	-0.005	-0.010
B^0 p_T re-weighting	0.004	0.018	0.004	-0.004	0.003	-0.002	-0.011
No B^0 p re-weighting	-0.001	-0.011	0.009	-0.002	0.005	-0.001	-0.009
IP Correction	0.002	-0.005	0.004	0.008	0.013	-0.031	0.024
Mass resolution Up	0.008	0.022	0.008	-0.003	0.001	0.005	0.001
Mass resolution Down	0.002	0.003	-0.001	-0.004	0.001	0.002	-0.016
S-wave component	0.010	0.034	-0.038	0.048	0.032	0.055	-0.013
Order 1 background model	0.002	-0.007	-0.002	-0.006	0.000	0.002	-0.008
Order 3 background model	0.001	0.005	-0.008	-0.000	0.002	0.005	-0.003
Kaon misidentification Up	0.002	0.002	-0.002	0.001	0.007	-0.001	-0.011
Kaon misidentification Down	0.003	-0.006	0.000	0.003	0.004	0.001	0.002
Class II and III Up	0.009	0.027	-0.011	0.017	0.014	0.020	-0.003
Class II and III Down	0.000	0.014	-0.009	0.009	0.010	0.015	-0.026

Table H.6: Impact on the measured value of A^2_{T} associated with each systematic variation.

Systematic source	q^2 bin						
	$q^2_{\text{ana}\,1}$	$q^2_{\text{ana}\,2}$	$q^2_{\text{ana}\,3}$	$q^2_{\text{ana}\,4}$	$q^2_{\text{ana}\,5}$	$q^2_{\text{ana}\,6}$	$q^2_{\text{ana}\,T}$
Nominal	-0.106	-0.113	0.046	-0.156	-0.014	-0.103	0.346
Standard Error	0.004	0.010	0.005	0.005	0.004	0.005	0.007
AC $\cos\theta_L$ Up	0.004	0.016	-0.005	0.005	0.001	0.004	-0.012
AC $\cos\theta_L$ Down	-0.002	0.006	-0.001	0.000	-0.007	0.002	0.002
AC $\cos\theta_K$ Up	0.014	0.006	0.006	0.002	0.002	0.008	-0.011
AC $\cos\theta_K$ Down	0.010	-0.003	0.004	0.006	-0.001	0.013	0.006
AC $\cos\theta_L$ Up $\cos\theta_K$ Up	0.004	-0.008	0.001	0.011	0.006	0.005	-0.013
AC $\cos\theta_L$ Up $\cos\theta_K$ Down	0.005	-0.008	-0.001	-0.001	0.005	0.007	-0.014
AC $\cos\theta_L$ Down $\cos\theta_K$ Up	0.001	-0.008	0.005	0.004	-0.010	0.012	0.003
AC $\cos\theta_L$ Down $\cos\theta_K$ Down	-0.000	-0.003	-0.001	0.009	-0.004	0.003	-0.003
AC Non-factorisable Up	0.012	-0.010	-0.006	0.013	0.003	0.000	0.005
AC Non-factorisable Down	0.002	-0.005	0.010	0.003	-0.001	-0.003	-0.001
AC q^2 binning +1	0.002	-0.008	0.016	0.008	-0.004	-0.002	-0.003
AC q^2 binning -1	-0.001	-0.010	0.010	0.003	0.002	0.004	-0.011
Trigger efficiency Up	-0.003	0.002	0.004	0.001	0.001	0.006	-0.008
Trigger efficiency Down	0.004	-0.001	-0.004	-0.004	0.003	0.001	-0.006
DLL binnning +1	0.002	-0.002	-0.003	-0.006	0.002	0.013	0.000
DLL binning -1	0.006	-0.013	0.005	0.007	-0.004	-0.006	-0.002
IsMuon efficiency Up	0.010	0.000	0.001	0.009	-0.000	0.002	-0.008
IsMuon efficiency Down	0.002	0.004	-0.005	0.000	-0.001	0.000	-0.006
Tracking efficiency Up	0.006	0.015	0.005	0.001	0.002	-0.000	0.000
Tracking efficiency Down	0.002	0.031	0.003	0.006	-0.001	0.002	-0.016
B^0 p_T re-weighting	-0.001	-0.005	0.009	0.006	-0.006	0.008	-0.009
No B^0 p re-weighting	0.002	-0.034	0.007	0.004	-0.001	-0.006	0.001
IP Correction	0.007	0.001	-0.002	0.004	0.004	-0.001	0.002
Mass resolution Up	-0.001	0.012	0.015	-0.002	-0.002	0.001	-0.010
Mass resolution Down	0.004	0.022	0.007	-0.000	-0.001	0.001	-0.007
S-wave component	0.016	0.013	-0.008	0.019	-0.001	0.016	-0.049
Order 1 background model	0.002	0.003	-0.001	0.005	0.005	0.008	-0.019
Order 3 background model	0.003	0.008	0.002	0.004	-0.001	0.006	-0.013
Kaon misidentification Up	0.002	0.018	-0.006	0.002	-0.002	0.001	-0.006
Kaon misidentification Down	0.002	-0.016	0.013	0.001	0.005	0.008	-0.004
Class II and III Up	0.014	-0.008	0.010	0.013	0.000	0.008	-0.023
Class II and III Down	0.015	0.014	-0.003	0.009	-0.001	-0.005	-0.035

Table H.7: Impact on the measured value of $A_{\mathrm{T}}^{\mathrm{I}}$ associated with each systematic variation.

Appendix I

Copyright Statement

Confirmation Number: 11096609
Order Date: 06/03/2013

Customer Information

Customer: Christopher Parkinson
Account Number: 3000663189
Organization: Christopher Parkinson
Email: c.parkinson09@imperial.ac.uk
Phone: +44 7582982594
Payment Method: Invoice

Order Details

Journal of high energy physics : JHEP : a refereed journal written, run, and distributed by electronic means

Billing Status: **N/A**

Order detail ID:	63718225
Article Title:	Cornering new physics in b → s transitions
Author(s):	Altmannshofer, Wolfgang ; Straub, David M.
DOI:	10.1007/JHEP08(2012)121
Date:	Aug 01, 2012
ISSN:	1029-8479
Publication Type:	e-Journal
Volume:	2012
Issue:	8
Start page:	1
Publisher:	SISSA
Author/Editor:	International School for Advanced Studies (Trieste, Italy) ; Institute of Physics (Great Britain)

Permission Status: **Granted**
Permission type: Republish or display content
Type of use: use in a thesis/dissertation
Order License Id: 3161371405707

Portion	Full text
Number of copies	1
Author of this Springer article	No
Title of your thesis / dissertation	The angular analysis of the B0 to K*0μ+μ- decay at LHCb
Expected completion date	Jun 2013
Estimated size(pages)	150

Note: This item was invoiced separately through our **RightsLink service.** More info **$ 0.00**

Total order items: 1 **Order Total: $0.00**