LHCb-ANA-2012-051 v3r2
February 15, 2013

Angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at LHCb with $1\,\mathrm{fb}^{-1}$

E. Ben-Haim[1], T. Blake[2], M. De Cian[3], U. Egede[4], J. Lefrancois[5], C. Parkinson[4], M. Patel[4], F. Polci[1], C. Salzmann[3], M. H. Schune[5], J. Serrano[6], N. Serra[3], A. Shires[4]

[1] *LPNHE, Université Pierre et Marie Curie, Université Paris Diderot, CNRS/IN2P3, Paris, France*
[2] *CERN, Geneva, Switzerland*
[3] *Universität Zürich, Switzerland*
[4] *Imperial College, London, UK*
[5] *Universite de Paris-Sud 11, France*
[6] *Universite d'Aix - Marseille, France*

Abstract

The angular distribution and differential branching fraction of the $B^0 \to K^{*0}\mu^+\mu^-$ decay are studied using a data sample, collected by the LHCb experiment, that corresponds to an integrated luminosity of $1\,\mathrm{fb}^{-1}$. A first measurement of the zero-crossing point of the forward-backward asymmetry of the dimuon system is also presented.

Contents

Changes since v2rX:

- There have been a large number of cosmetic changes made in this draft of the ANA note: the zero crossing point measurement has been moved after the angular analysis results; the discussion of the S-wave and the threshold terms has been moved after the angular analysis results. There have also been changes made to the text in several places to hopefully improve the readability of the document.

- A bug has been found and corrected in the estimation of the zero-crossing point of A_{FB}. This results in a small change in the zero-crossing point, changing the value of the crossing point from $5.0\,^{+0.9}_{-1.4}$ to 4.9 ± 0.9.

  The bug related to the use of weighted datasets in RooFit. It was discovered that when cloning a weighted dataset, information about the weights was lost (even though the dataset still had a flag set to say that it was weighted). Without the weights applied the forward backward asymmetry is reduced, reducing the gradient of A_{FB} in the region around the zero-crossing point and increasing the error on q_0^2. As expected, the value of q_0^2 itself is almost unchanged by turning on/off the weights to correct for the acceptance correction. The effect is largest for low q^2 where the acceptance effects in $\cos\theta_\ell$ can be large.

- A p-value of the data with respect to the SM hypothesis has been calculated for the q^2 bins using toy pseudo-experiments (Sec. 15.5).

- The systematic uncertainties on the angular observables have been re-evaluated using toy-experiments (Sec. 18.13).

- A summary of the final results has been added.

Changes since v3r0:

- Two problems were spotted with the systematic Tables. 57-65 in Appendix H:

  1. There was a problem identified with the systematic associated to the B p_T re-weighting (due to a broken ROOT ntuple). The large systematic uncertainty that (mistakenly) appeared in the `v3r0` has been reduced to a negligible level.
  2. Two bugs were also identified in the script that makes the tables. The first bug resulted in the systematic uncertainties being assigned with the wrong sign. The second bug resulted in the

sign and magnitude of some of the systematic uncertainties being assigned the wrong value. The overall impact of the two bugs does not significantly change the conclusions that we drew from Appendix H.

- The text describing the systematic uncertainties has also been updated in an attempt to make the description more complete.

Changes since v3r1:

- A true p-value test has been added in Sec. 15.5. This test is based on the point-to-point dissimilarity method described in Ref. [1].
- The differential branching fraction description has been re-written.
- A key has been added linking the description of the systematic uncertainties in the text to the tables of numbers in Secs. 7.5 and 19.0.1.
- Sec. 9.3 has been added, showing the signal angular resolution obtained using simulated events.
- A short paragraph explaining the differences between the zero crossing result in this ANA note and the preliminary result (in which the bug described above was present) gas been added.

1 Introduction

This analysis note describes the angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ with $1\,\mathrm{fb}^{-1}$ of integrated luminosity collected by the LHCb experiment in 2011. This data set corresponds to the entirety of the `Reco 12-Stripping 17` dataset.

1.1 Angular observables

The decay $B^0 \to K^{*0}\mu^+\mu^-$ is a flavour changing neutral current process that proceeds via electroweak box or penguin diagrams in the Standard Model (SM). Beyond the SM, new particles can enter in loop-order diagrams with comparable amplitudes and lead to deviations from SM predictions. A number of angular observables in $B^0 \to K^{*0}\mu^+\mu^-$ decays can be theoretically predicted, with good control over the relevant form-factor uncertainties. These observables include the forward-backward asymmetry of the dimuon system, A_{FB}, and the fraction of longitudinal polarisation of the K^{*0}, F_{L}, as a function of the dimuon invariant mass-squared (q^2). This pair of observables has previously been measured by LHCb with $370\,\mathrm{pb}^{-1}$ [2][3] of integrated luminosity and by BaBar [4], Belle [5] and CDF [6][7]. A preliminary result has already been presented by LHCb with $1\,\mathrm{fb}^{-1}$ [8].

In the SM, A_{FB} varies as a function of q^2 and changes sign at a well defined point, q_0^2. This zero-crossing point comes from the interplay between the $\mathcal{O}_7$ (electromagnetic penguin) operator, which dominates as $q^2 \to 0$, and $\mathcal{O}_9$ and $\mathcal{O}_{10}$ (the vector and axial-vector) operators, which dictate the behaviour at high-q^2. In the SM the zero-crossing point is predicted to be [9]:

$$q^2_{0,\mathrm{S.M.}} = 3.97\, \underbrace{{}^{+0.03}_{-0.03}}_{\mathrm{F.F.}} \overbrace{{}^{+0.09}_{-0.09}}^{S.L.} \underbrace{{}^{+0.29}_{-0.27}}_{S.D.}\ \mathrm{GeV}^2/c^4$$

where the three uncertainties come from: the uncertainty on the form-factors (F.F.); the uncertainty on the unknown, 'sub-leading' (S.L.), Λ/m_b corrections; and the uncertainty on the short distance parameters (S.D.), including the uncertainty on m_t and m_W and on the scale-μ.

A_{FB} and F_{L} can be extracted from fits to the angular distribution of the muons, kaon and pion from the dimuon and K^{*0} decays. Two additional observables can be extracted from a fit to the data if the angle, ϕ, between the decay planes of the dimuon and the K^{*0} systems in the B^0 rest frame, is included. These observables are A_T^2, the asymmetry between the transverse K^{*0} amplitudes and A_{Im}, formed from the imaginary components of the transversity amplitudes of the K^{*0} [10]. The four angular observables are discussed in greater detail later in this note. A_T^2 in particular can have large sensitivity to the presence of new virtual particles that can modify the contribution from right-handed currents ($\mathcal{C}_7'$, $\mathcal{C}_9'$ and $\mathcal{C}_{10}'$). The observable $S_3 = \frac{1}{2}(1-F_L)A_T^2$ is sometimes used in literature instead of A_T^2 [11]. It has been shown in several papers [10, 12] that hadronic uncertainties cancel out, to a certain extent, when ratios of observables with the same form factor dependence are used. The observable A_T^2 is an example of these 'clean' observables. Other observables are $A_T^{Re} = (4/3) \times A_{FB}/(1-F_L)$ and $A_T^{Im} = 2 \times A_{Im}/(1-F_L)$. We will refer to the 'clean' set of observables A_T^2, A_T^{Im} and A_T^{Re} as transverse observables. The different choices of variable will be discussed in much greater detail later in this document.

1.2 Analysis strategy

The analysis strategy follows that outlined in Ref. [13]. A cut based preselection and multivariate selection are performed to reject combinatorial background (Sec. 3). Specific peaking backgrounds are then rejected using mass and particle identification criteria (Sec. 3.4). The q^2 regions which are dominated by J/ψ and $\psi(2S)$ resonances, which are difficult to be treated theoretically, are removed (Sec. 3.4). The effect of the event reconstruction, trigger and candidate selection on the angular distributions of the B^0 daughters is then accounted for by performing an acceptance correction using simulated events (Sec. 11). The simulation used has a set of data-derived corrections applied which remove the effect of data-simulation differences which are observed in control channels (see Sec. 10). Finally, in each q^2 bin, a fit is made to the angular distribution of the daughter particles (the kaon, pion and the muons) and the $K^+\pi^-\mu^+\mu^-$ [1] invariant mass to separate signal and background and to estimate the angular observables (Secs. 4 and 9). The angular basis is defined such that CP averaged quantities are measured throughout unless explicitly stated.

The decay $B^0 \to K^{*0} J/\psi$ is used throughout the analysis as a high statistics control channel, both for branching fraction normalisation and for validating the acceptance correction and the fitting procedure. $B^0 \to K^{*0} J/\psi$ events are selected using the same trigger, stripping and offline selection requirements as the signal, but with the J/ψ-veto reversed to reject $B^0 \to K^{*0}\mu^+\mu^-$ candidates.

In summary, this analysis note covers four separate analyses of the $B^0 \to K^{*0}\mu^+\mu^-$ data set. These are:

1. a measurement of the differential branching fraction of $B^0 \to K^{*0}\mu^+\mu^-$ in bins of q^2;

[1] Charge conjugation is implied throughout, unless explicitly stated otherwise.

2. a measurement of A_T^2, A_T^{Re}, A_T^{Im} (or equivalently S_3, A_{FB} and S_9) and $F_{\rm L}$ in bins of q^2;

3. a measurement of A_9, a T-odd CP asymmetry between B^0 and $\overline{B}^0$ decays;

4. a measurement of the zero-crossing point of $A_{\rm FB}$ from an "unbinned counting experiment".

The measurement of the differential branching fraction is described in Sec. 7. The extraction of the angular observables is described in Sec. 9. The zero-crossing point extraction is described in Sec. 21.

The use of the transverse observables, has implications in the fit, since the transverse variables appear as e.g. $(1 - F_L(q^2))A_T^2(q^2)$ in the angular distribution. This is discussed in more details in Section 8.7.

The contribution of a possible S-wave $K^+ \pi^-$ system interfering with the $K^{*0}(892)$, leading to a modified angular distribution, is also explored and discussed in Section **??**. In all previous analysis of $B^0 \to K^{*0}\mu^+\mu^-$, terms proportional to $m^2_{\mu^+\mu^-}/q^2$ in the angular distribution have been completely neglected. For the first time, at low-q^2 an attempt is made to account for the effect of neglecting these terms. This is discussed in detail in Section 17.

To summarise the main differences with the preliminary results shown at Moriond 2012, are:

1. Transverse observables are measured, as well as the non transverse observables already measured for the preliminary result. This is motivated by the fact that for transverse observables there is a reduced form factor dependence, making this observables cleaner from the theoretical point of view. A discussion on the observables and the implications can be found in Sec. 8.6 and Sec. 8.7.

2. The T-odd asymmetry A_9 is measured.

3. The S-wave contribution is estimated using the asymmetry in $\cos\theta_K$, and added as systematic. This is described in Sec. 16.

4. The effect of the threshold terms, arising from non-zero lepton masses, are considered in the lowest q^2 bin. A correction is applied and described in Sec. 17.

5. The Feldman-Cousins method is used to evaluate the uncertainty on the observables, in contrast with the `MINOS` error used for the preliminary result. This is described in Sec. 15.

6. The statistical uncertainty on the zero-crossing point is reduced. Due to a wrong behaviour of the code that calculated the statistical uncertainty on the zero-crossing point for the preliminary result, the weights were not included in the computation.

1.3 Data sets

This analysis is based on data corresponding to $1\,\mathrm{fb}^{-1}$ of integrated luminosity collected by the LHCb detector in 2011. Candidates have been reconstructed with `Reco 12` and stripped with `Stripping 17`. The multivariate selection described in Sec. 3.3 has been tuned using $36\,\mathrm{pb}^{-1}$ of integrated luminosity from `Reco 08` collected by LHCb in 2010. The data used to tune the multivariate selection is not used in the subsequent analysis. The multivariate selection is the same as described in Ref. [2].

The signal acceptance correction is evaluated using 50 M fully simulated $B^0 \to K^{*0}\mu^+\mu^-$ Monté Carlo (MC) events from MC10. These events have been generated as a phase-space decay, neglecting the physics in the angular distribution. In addition samples of $\mathcal{O}(1\,M)$, fully simulated, exclusive decays from MC10 are used to understand the contribution of peaking backgrounds to the final analysis.

2 Mass windows and q^2-binning

This section describes the K^+ π^- $\mu^+\mu^-$ and K^+ π^- mass windows used in the analysis. It also describes the choice of q^2-binning.

2.1 Definition of mass windows used in the analysis

Candidates are only considered for the analysis if they have a $K^+\pi^-\mu^+\mu^-$ invariant mass $m_{K^+\pi^-\mu^+\mu^-} > 5150\,\mathrm{MeV}/c^2$ and a $K^+\pi^-$ invariant mass $792 < m_{K^+\pi^-} < 992\,\mathrm{MeV}/c^2$ ($\pm 100\,\mathrm{MeV}/c^2$ from the nominal K^{*0} mass). Candidates are considered as being in a 'signal' mass window if the $K^+\pi^-\mu^+\mu^-$ invariant mass is in the range $5230 < m_{K^+\pi^-\mu^+\mu^-} < 5330\,\mathrm{MeV}/c^2$. The term upper sideband is used to refer to events with $K^+\pi^-\mu^+\mu^-$ invariant masses $5350 < m_{K^+\pi^-\mu^+\mu^-} < 5800\,\mathrm{MeV}/c^2$. The term lower sideband is used to refer to events with $K^+\pi^-\mu^+\mu^-$ invariant masses $5150 < m_{K^+\pi^-\mu^+\mu^-} < 5230\,\mathrm{MeV}/c^2$.

2.2 q^2-Binning

The choice of q^2 binning remains the same for this analysis as described in Ref. [13], apart for the treatment of the first q^2-bin, which is now restricted to $q^2 > 0.1\,\mathrm{GeV}^2/c^4$. This is motivated by the fact that the below $0.1\,\mathrm{GeV}^2/c^4$ the efficiency to reconstruct, trigger and select the $B^0 \to K^{*0}\mu^+\mu^-$ decay varies rapidly (making it difficult to appropriately model the acceptance).

Requiring that $q^2 > 0.1\,\mathrm{GeV}^2/c^4$ also significantly reduces the impact of the threshold terms that appear in the angular distribution at low-q^2. The q^2 binning is shown in Table. 1. This binning scheme was designed to match the binning used by BaBar, Belle and CDF. Due to limited MC-statistics the upper q^2 bin is limited to the range $16.0 < q^2 < 19.0\,\mathrm{GeV}^2/c^4$ and is not extended to the kinematic limit. Results will also be quoted in the theoretically favoured $1 < q^2 < 6\,\mathrm{GeV}^2/c^4$ range, which is far enough from the photon pole (at $q^2 \sim 0$) and the $c\bar{c}$ resonances for QCD factorisation to be used reliably. It is also relatively free from contributions from light-resonances. Further, for $q^2 > 1\,\mathrm{GeV}^2/c^4$, the threshold terms in the angular distribution can be neglected.

Binning	q^2 region (GeV^2/c^4)
q^2-binning scheme	$0.1 < q^2 < 2$
	$2 < q^2 < 4.3$
	$4.3 < q^2 < 8.68$
	$10.09 < q^2 < 12.86$
	$14.18 < q^2 < 16$
	$16 < q^2 < 19$
	$1 < q^2 < 6$

Table 1: Definition of q^2 bins used in the analysis. These include six q^2 bins covering $0.1 < q^2 < 19$ GeV^2/c^4 and the theoretically favoured region $1 < q^2 < 6$ GeV^2/c^4.

3 Selection

The offline event selection procedure follows that described in Ref. [14]. The only significant difference is an introduction of a cut on the transverse momentum of the four daughter particles (the kaon, pion and two muons), with $p_T > 250\,\mathrm{MeV}/c$, at the stripping level. This cut has a small impact on the input and output of the subsequent multivariate selection (based on a BDT). The stripping and offline selections are described briefly below. In addition to the MVA selection, cuts are applied to remove specific "peaking" backgrounds. These criteria are detailed in Sec. 3.4 and have been updated from the $0.37\,\mathrm{fb}^{-1}$ analysis [2]to reflect changes in the particle identification performance between `Reco 10` and `Reco 12`.

3.1 Trigger

Candidates are only considered for the offline analysis if they have passed through the following triggers: `L0Muon` at L0; `Hlt1TrackAllL0` or `Hlt1TrackMuon` at HLT 1; `Hlt2Topo[2,3,4]BodyBBDT`, `Hlt2TopoMu[2,3,4]BodyBBDT`, `Hlt2SingleMuon` or `Hlt2DiMuonDetached` at HLT 2. At all stages the offline-candidates are required to be `TOS`, i.e. the trigger decision is due solely to the presence of the candidate in the event. The trigger requirements are unchanged from the preliminary result with $1\,\mathrm{fb}^{-1}$ [14]. This choice of triggers only selects candidates in events with an SPD multiplicity < 600.

3.2 Stripping and pre-selection

This analysis uses candidates from the `StrippingBd2KstarMuMu` stripping line in `Reco 12-Stripping 17`. The cut based selection used in the stripping is close to that of the previous analysis (`Reco 10-Stripping 13b`). The only difference is a $p_T > 250\,\mathrm{MeV}/c$ cut on the muons, kaon and pion. The stripping selection requirements are included for reference in Table. 2.

Candidates from the stripping line are required to pass a further cut-based pre-selection (prior to the multivariate selection) to remove pathological events. These requirements are summarised in Table. 3.

3.3 Multivariate Offline Selection

The combinatorial background is reduced offline using a multivariate classifier: a boosted decision tree (BDT). The training and validation of the BDT is detailed in Ref. [14]. Briefly, the following information is input to the BDT:

Particle	Selection Requirement
B^0	$4850 < m_{K^+\pi^-\mu^+\mu^-} < 5780$ MeV/c^2
B^0	DIRA > 0.9999
B^0	Vertex χ^2/NDOF < 6
B^0	IP $\chi^2 < 16$
B^0	FD $\chi^2 > 121$
K^{*0}	$600 < m_{K^+\pi^-} < 2000$ MeV/c^2
K^{*0}	Vertex χ^2/NDOF < 12
K^{*0}	FD $\chi^2 > 9$
$\mu^+\mu^-$	FD $\chi^2 > 9$
$\mu^+\mu^-$	Vertex χ^2/NDOF < 12
Track	χ^2/ dof < 5
Track	IP $\chi^2 > 9$
Track	$p_{\mathrm{T}} > 250\,\mathrm{MeV}/c^2$
$\mu^\pm$	`IsMuonLoose` True

Table 2: Cut based selection used in `StrippingBd2KstarMuMu` for `Stripping 17`.

Particle	Selection Requirement
Track	$0 < \theta < 400$ mrad
Track	KL Distance > 5000
Track Pairs	$\theta > 1$ mrad
$\mu^+\mu^-$	`IsMuon` True
K	`hasRich` True
K	$\mathrm{DLL}_{K\pi} >$ -5
π	`hasRich` True
π	$\mathrm{DLL}_{K\pi} < 25$
PV	$\lvert X- < X > \rvert < 5\,\mathrm{mm}$
PV	$\lvert Y- < Y > \rvert < 5\,\mathrm{mm}$
PV	$\lvert Z- < Z > \rvert < 200\,\mathrm{mm}$

Table 3: Pre-selection cuts applied to stripped candidates.

- the B^0 pointing to the primary vertex, flight-distance and IP χ^2 with respect to the primary vertex, p_T and vertex quality (χ^2);
- the K^{*0} and dimuon flight-distance and IP χ^2 with respect to the primary vertex (associated to the B^0), p_T and vertex quality (χ^2);
- the impact parameter χ^2 and the $\Delta LL(K-\pi)$ and $\Delta LL(\mu-\pi)$ of the four final state particles.

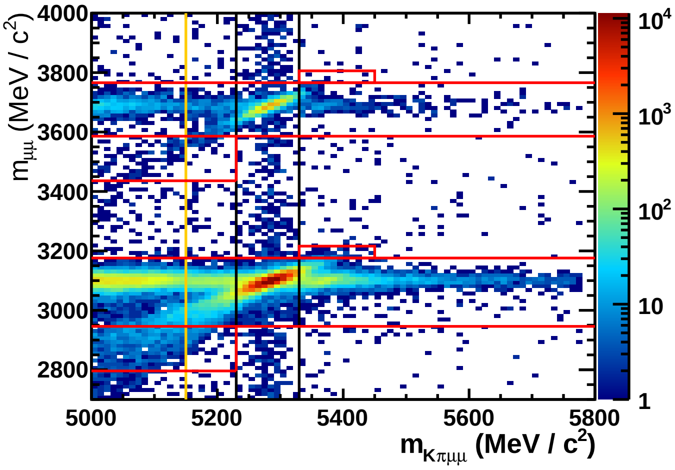
When training the BDT selection, $B^0 \to K^{*0}J/\psi$ candidates from the 2010 data were used as a proxy for the signal and $B^0 \to K^{*0}\mu^+\mu^-$ candidates from the upper mass sideband were used as a background sample. Half of the candidates were used for training (corresponding to $18\,\mathrm{pb}^{-1}$) and the remaining half used to test the performance of the BDT.

3.4 Specific background and vetoes

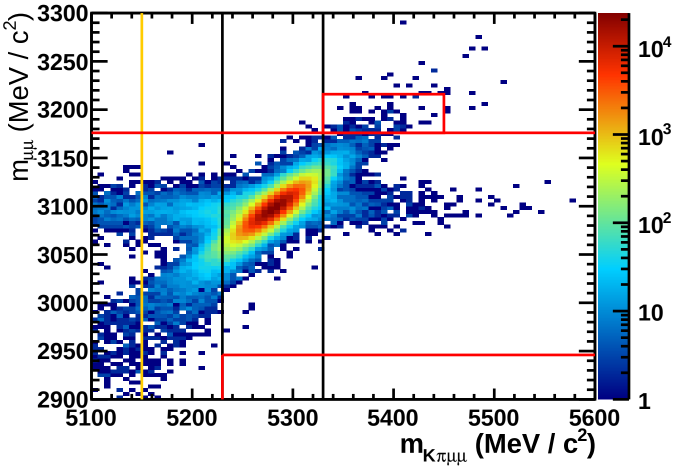
The decays $B^0 \to K^{*0}J/\psi$ and $B^0 \to K^{*0}\psi(2S)$ are treated separately in the analysis due to the different underlying physics that contributes in the decays. Event in the regions $2946 < m_{\mu^+\mu^-} < 3176$ MeV/c^2 and $3586 < m_{\mu^+\mu^-} < 3766$ MeV/c^2 for $B^0 \to K^{*0}J/\psi$ and $B^0 \to K^{*0}\psi(2S)$ are removed from the analysis. In addition the vetoes were extended to the region $2796 < m_{\mu^+\mu^-} < 3176$ MeV/c^2 and $3436 < m_{\mu^+\mu^-} < 3766$ MeV/c^2 for the events $m_{K\pi\mu^+\mu^-} < 5230$ MeV/c^2, to account for the radiative tail of the J/ψ decay. The vetoes were also extended to the region $3176 < m_{\mu^+\mu^-} < 3201\,\mathrm{MeV}/c^2$, to account for a misreconstructed tail of the J/ψ decay. This is shown in Fig. 1. Combinatorial background events are also removed by extending the veto regions. In order to correct for this, the remaining candidates in the bins of q^2 adjacent to the J/ψ and $\psi(2S)$ in the affected $K^+\pi^-\mu^+\mu^-$ invariant masses regions are re-weighted according to the fraction of the q^2 bin removed by the extending the vetoes. This re-weighting assumes that the background candidates are uniformly distributed in q^2 within the q^2 bin. This assumptions seems to hold well at the current level of precision.

In addition a number of specific backgrounds were considered in this analysis and the following additional vetoes have been applied:

- $B^0 \to K^*\mu^+\mu^-$ with $K \leftrightarrow \pi$ misidentification. This is dealt with by requiring $KDLL_{K\pi} + 10 < \pi DLL_{K\pi}$ for events where the $K^+\pi^-$ mass is in the range $792 < m_{K(\to\pi)\pi(\to K)} < 992$ after swapping the kaon and pion mass hypothesis.
- $B^0 \to J/\psi K^*$ where a muon is misidentified and swapped with the pion or kaon. This background is removed by rejecting candidates where



(a) Data



(b) MC10

Figure 1: The $K\pi\mu^+\mu^-$ versus $\mu^+\mu^-$ invariant mass distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates that lie close to the J/ψ mass in the data (left) and in $B^0 \to K^{*0}J/\psi$ MC (right). The charmonium veto regions are indicated by the red lines. The yellow line indicates the extent of the lower mass sideband used for the angular analysis.

the pion/kaon passes the IsMuon requirements or has $DLL_{\mu\pi} > 5.0$ if the $K^+\mu^-$ or $\pi^-\mu^+$ mass is in the range $[3036, 3156]$ MeV/c^2, after exchanging the π/K with the muon mass hypothesis.

- $B_s \to \phi\mu^+\mu^-$ where a K from the ϕ-meson is misidentified as a π. Such events are removed by applying the following cuts for events that fall in the region $5321 < m_{KK\pi\pi} < 5411$ MeV/c^2: $\pi DLL_{K\pi} > -50$ for events in the region $1010 < m_{KK} < 1030$ MeV/c^2 and $\pi DLL_{K\pi} > 20$ for events in the region $1030 < m_{KK} < 1075$ MeV/c^2.

- $B^+ \to K^+\mu^+\mu^-$ combined with a soft pion coming from elsewhere in the event. This background peaks on the right of the signal window, in the upper mass sideband, and is removed by vetoing the region of $K^+\mu^+\mu^-$ invariant mass $5220 < m_{K\mu^+\mu^-} < 5340\,\mathrm{MeV}/c^2$.

- $\Lambda_b \to pK^-\mu^+\mu^-$ where either the proton is identified as a pion or the proton is identified as a kaon and the kaon as a pion. This background is removed by rejecting candidates with $\pi/KDLL_p > 20$ and both $5575 < m_{K^+p^-\mu^+\mu^-} < 5665\,\mathrm{MeV}/c^2$ and $1490 < m_{K^+p^-} < 1550\,\mathrm{MeV}/c^2$, after exchanging the pion mass with the proton (or pion with kaon, kaon with proton) mass hypothesis.

Peaking backgrounds from $B^0 \to \rho^0\mu^+\mu^-$, $B^+ \to K^{*+}\mu^+\mu^-$, $B_s \to f_0\mu^+\mu^-$ and $B_s^0 \to K^{*0}\mu^+\mu^-$ have also been studied using simulated events (correcting for the PID performances observed in data) and found to be negligible.

Partially reconstructed $B \to K^+\pi^-\mu^+\mu^- + X$ where one or more particles from a B-meson decay are not reconstructed are removed by requiring that candidates have an invariant mass $m_{K^+\pi^-\mu^+\mu^-} > 5150\,\mathrm{MeV}/c^2$. Finally cascade decays where B^0 decays semileptonically to a D meson that in turn decays semileptonically, sits in the lower mass sideband. This background is largely removed by requiring $m_{K^+\pi^-\mu^+\mu^-} > 5150\,\mathrm{MeV}/c^2$. This has been validated using older MC studies [15]. Further it has been checked that the angular distribution of candidates below the signal mass window, but with $m_{K^+\pi^-\mu^+\mu^-} > 5150\,\mathrm{MeV}/c^2$, is consistent with those appearing in the upper mass sideband.

The background from a possible broad S-wave $K^+\pi^-$ system or from the tail of $K_0^*(1430)$ is discussed in Sec 16.

The level of peaking background remaining after applying the full selection requirements and vetoes, is given in Table 4. These backgrounds are ignored in the subsequent angular analysis, but are including in the branching fraction determination. A systematic uncertainty is assigned to the result of the angular analysis to reflect the assumption that these backgrounds can be neglected.

The level of $\Lambda_b \to pK^-\mu^+\mu^-$ was estimated using $\Lambda_b \to pK^- J/\psi$ events in data. These decays were isolated in data in the upper B mass sideband. The level of events inside the B mass window was extracted using the B mass distribution of $\Lambda_b \to pK^-\mu^+\mu^-$ simulated events. From this the ratio of $\Lambda_b \to pK^- J/\psi$ and $B^0 \to K^{*0} J/\psi$ in data in the signal region was found to be approximately 1.5%. Assuming the same ratio for the $\mu^+\mu^-$ mode, the level of $\Lambda_b \to pK^-\mu^+\mu^-$ events is 1.5% of the signal yield. The veto applied (above) rejects 50% of simulated $\Lambda_b \to pK^-\mu^+\mu^-$ events, reducing this peaking background to the level of $\approx 0.75\%$.

Background	Background Level (%)	Signal Loss (%)
$B^0 \to K^{*0}\mu^+\mu^-$ (with $K \leftrightarrow \pi$)	0.85 ± 0.02	0.11
$B^0 \to K^{*0} J/\psi$ (with $\pi \leftrightarrow \mu$)	0.27 ± 0.08	0.05
$B^0 \to K^{*0} J/\psi$ (with $K \leftrightarrow \mu$)	0.00 ± 0.00	0.03
$B^0_s \to \phi\mu^+\mu^-$	1.23 ± 0.50	0.32
$B^+ \to K^+\mu^+\mu^-$	0.14 ± 0.03	–
$\Lambda_b \to pK^-\mu^+\mu^-$	0.75 ± 0.15	0.47
Total	3.24 ± 0.53	0.98

Table 4: The level of exclusive peaking backgrounds with respect to the $B^0 \to K^{*0}\mu^+\mu^-$ signal (as scaled from the relative efficiency in MC and the PDG branching fraction).

3.5 Multiple Candidates

After applying the multivariate selection and peaking background vetoes it is still possible to have multiple candidates in the final data sample. This includes situations where the K and the π are swapped (as only a loose PID requirement is made). Multiple candidates surviving the selection were treated by weighting each candidate by the inverse of the number of candidates in that event. After the selection 98% (98%) of events in the $B^0 \to K^{*0}\mu^+\mu^-$ ($B^0 \to K^{*0} J/\psi$) signal mass window have just one candidate. In the upper mass sideband, 98% (97%) of events have just one candidate.

4 $K^+ \pi^- \mu^+\mu^-$ and $K^+ \pi^-$ invariant mass distributions

4.1 $K^+\pi^-\mu^+\mu^-$ invariant mass distribution

The mass model used for the signal and background is explored using $B^0 \to K^{*0} J/\psi$ events and $B^0 \to K^{*0}\mu^+\mu^-$ MC. The background mass distribution is parametrised by an exponential to model the combinatorial background. In the $B^0 \to K^{*0}\mu^+\mu^-$ analysis candidates are only considered if they have $m_{K^+\pi^-\mu^+\mu^-} > 5150\,\mathrm{MeV}/c^2$. In this section, this requirement has been relaxed to highlight the contribution from partially reconstructed B decays. A `RooExpAndGauss` model is used to model this background shape, describing an exponential rise to a threshold with a Gaussian fall off above the threshold. This is empirically is seen to describe well the data for $m_{K^+\pi^-\mu^+\mu^-} < 5150\,\mathrm{MeV}/c^2$.

The signal mass distribution is parametrised by the sum of two Crystal Ball shapes [16], with both tails on the left hand side of the distribution. The nominal B^0 mass, μ_{B^0}, and shape parameters α and n are assumed to be common between the two crystal ball shapes, but the widths of the distributions σ_1 and σ_2 are allowed to float in the fit to $B^0 \to K^{*0} J/\psi$. The signal shape parameters are then fixed to their best fit values when fitting the invariant mass distribution of $B^0 \to K^{*0}\mu^+\mu^-$ decays. Again, the choice of signal model is empirical and we use the minimal model that well describes the mass distribution in data and in SM $B^0 \to K^{*0}\mu^+\mu^-$ MC. The $K^+\pi^-\mu^+\mu^-$ invariant mass distribution of $B^0 \to K^{*0} J/\psi$ decays in the J/ψ mass window is shown in Fig. 2. A fit to the data with the full double Crystal Ball model is overlaid. For $B^0 \to K^{*0} J/\psi$ a second signal component is included for $B^0_s \to \overline{K}^{*0} J/\psi$ decays that is suppressed by f_s/f_d and a CKM factor. In the fit the fraction of B^0_s decays is constrained from Ref. [17] to be $0.7 \pm 0.2\%$. This B^0_s contribution is not included in the fit to $B^0 \to K^{*0}\mu^+\mu^-$.

The q^2-dependence of the $K^+\pi^-\mu^+\mu^-$ invariant mass distribution is explored using SM MC. There is a small difference in the signal mass resolution between low and high-q^2. Differences are visible at the level of 5%, but there is no dramatic worsening of the resolution in q^2. This is treated as a source of systematic.

4.2 $K^+\pi^-$ invariant mass distribution

Fig. 3 shows the two dimensional, $K^+\pi^-\mu^+\mu^-$ versus $K^+\pi^-$ invariant mass distribution for $B^0 \to K^{*0}\mu^+\mu^-$ candidates and J/ψ candidates. The contribution from the $K^{*0}(892)$ is visible in both figures as are contributions from higher K^* states around the $K^*(1430)$. There is also clear evidence for a broad structure that extends between the $K^{*0}(892)$ and the $K^*(1430)$ that can be partially attributed to the tails of the $K^{*0}(892)$ and the higher states and to the presence of a $K\pi$ S-wave. No attempt is made here to disentangle the overlapping higher mass states. The effect of a $K\pi$ S-wave is discussed later.

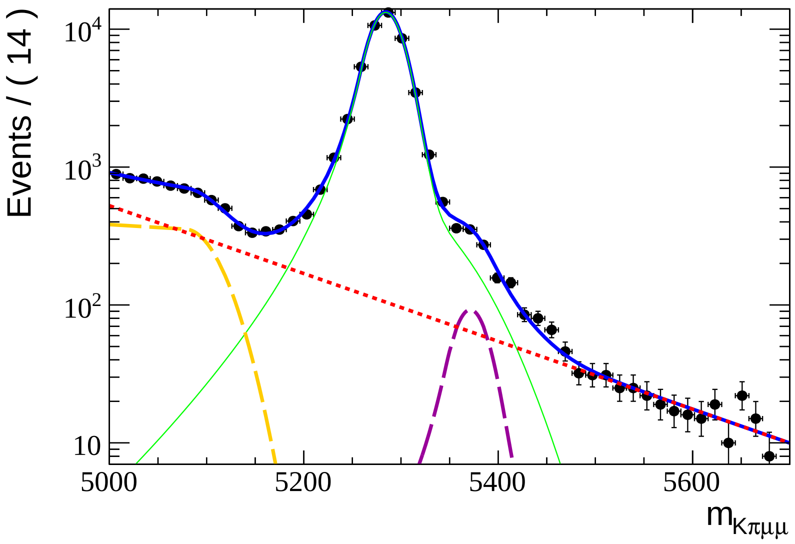
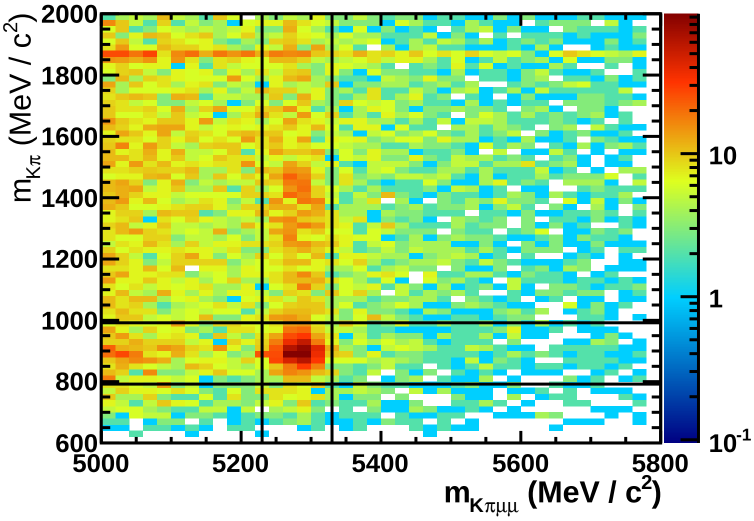
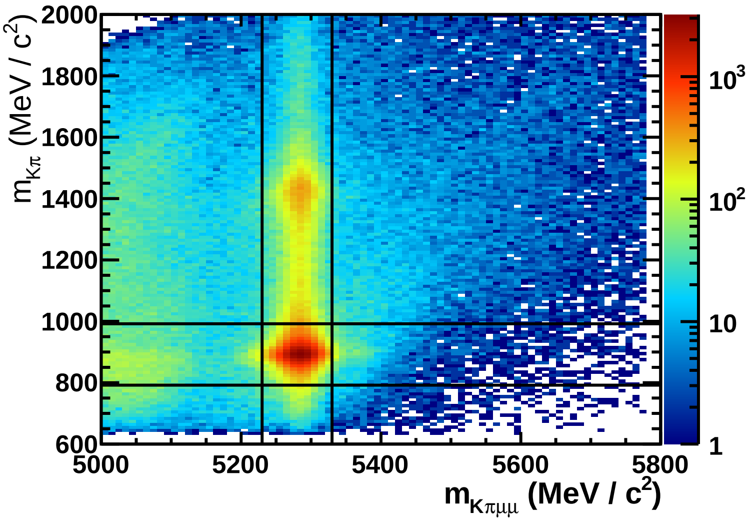


Figure 2: The $K^+\pi^-\mu^+\mu^-$ invariant mass of $B^0 \to K^{*0} J/\psi$ candidates fitted with a: double Crystal Ball shape for the signal component (thin-green line) and $B_s^0 \to \overline{K}^{*0} J/\psi$ (long-dashed purple line); an exponential shape to model combinatorial background (dotted-red line) and a `RooExpAndGauss` shape to model low-mass partially reconstructed backgrounds (dashed-yellow line). The full fit model (blue line) has a $P(\chi^2) = 6\%$.



(a) $B^0 \to K^{*0}\mu^+\mu^-$



(b) $B^0 \to K^{*0}J/\psi$

Figure 3: The $K^+\pi^-\mu^+\mu^-$ versus $K^+\pi^-$ invariant mass distribution for candidates outside the J/ψ and $\psi(2S)$ vetoes (left) and for candidates in the J/ψ veto region (right). The solid lines represent the signal $K^+\pi^-\mu^+\mu^-$ and the $K^+\pi^-$ mass window used in the subsequent analysis.

5 Event yields

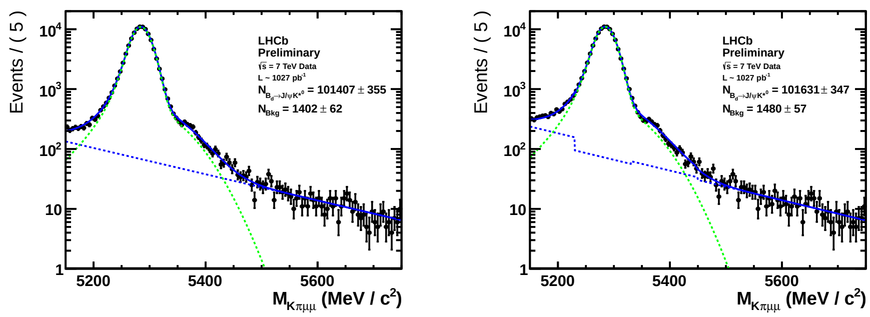
The $K^+\pi^-\mu^+\mu^-$ invariant mass distribution of $B^0 \to K^{*0}J/\psi$ candidates is shown in Fig. 4. The same selection, including the peaking vetoes (apart for the J/ψ veto) are applied to the $B^0 \to K^{*0}J/\psi$ and to the signal. The yield of $B^0 \to K^{*0}J/\psi$ in about 1fb^{-1} is 101407 ± 355 events, which is in agreement with what is expected. The line-shape from a fit to the distribution is then used to estimate the $B^0 \to K^{*0}\mu^+\mu^-$ yield in the full q^2 window and in each of the six bins used in the angular analysis. In the fit to $B^0 \to K^{*0}\mu^+\mu^-$, the shape parameters are floated, but constrained to the result of the fit to $B^0 \to K^{*0}J/\psi$. This implicitly assumes that the width of the signal distribution is independent of q^2 (see Sec. 4). The effect from multiple candidates has been neglected here.

The $K^+\pi^-\mu^+\mu^-$ invariant mass distribution, after applying the vetoes for peaking backgrounds, of $B^0 \to K^{*0}\mu^+\mu^-$ candidates is shown in Fig. 5. The $K^+\pi^-\mu^+\mu^-$ invariant mass distributions of the six q^2 bins are shown in Figs. 6(a)-(f). Table. 5 lists the signal and background yield in a $\pm 50\,\text{MeV}/c^2$ signal mass window in each of the q^2-bins. Note, the uncertainty on the background yield appearing in the table is smaller than the square-root of the background yield as it is scaled appropriately from the background yield, in the full mass window. In total, 883 signal candidates are seen with $0.1 < q^2 < 19\,\text{GeV}^2/c^4$. The results of these fits are provided for reference only, they are not used in the angular analysis, where the inclusion of the signal angular distribution and re-weighting of the candidates for the detector acceptance can impact the signal-to-background ratio.

The yield has scaled as expected from the $0.37\,\text{fb}^{-1}$ analysis where 337 signal candidates were observed in the signal mass window.

q^2 (GeV^2/c^4) range	Signal Yield	Background Yield
$0.1 < q^2 < 2$	139.9 ± 13.4	26 ± 3.7
$2 < q^2 < 4.3$	72.6 ± 10.8	35.6 ± 4.2
$4.3 < q^2 < 8.68$	270.8 ± 18.9	56 ± 5.5
$10.09 < q^2 < 12.86$	168.1 ± 15	39 ± 4.5
$14.18 < q^2 < 16$	115.1 ± 11.7	14.2 ± 2.9
$16 < q^2 < 19$	116.3 ± 12.5	23.1 ± 3.6
$1 < q^2 < 6$	197 ± 17.1	72.2 ± 5.9
$0.1 < q^2 < 19$	883.3 ± 34.3	193.8 ± 10.2

Table 5: The signal and background yields resulting from a fit to the $K^+\pi^-\mu^+\mu^-$ invariant mass distributions of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the six q^2-bins used in the analysis, the theoretically 'favoured' $1 < q^2 < 6\,\mathrm{GeV}^2/c^4$ range and in the full q^2-range.



(a) $3036 < m_{J/\psi} < 3156\,\mathrm{MeV}/c^2$ (b) Inverted veto

Figure 4: The $K^+\pi^-\mu^+\mu^-$ invariant mass distribution of $B^0 \to K^{*0} J/\psi$ candidates in the data after the full selection has been applied. The fitted signal (green dotted) and background shapes are is described in Sec. 4. The left plot requires candidates in the di-mu mass region $3036 < m_{J/\psi} < 3156\,\mathrm{MeV}/c^2$ as in the previous analysis. The right plot applies the inverse of the J/ψ veto region, in order to fully capture the radiative tail. The background model is modified to account for the additional combinatorial background.

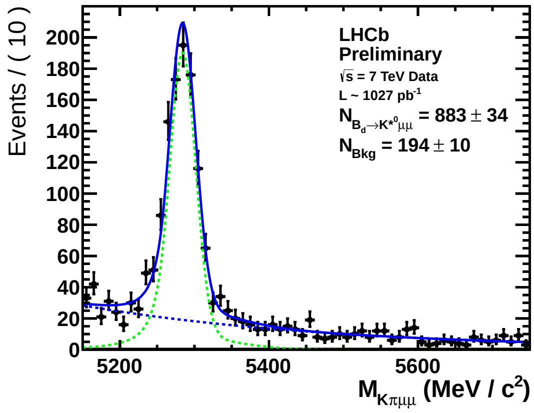
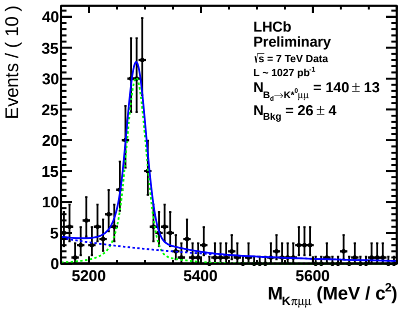
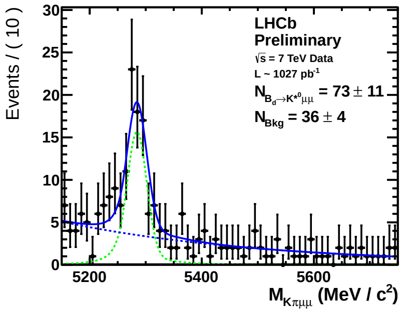


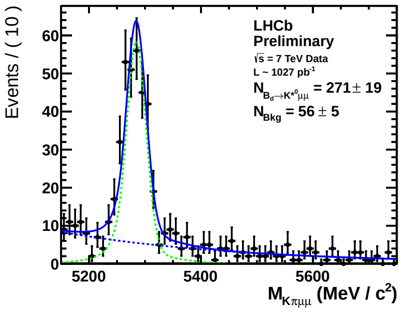
Figure 5: The $K^+\pi^-\mu^+\mu^-$ invariant mass distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates, in the range $0.1 < q^2 < 19\,\mathrm{GeV}^2/c^4$, in the data after the full selection has been applied. The fitted signal (green dotted) and background shapes are is described in Sec. 4.



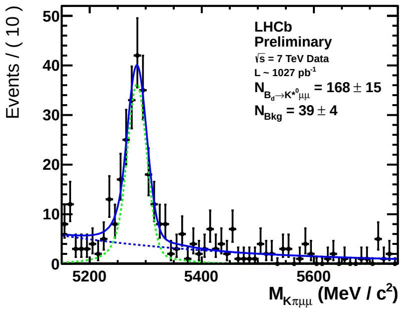
(a) $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$



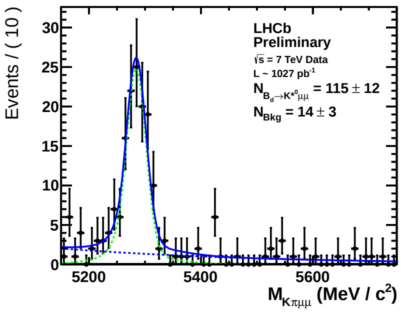
(b) $2 < q^2 < 4.3\,\mathrm{GeV}^2/c^4$



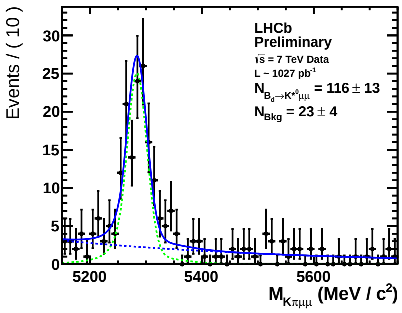
(c) $4.3 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$



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



(e) $14.18 < q^2 < 16\,\mathrm{GeV}^2/c^4$



(f) $16 < q^2 < 19\,\mathrm{GeV}^2/c^4$

Figure 6: The $K^+\pi^-\mu^+\mu^-$ invariant mass distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the data in the six q^2-bins used in the analysis. The fitted signal (green dotted) and background shapes are is described in Sec. 4. The signal has a significance greater than 5 "sigma" in all six q^2-bins.

6 q^2 spectrum of signal candidates

The q^2 spectrum of signal candidates is unfolded using the *sPlot* technique with the $K^+\pi^-\mu^+\mu^-$ invariant mass as the discriminating variable. The resulting distribution is shown in Fig. 7.

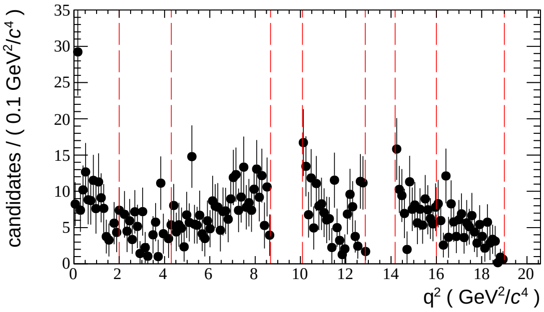


Figure 7: The background subtracted q^2 distribution of $B^0 \to K^{*0}\mu^+\mu^-$ signal candidates obtained using the *sPlot* technique. The dashed lines indicate the boundaries between the different q^2 bins used in this analysis.

If the background subtraction is performed independently in the q^2 bins, the average q^2 value of the signal in each q^2 bin is given in Table. 6.

	$< q^2 >$
$0.10 < q^2 < 2.00\,\text{GeV}^2/c^4$	$0.8\,\text{GeV}^2/c^4$
$2.00 < q^2 < 4.30\,\text{GeV}^2/c^4$	$3.1\,\text{GeV}^2/c^4$
$4.30 < q^2 < 8.68\,\text{GeV}^2/c^4$	$6.7\,\text{GeV}^2/c^4$
$10.09 < q^2 < 12.86\,\text{GeV}^2/c^4$	$11.3\,\text{GeV}^2/c^4$
$14.18 < q^2 < 16.00\,\text{GeV}^2/c^4$	$15.0\,\text{GeV}^2/c^4$
$16.00 < q^2 < 19.00\,\text{GeV}^2/c^4$	$17.2\,\text{GeV}^2/c^4$
$1.00 < q^2 < 6.00\,\text{GeV}^2/c^4$	$3.5\,\text{GeV}^2/c^4$

Table 6: The background subtracted mean q^2 value of $B^0 \to K^{*0}\mu^+\mu^-$ signal candidates in the q^2 bins. The values have been obtained using the *sPlot* technique.

7 Differential branching fraction

The differential branching fraction as a function of q^2, $\mathrm{d}\mathcal{B}/\mathrm{d}q^2$ receives similar enhancements from "new physics" to the angular observables. However, the sensitivity to the "new physics" in $\mathrm{d}\mathcal{B}/\mathrm{d}q^2$, is limited by the large uncertainty ($\mathcal{O}(30\%)$) on the hadronic form factors.

The partial branching fraction, $\mathcal{B}_k$, in the q^2 bin can be estimated by comparing the yield of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the q^2 bin to the number of $B^0 \to K^{*0}J/\psi$ candidates in the total sample. The partial branching fraction is then given by

$$\mathcal{B}_k = \mathcal{B}(B^0 \to K^{*0}J/\psi) \times \mathcal{B}(J/\psi \to \mu^+\mu^-) \times \frac{N_{K^{*0}\mu^+\mu^-;\,k}}{N_{K^{*0}J/\psi}} \frac{\varepsilon_{K^{*0}J/\psi}}{\varepsilon_{K^{*0}\mu^+\mu^-;k}} \ ,$$

where $N_{K^{*0}\mu^+\mu^-;k}$ is the number of $B^0 \to K^{*0}\mu^+\mu^-$ candidates in bin k, $N_{K^{*0}J/\psi}$, is the number of $B^0 \to K^{*0}J/\psi$ candidates in the full data sample and $\epsilon_{K^{*0}J/\psi}/\varepsilon_{K^{*0}\mu^+\mu^-;k}$ is the ratio of efficiencies between the two decays. This last number would traditionally be take from MC samples. Unfortunately, whilst $\varepsilon_{K^{*0}J/\psi}$ is known precisely from simulated events, $\varepsilon_{K^{*0}\mu^+\mu^-;k}$ is poorly known because it depends on the unknown angular distribution and q^2 spectrum.

To avoid making any assumption about the unknown angular distribution of the $B^0 \to K^{*0}\mu^+\mu^-$ decay, event-by-event weights (see Sec. 11) are used to estimate the average efficiency of signal candidates in each q^2 bin. The procedure is described below.

7.1 Determining $\mathrm{d}\mathcal{B}/\mathrm{d}q^2$ using event-by-event weights

The yield in each q^2 bin is extracted by using an extended unbinned maximum likelihood fit to the $K^+\pi^-\mu^+\mu^-$ invariant mass distribution to the candidates in the q^2 bin. In this likelihood fit, the candidates are weighted to account for the detector acceptance in the same manner in which they are for the angular analysis. As in the angular analysis the weights are normalized to be on average one, i.e. that

$$\sum_{i=0}^{N_k} \alpha_k w_i = N_k \tag{1}$$

where w_i is the event-by-event weight. The factor α used for the normalization of the event weights. The procedure to calculate the partial branching fraction in each bin then consists of the following steps:

- Each event is weighted in the extended likelihood fit to the $K^+\pi^-\mu^+\mu^-$ invariant mass;

- The weights are normalised such that the sum of the weights is the number of events (scaling the weights by a normalisation factor $\alpha_{K^{*0}\mu^+\mu^-}$);

- The procedure is repeated for $B^0 \to K^{*0} J/\psi$ (with a normalisation factor $\alpha_{K^{*0}J/\psi}$);

- The differential branching fraction is extracted from the number of events that come from the two likelihood fits and the ratio of the normalisation factors.

In the q^2 bin, $\mathcal{B}_k$ is then given by

$$\mathcal{B}_k = \mathcal{B}(B^0 \to K^{*0} J/\psi) \times \mathcal{B}(J/\psi \to \mu^+\mu^-) \times \frac{N'_{K^{*0}\mu^+\mu^-;k}}{N'_{K^{*0}J/\psi}} \frac{\alpha_{K^{*0}J/\psi}}{\alpha_{K^{*0}\mu^+\mu^-;k}} , \quad (2)$$

where $N'_{K^{*0}\mu^+\mu^-;k}$ and $N'_{K^{*0}J/\psi}$ denote the $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to K^{*0} J/\psi$ event yields in the q^2 bin that come from the weighted likelihood fit.

The resulting differential branching fraction in the q^2 bin is then given by

$$\frac{d\mathcal{B}_k}{dq^2} = \frac{1}{q^2_{\mathrm{max.};k} - q^2_{\mathrm{min.};k}} \mathcal{B}_k \quad .$$

The contributions from the decays $B^0_s \to \overline{K}^{*0}\mu^+\mu^-$ and $B^0_s \to \phi\mu^+\mu^-$ (where one kaon is identified as a pion) are included in the fit, but are fixed to the expected level of background from Sec. 3.4. $B^0_s \to \overline{K}^{*0}\mu^+\mu^-$ is assumed to be at the level of $f_{B^0_s} = 1 \pm 1\%$ ($\simeq (f_s/f_d)|V_{td}/V_{ts}|^2$) of the signal. $B^0_s \to \phi\mu^+\mu^-$ at the level of $f_\phi = 1.2 \pm 0.5\%$ of the signal. The line-shape of the $\overline{B}^0_s \to K^{*0}\mu^+\mu^-$ is assumed to be the same as the $B^0 \to K^{*0}\mu^+\mu^-$ signal. A template for the shape of the $B^0_s \to \phi\mu^+\mu^-$ line-shape has been taken from SM MC. The uncertainty on the line-shape of this background is small compared to the uncertainty on the yield, therefore no systematic uncertainty on the shape is considered, but the level of each background is varied within its uncertainty.

7.2 Unbinned maximum likelihood fit for the differential branching fraction

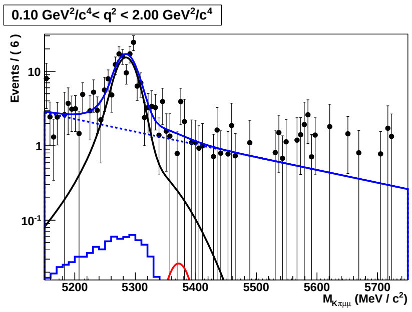
Summarising the contributions, the log-likelihood is given by:

$$
\begin{aligned}
-\log L = -\sum_{i=0}^{N} \alpha w_i \log \Bigg[& \frac{N'_{\text{sig}}}{(1+f_\phi+f_{B_s^0})N'_{\text{sig}}+N'_{\text{bkg}}} M(m_{K^+\pi^-\mu^+\mu^-}|\sigma_1,\sigma_2,\alpha,n) \quad + \\
& \frac{f_{B_s^0}\times N'_{sig}}{(1+f_\phi+f_{B_s^0})N'_{\text{sig}}+N'_{\text{bkg}}} M(m_{K^+\pi^-\mu^+\mu^-}|\sigma_1,\sigma_2,\alpha,n) \quad + \\
& \frac{f_\phi\times N'_{sig}}{(1+f_\phi+f_{B_s^0})N'_{\text{sig}}+N'_{\text{bkg}}} F_\phi(m_{K^+\pi^-\mu^+\mu^-}) \quad + \\
& \frac{N'_{\text{bkg}}}{(1+f_\phi+f_{B_s^0})N'_{\text{sig}}+N'_{\text{bkg}}} E(m_{K^+\pi^-\mu^+\mu^-}|p_0) \Bigg] \quad - \\
& \log P(N|(1+f_\phi+f_{B_s^0})N'_{\text{sig}}+N'_{\text{bkg}})
\end{aligned}
\tag{3}
$$

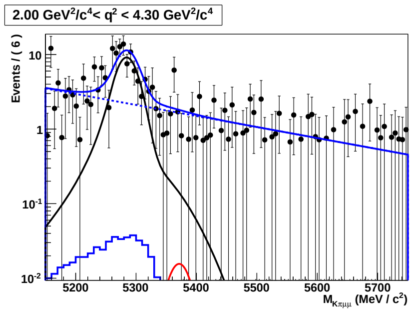
where $M(m_{K^+\pi^-\mu^+\mu^-}|\sigma_1,\sigma_2,\alpha,n)$ is the double crystal ball mass model for the signal described above, $E(m_{K^+\pi^-\mu^+\mu^-}|p_0)$ is an exponential model for the combinatorial background, $N'_{\text{sig.}}$ is the effective number of signal candidates and N'_{bkg} the effective number of background candidates. F_ϕ, is the template for the $B_s^0 \to \phi\mu^+\mu^-$ line-shape. The $B_s^0 \to K^{*0}\mu^+\mu^-$ line-shape is fixed to be the same as the signal line-shape, but is shifted in $K^+\pi^-\mu^+\mu^-$ invariant mass by the $B_s^0 - B^0$ mass difference. The weights are normalised as described above.

7.3 Results from fits to the $1\,\text{fb}^{-1}$ data sample

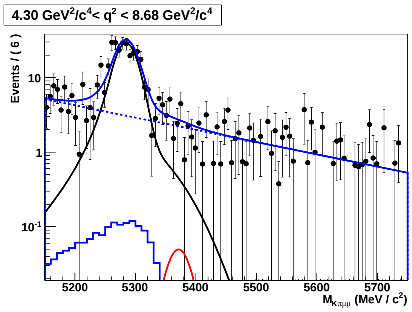
The differential branching ratio as a function of q^2 is summarised in Table 7. It is consistent with previous results (from LHCb, the B-factories and CDF) and with the SM prediction.



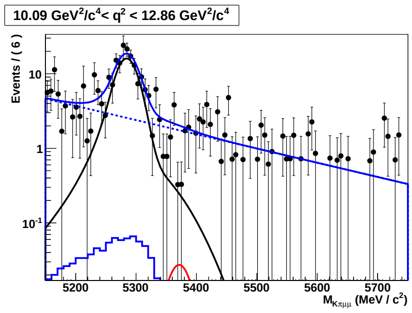
(a) $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$



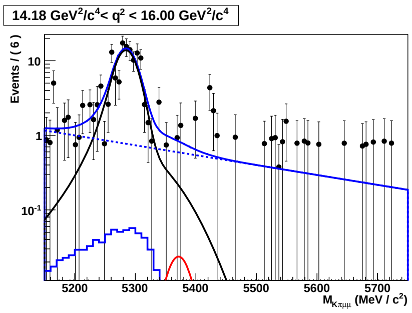
(b) $2 < q^2 < 4.3\,\mathrm{GeV}^2/c^4$



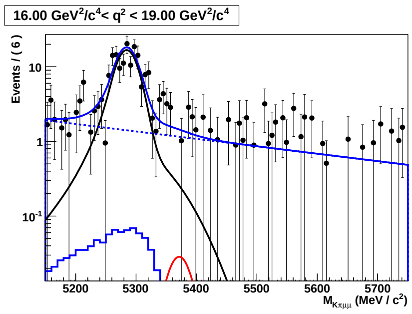
(c) $4.3 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$



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



(e) $14.18 < q^2 < 16\,\mathrm{GeV}^2/c^4$



(f) $16 < q^2 < 19\,\mathrm{GeV}^2/c^4$

Figure 8: Mass fit to the invariant $K^+\pi^-\mu^+\mu^-$ mass used to determine the differential branching ratio. The mass fit is described in more detail in Section 4.

q^2-bin	$d\mathcal{B}/dq^2(10^{-7}c^4/GeV^2)$
$0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$	0.61 ± 0.08
$2.00 < q^2 < 4.30\,\mathrm{GeV}^2/c^4$	0.30 ± 0.05
$4.30 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$	0.50 ± 0.05
$10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$	0.43 ± 0.05
$14.18 < q^2 < 16.00\,\mathrm{GeV}^2/c^4$	0.55 ± 0.07
$16.00 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$	0.38 ± 0.05
$1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$	0.35 ± 0.04

Table 7: The measured differential branching fraction for $B^0 \to K^{*0}\mu^+\mu^-$ in bins of q^2. The errors are purely statistical and are the result of the fit described in the text.

7.4 Cross check of the differential branching fraction

As a cross check, the differential branching ratio was calculated from the event yields in Sec. 5, taking an average efficiency for the signal candidates in the q^2 bin, rather than weighting the candidates in the fit. The average efficiency is estimated in two ways: firstly using SM MC and secondly using the *sPlot* technique [18] to unfold the efficiency distribution of the signal. The two approaches, of weighting in or after the fit, lead to consistent results.

The error estimates on N'_{sig} coming from the weighted-likelihood fit are shown to be reliable using toy-experiments. Unlike the angular analysis, the weights are uncorrelated to the $K^+\pi^-\mu^+\mu^-$ inviariant mass distribution and the naive scaling of the weights by α is appropriate,

7.5 Systematic uncertainties

In this section the result of the measurement of the differential branching ratio including the systematic uncertainty is shown.

The systematic uncertainty on $\mathrm{d}\mathcal{B}/\mathrm{d}q^2$ has been estimated by repeating the fits to the $K^+\pi^-\mu^+\mu^-$ invariant mass with a different, systematically varied acceptance correction. The difference between $\mathrm{d}\mathcal{B}/\mathrm{d}q^2$ in the fit with the varied acceptance and the nominal one is assigned as a systematic uncertainty. A complete description of the acceptance variations that are tried can be found in Sec. 18. The mass fits have also been repeated after changing the peaking background level by one sigma of the estimated uncertainty. This variation has a negligible effect on the $\mathrm{d}\mathcal{B}/\mathrm{d}q^2$. A 5% variation of the signal mass resolution has also been considered.

Finally, a one side systematic uncertainty is assigned to $\mathrm{d}\mathcal{B}/\mathrm{d}q^2$ to account for the possible S-wave contamination in the $B^0 \to K^{*0}\mu^+\mu^-$ decay. The S-wave is indistinguishable from the signal in $K^+\pi^-\mu^+\mu^-$ and will lead to a small over-estimate of the differential branching fraction. There will also be an S-wave contamination in the normalisation channel ($B^0 \to K^{*0} J/\psi$). This contamination is however accounted for in the branching fraction that we use for normalisation, which in reality corresponds to $\mathcal{B}(B^0 \to K^+\pi^- J/\psi)$ in the same $\pm 100\,\mathrm{MeV}/c^2$ mass window used in our analysis. An upper limit on the S-wave contamination to $B^0 \to K^{*0}\mu^+\mu^-$ is determined to be $F_S \lesssim 0.07$ at 68% confidence level (see Sec. 16 for details).

The dominant source of systematic uncertainty arises from the 4% uncertainty on the $B^0 \to K^{*0} J/\psi$ and $J/\psi \to \mu^+\mu^-$ branching fractions. The resulting differential branching fraction, including the full list set of systematic uncertainties is summarised in Table. 8. A breakdown of the contributions to the total systematic uncertainty is given in Table. 9.

q^2-bin	$d\mathcal{B}/dq^2 (10^{-7} c^4/\,\mathrm{GeV}^2)$
$0.10 < q^2 < 2.00\,\mathrm{GeV}^2/c^4$	$0.61 \pm 0.08 \pm 0.05^{+0.0}_{-0.05}$
$2.00 < q^2 < 4.30\,\mathrm{GeV}^2/c^4$	$0.30 \pm 0.05 \pm 0.03^{+0.0}_{-0.02}$
$4.30 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$	$0.50 \pm 0.05 \pm 0.04^{+0.0}_{-0.04}$
$10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$	$0.43 \pm 0.05 \pm 0.04^{+0.0}_{-0.03}$
$14.18 < q^2 < 16.00\,\mathrm{GeV}^2/c^4$	$0.57 \pm 0.07 \pm 0.04^{+0.0}_{-0.05}$
$16.00 < q^2 < 19.00\,\mathrm{GeV}^2/c^4$	$0.42 \pm 0.05 \pm 0.04^{+0.0}_{-0.03}$
$1.00 < q^2 < 6.00\,\mathrm{GeV}^2/c^4$	$0.35 \pm 0.04 \pm 0.04^{+0.0}_{-0.03}$

Table 8: The measured differential branching fraction for $B^0 \to K^{*0}\mu^+\mu^-$ in bins of q^2. The first error is statistical, the second systematic, the third error is due to the S-wave contribution.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal	0.61	0.30	0.49	0.42	0.54	0.37	0.34
B^0 p_T re-weighting [I]	0.61	0.30	0.49	0.42	0.54	0.38	0.34
PID performance -5% [G]	0.62	0.30	0.49	0.42	0.53	0.37	0.35
PID performance -10% [G]	0.62	0.29	0.49	0.42	0.53	0.38	0.35
PID performance -30% [G]	0.62	0.29	0.49	0.42	0.53	0.38	0.35
PID performance +5% [G]	0.62	0.30	0.49	0.42	0.53	0.37	0.35
PID performance +10% [G]	0.63	0.30	0.50	0.42	0.53	0.36	0.35
PID performance +30% [G]	0.63	0.30	0.50	0.42	0.53	0.36	0.35
Removal of soft tracks [N]	0.60	0.30	0.49	0.43	0.55	0.39	0.34
IsMuon efficiency Down [E]	0.61	0.29	0.49	0.43	0.54	0.38	0.34
IsMuon efficiency Up [E]	0.61	0.30	0.50	0.42	0.53	0.37	0.35
B^0 p re-weighting [I]	0.61	0.30	0.49	0.42	0.54	0.38	0.35
IP Smearing [H]	0.61	0.29	0.49	0.43	0.54	0.38	0.34
AC CTK Down [H]	0.63	0.31	0.51	0.44	0.56	0.39	0.36
AC CTK Up [C]	0.60	0.28	0.47	0.41	0.52	0.36	0.33
AC CTL Down [C]	0.63	0.30	0.51	0.44	0.56	0.39	0.35
AC CTL Up [C]	0.60	0.29	0.48	0.41	0.52	0.36	0.33
AC CTL Down CTK Down [C]	0.65	0.32	0.53	0.46	0.58	0.40	0.37
AC CTL Up CTK Up [C]	0.58	0.28	0.46	0.40	0.51	0.35	0.32
AC Non-factorisable Up [C]	0.61	0.29	0.49	0.42	0.54	0.38	0.34
AC Non-factorisable Down [C]	0.61	0.30	0.49	0.43	0.54	0.37	0.35
Tracking efficiency Down [F]	0.62	0.30	0.49	0.42	0.54	0.37	0.35
Tracking efficiency Up [F]	0.61	0.29	0.49	0.42	0.54	0.38	0.34
Trigger efficiency Down [D]	0.61	0.30	0.49	0.42	0.54	0.38	0.35
Trigger efficiency Up [D]	0.61	0.30	0.49	0.42	0.54	0.38	0.35
Signal mass width Up [J]	0.62	0.30	0.50	0.43	0.55	0.38	0.35
Signal mass width Down [J]	0.60	0.29	0.48	0.42	0.53	0.37	0.34
Peaking Bkg [L]	0.62	0.30	0.50	0.43	0.55	0.38	0.35

Table 9: Variation of $\mathrm{d}\mathcal{B}/\mathrm{d}q^2$ when systematically varying fit parameters or the weights applied to the input data set. The letter is a key corresponding to the text in Sec. 18.

The result of the differential branching fraction measurement in the six q^2-bins is shown in Fig. 9 .The SM prediction, and the prediction rate-averaged over the q^2 bin, are also indicated on the figure. No SM prediction is included for the region between the $c\bar{c}$ resonances where the assumptions made in the prediction break down.

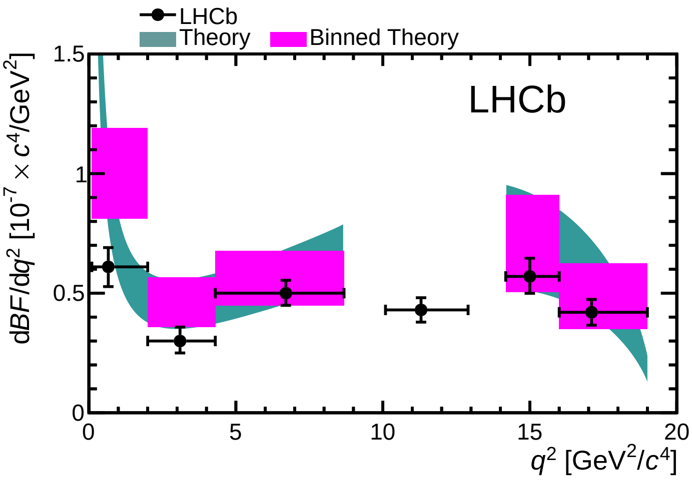


Figure 9: Differential branching fraction as a function of q^2. Points include both statistical and systematic uncertainties. The theory predictions are described in Ref. [19].

8 Signal angular distribution

8.1 Angular basis

$B^0 \to K^{*0}(\to K\pi)\mu^+\mu^-$ is treated as a pseudo-scalar to vector-vector decay and the angular distribution expressed in the Helicity angular basis (the decay amplitudes are however typically given as Transversity amplitudes). In this basis the decay of the B^0, K^{*0} and dimuon pair are each defined by a 'polar' and 'azimuthal' angle. Taking the decay of the K^{*0} as an example, the 'polar' angle is the angle between the K^+ direction in the rest frame of the K^{*0} and the direction of the K^{*0} in the rest frame of its parent, the B^0. The corresponding 'azimuthal' angle is a rotation of the plane containing the K^+ and π^- around the axis defined by the K^{*0} direction in the B^0 frame. This leads to an angular basis with six angles. In practice the physics content of the decay can be expressed in terms of just three: θ_ℓ, $\theta_K$and ϕ. The angle ϕ is the angle between the planes defined by the $\mu^+\mu^-$ and the $K\pi$ in the B^0 rest frame and is related to the 'azimuthal' angles of the K^{*0} and the dimuon in their respective frames. The transformation between the B^0 and $\overline{B}^0$ is made using the $\mathcal{CP}$ operator, i.e. by exchanging particles for their anti-particles and by reversing the particle momentum vectors.

8.1.1 Nomenclature

In the remainder of this note the momentum vector of a particle a in the rest frame of f is expressed as $\vec{p}_a^{\,f}$ and the sum of, and difference between, the momentum of two particles (a and b) in this frame as:

$$\vec{p}_{ab}^{\,f} = \vec{p}_a^{\,f} + \vec{p}_b^{\,f} \quad \text{and} \quad \vec{q}_{ab}^{\,f} = \vec{p}_a^{\,f} - \vec{p}_b^{\,f} \quad .$$

The unit normal vector to the plane containing a and b in the rest frame of f can then also be defined as:

$$\hat{n}_{ab}^{f} = \frac{\vec{p}_a^{\,f} \times \vec{p}_b^{\,f}}{|\vec{p}_a^{\,f} \times \vec{p}_b^{\,f}|} \quad .$$

8.1.2 The angle θ_ℓ

For the B^0 decay the angle θ_ℓ is defined by the angle between the vector defining the direction of the μ^+ in the dimuon rest frame and the direction of the dimuon in the B^0 rest frame. Equivalently this is the angle between the μ^+ and the direction opposite that of the B^0 in the dimuon rest frame:

$$\cos\theta_\ell = \frac{\vec{p}^{\,\mu\mu}_{\mu^+}\cdot\vec{p}^{\,B}_{\mu^+\mu^-}}{|\vec{p}^{\,\mu\mu}_{\mu^+}||\vec{p}^{\,B}_{\mu^+\mu^-}|}$$

or equivalently

$$\cos\theta_\ell = \frac{\vec{q}^{\,\mu\mu}_{\mu^+\mu^-}\cdot\vec{p}^{\,B}_{\mu^+\mu^-}}{|\vec{q}^{\,\mu\mu}_{\mu^+\mu^-}||\vec{p}^{\,B}_{\mu^+\mu^-}|} = -\frac{\vec{q}^{\,\mu\mu}_{\mu^+\mu^-}\cdot\vec{p}^{\,\mu\mu}_{B}}{|\vec{q}^{\,\mu\mu}_{\mu^+\mu^-}||\vec{p}^{\,\mu\mu}_{B}|} = -\frac{\vec{q}^{\,\mu\mu}_{\mu^+\mu^-}\cdot\vec{p}^{\,\mu\mu}_{K^+\pi^-}}{|\vec{q}^{\,\mu\mu}_{\mu^+\mu^-}||\vec{p}^{\,\mu\mu}_{K^+\pi^-}|} \quad .$$

For the $\overline{B}^0$ decay the angle is instead defined by the angle between the μ^- in the $\mu^+\mu^-$ rest frame and the direction of the dimuon pair in the rest frame of the $\overline{B}^0$:

$$\cos\theta_L = \frac{\vec{p}^{\,\mu\mu}_{\mu^-}\cdot\vec{p}^{\,B}_{\mu^+\mu^-}}{|\vec{p}^{\,\mu\mu}_{\mu^-}||\vec{p}^{\,B}_{\mu^+\mu^-}|} = -\frac{\vec{p}^{\,\mu\mu}_{\mu^+}\cdot\vec{p}^{\,B}_{\mu^+\mu^-}}{|\vec{p}^{\,\mu\mu}_{\mu^+}||\vec{p}^{\,B}_{\mu^+\mu^-}|} \quad .$$

8.1.3 The angle θ_K

For the $B^0/\overline{B}^0$ the angle θ_K is defined by the angle between the vector defining the direction of the K in the $K^{*0}/\overline{K}^{*0}$ rest frame and the direction of the $K^{*0}/\overline{K}^{*0}$ in the B rest frame:

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

or

$$\cos\theta_K = \frac{\vec{q}^{\,K\pi}_{K\pi}\cdot\vec{p}^{\,B}_{K\pi}}{|\vec{q}^{\,K\pi}_{K\pi}||\vec{p}^{\,B}_{K\pi}|} = -\frac{\vec{q}^{\,K\pi}_{K\pi}\cdot\vec{p}^{\,K\pi}_{B}}{|\vec{q}^{\,K\pi}_{K\pi}||\vec{p}^{\,K\pi}_{B}|} = -\frac{\vec{q}^{\,K\pi}_{K\pi}\cdot\vec{p}^{\,K\pi}_{\mu^+\mu^-}}{|\vec{q}^{\,K\pi}_{K\pi}||\vec{p}^{\,K\pi}_{\mu^+\mu^-}|} \quad .$$

8.1.4 The angle ϕ

The angle ϕ is given by the angle between the plane defined by the daughters of the dimuon and the daughters of the K^{*0}. In the case of the B^0 this is:

$$\cos\phi = \hat{n}^{B}_{\mu^+\mu^-}\cdot\hat{n}^{B}_{K^+\pi^-} \quad \text{and} \quad \sin\phi = \left(\hat{n}^{B}_{\mu^+\mu^-}\times\hat{n}^{B}_{K^+\pi^-}\right)\cdot\frac{\vec{p}^{\,B}_{K^+\pi^-}}{|\vec{p}^{\,B}_{K^+\pi^-}|}$$

For the $\overline{B}^0$ decay the $\mathcal{C}$ operator exchanges the μ^+ and μ^-. After applying the $\mathcal{P}$ to reverse the momentum directions:

$$\cos\phi = \hat{n}^B_{\mu^-\mu^+} \cdot \hat{n}^B_{K^-\pi^+} = -\hat{n}^B_{\mu^+\mu^-} \cdot \hat{n}^B_{K^-\pi^+}$$

as the $\mathcal{P}$ operator leaves $\hat{n}^B_{\mu^-\mu^+}$ unchanged:

$$\mathcal{P}(\hat{n}^B_{\mu^-\mu^+}) = \hat{n}^B_{\mu^-\mu^+}$$

and

$$\sin\phi = -\left(\hat{n}^B_{\mu^-\mu^+} \times \hat{n}^B_{K^-\pi^+}\right) \cdot \frac{\vec{p}^{\,B}_{K^-\pi^+}}{|\vec{p}^{\,B}_{K^-\pi^+}|} = +\left(\hat{n}^B_{\mu^+\mu^-} \times \hat{n}^B_{K^-\pi^+}\right) \cdot \frac{\vec{p}^{\,B}_{K^-\pi^+}}{|\vec{p}^{\,B}_{K^-\pi^+}|} \quad .$$

8.2 Differential angular distribution

The differential angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ candidates when neglecting terms proportional to $\sqrt{m_\mu^2/q^2}$ or m_μ^2/q^2 is given by:

$$\begin{aligned}
\frac{\mathrm{d}^4\Gamma[B^0 \to K^{*0}\mu^+\mu^-]}{d\cos\theta_\ell\, d\cos\theta_K\, d\phi\, dq^2} = \frac{9}{32\pi} \big[& I_1^s \sin^2\theta_K + I_1^c \cos^2\theta_K + \\
& (I_2^s \sin^2\theta_K + I_2^c \cos^2\theta_K)\cos 2\theta_\ell + \\
& I_3 \sin^2\theta_K \sin^2\theta_\ell \cos 2\phi + I_4 \sin 2\theta_K \sin 2\theta_\ell \cos\phi + \\
& I_5 \sin 2\theta_K \sin\theta_\ell \cos\phi + I_6 \sin^2\theta_K \cos\theta_\ell + \\
& I_7 \sin\theta_\ell \sin 2\theta_K \sin\phi + I_8 \sin 2\theta_K \sin 2\theta_\ell \sin\phi + \\
& I_9 \sin^2\theta_K \sin^2\theta_\ell \sin 2\phi \ \big]
\end{aligned}$$

where I_1 through I_9 are:

$$
\begin{aligned}
I_1^c &= \left(|A_{0L}|^2 + |A_{0R}|^2\right) \\
I_1^s &= \frac{3}{4}\left(|A_{\parallel L}|^2 + |A_{\parallel R}|^2 + |A_{\perp L}|^2 + |A_{\perp R}|^2\right) \\
I_2^c &= -\left(|A_{0L}|^2 + |A_{0R}|^2\right) \\
I_2^s &= \frac{1}{4}\left(|A_{\parallel L}|^2 + |A_{\parallel R}|^2 + |A_{\perp L}|^2 + |A_{\perp R}|^2\right) \\
I_3 &= \frac{1}{2}\left(|A_{\perp L}|^2 - |A_{\parallel L}|^2 + |A_{\perp R}|^2 - |A_{\parallel R}|^2\right) \\
I_4 &= \frac{1}{\sqrt{2}}\left(Re(A_{0L}A^*_{\parallel L}) + Re(A_{0R}A^*_{\parallel R})\right) \\
I_5 &= \sqrt{2}\left(Re(A_{0L}A^*_{\perp L}) - Re(A_{0R}A^*_{\perp R})\right) \\
I_6 &= 2\left(Re(A_{\parallel L}A^*_{\perp L}) - Re(A_{\parallel R}A^*_{\perp R})\right) \\
I_7 &= \sqrt{2}\left(Im(A_{0L}A^*_{\parallel L}) - Im(A_{0R}A^*_{\parallel R})\right) \\
I_8 &= \frac{1}{\sqrt{2}}\left(Im(A_{0L}A^*_{\perp L}) + Im(A_{0R}A^*_{\perp R})\right) \\
I_9 &= \left(Im(A_{\parallel L}A^*_{\perp L}) + Im(A_{\parallel R}A^*_{\perp R})\right)
\end{aligned}
$$

i.e. they depend on the K^{*0} transversity amplitudes, which in turn are sensitive to the contributions from NP. The L and R labels on the K^{*0} transversity amplitudes refer to the chirality of the lepton current, which can be both left- and right-handed.

Neglecting terms proportional to m_μ^2/q^2 and possible scalar and tensor amplitudes there are 6 complex amplitudes that appear in I_1 through I_9. In the most general case there would be $6+3(\text{tensor})+1(\text{scalar})+1(time-like)$ complex amplitudes.

The addition of a broad S-wave, with $K\pi$ system in a spin 0 state, modifies terms in $I_{1\dots9}$ according to:

$$
A_{0L,R}Y_1^0(\theta_K) \to \sum_{J=0,1} A_0^J Y_J^0(\theta_K)
$$

where the index J refers to the spin of the $K\pi$ system and the $Y_J^0(\theta_K)$ are spherical harmonics. There is no contribution from the S-wave to terms in $A_{\parallel L,R}$ and $A_{\perp L,R}$ (because these correspond to transverse polarisation of the $K^+\pi^-$ system). The S-wave contribution to the angular observables is discussed in detail below.

8.3 Combining B^0 and $\overline{B}^0$ decays

The angular basis has been defined starting with the B^0 decay and applying the CP transformation to go from the B^0 to the $\overline{B}^0$ decay. As a result, neglecting any production, detector or direct CP asymmetry, the combined angular distribution for the B^0 and the $\overline{B}^0$ is given by:

$$\frac{d[B^0 + \overline{B}^0]}{d\cos\theta_\ell \, d\cos\theta_K \, d\phi \, dq^2} = \frac{9}{32\pi} \sum_{i=1}^{9} (I_i + \bar{I}_i) f_i(\cos\theta_\ell, \cos\theta_K, \phi)$$

This is a different angular basis to the one that often appears in literature. Using the nomenclature of Ref. [11], this corresponds to describing the angular distribution by a sum of S_1 to S_9 when combining B^0 and $\overline{B}^0$ decays.

8.3.1 CP averages and CP asymmetries (A_9 vs S_9)

Whilst the angular basis differs from the theory convention, it is identical to that of BaBar, Belle and CDF for the angles θ_ℓ and θ_K. It does however differ from the CDF ϕ angle definition in Ref. [7]. The CDF ϕ definition does not obey the CP transformation needed to measure S_9. Instead under the CDF definition, the difference between B^0 and $\overline{B}^0$ decays is measured for all terms that are 'odd' in ϕ (terms 7, 8 and 9). Consequently under the CDF definition, for example, A_9 appears in place of S_9 in the angular distribution. Explicitly, in the absence of any production, detector or direct CP asymmetry:

$$S_9 = \frac{1}{2}\left(I_9 + \bar{I}_9\right) \text{ and } A_9 = \frac{1}{2}\left(I_9 - \bar{I}_9\right) \quad .$$

If production, detection or direct CP asymmetries become large then there will be a mixing between S_9 and A_9. This effect is neglected in this analysis. The angular distributions and the PID likelihoods for kaons and pions are compared for B^0 and $\overline{B}^0$, using the decay $B^0 \to K^{*0} J/\psi$, as shown in Appendix C. No significant discrepancy has been observed.

The observable A_9 is a T-odd CP asymmetry. This has little meaning for this self-tagging decay, but A_9 could, for example, also be measured in decays $B^0_s \to \phi\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ ($K^{*0} \to K^0_S\pi^0$) where it is not possible to unambiguously separate the B and $\overline{B}$ decays.

In terms of NP sensitivity the principle difference between S_9 and A_9 is that:

$$S_9 \propto \cos\lambda \sin\delta \quad \text{and} \quad A_9 \propto \sin\lambda \cos\delta \ ,$$

where δ is a strong phase and λ is the contribution from the weak phase. In the SM both the strong phase and the weak phase are small (the weak phase contribution comes from V_{ts}) and so A_9 and $S_9 \sim 0$. S_9 remains small in NP models. It is possible to fit for A_9 in place of S_9 in the LHCb convention by swapping the sign of ϕ ($\phi \rightarrow -\phi$) for $\overline{B}^0$ decays only. To avoid confusion below, the notation A_{Im} is adopted to refer to either S_9 or A_9 in the angular distribution.

While the principal difference between S_9 and A_9 is a simple sign change of the ϕ angle for B^0 and $\overline{B}^0$ decays, it has important experimental consequences. When measuring S_9, there is a need to understand the combined acceptance correction for the combination of B^0 and $\overline{B}^0$ decays. Conversely when measuring A_9 there is a need to understand the difference between the acceptance correction for B^0 and $\overline{B}^0$ decays.

If there were to be a significant production, detection or direct CP asymmetry between the B^0 and $\overline{B}^0$ that results in a different number of B^0 and $\overline{B}^0$ decays appearing in the angular analysis, then this would lead to a mixing between the A's and S's:

$$A_i^{\text{measured}} \approx A_i - S_i(A_{CP} + A_D + \kappa A_P)$$

where A_P is the B^0-$\overline{B}^0$ production asymmetry, A_D, the detection asymmetry, A_{CP} the direct CP asymmetry and κ is a factor to account for the dilution of A_P due to mixing.

8.4 Folding the ϕ-angle

The differential branching fraction can be greatly simplified by "folding" the ϕ-angle such that $\hat{\phi} = \phi + \pi$ if $\phi < 0$. This cancels terms with with $\cos\phi$ and $\sin\phi$ dependencies (but not $\cos 2\phi$ and $\sin 2\phi$), i.e. the terms I_4, I_5, I_6 and I_8 above. This cancellation dramatically simplifies the angular expression and leaves sensitivity to F_L, A_{FB} (through I_6), A_T^2 (through I_3) and A_{Im} (through I_9).

This simplification leads to:

$$\frac{1}{\Gamma}\frac{\mathrm{d}^4\Gamma}{\mathrm{d}\cos\theta_\ell\,\mathrm{d}\cos\theta_K\,\mathrm{d}\hat{\phi}\,\mathrm{d}q^2} = \frac{9}{16\pi}\Bigg[F_L\cos^2\theta_K + \frac{3}{4}F_T(1-\cos^2\theta_K) + \\ \frac{1}{4}F_T(1-\cos^2\theta_K)\cos 2\theta_\ell - \\ F_L\cos^2\theta_K\cos 2\theta_\ell \;+ \\ S_3(1-\cos^2\theta_\ell)(1-\cos^2\theta_K)\cos 2\hat{\phi} \;+ \\ \frac{4}{3}A_{FB}(1-\cos^2\theta_K)\cos\theta_\ell + \\ A_{Im}(1-\cos^2\theta_K)(1-\cos^2\theta_\ell)\sin 2\hat{\phi} \;\Bigg]$$

where A_{FB}, F_L, A_T^2 and A_{Im} are:

$$A_{FB} = \frac{3}{2}\frac{Re(A_{\parallel L}A^*_{\perp L}) - Re(A_{\parallel R}A^*_{\perp R})}{|A_{0L}|^2 + |A_{\parallel L}|^2 + |A_{\perp L}|^2 + |A_{0R}|^2 + |A_{\parallel R}|^2 + |A_{\perp R}|^2}$$

$$F_L = \frac{|A_{0L}|^2 + |A_{0R}|^2}{|A_{0L}|^2 + |A_{\parallel L}|^2 + |A_{\perp L}|^2 + |A_{0R}|^2 + |A_{\parallel R}|^2 + |A_{\perp R}|^2} = 1 - F_T$$

$$A_{Im} = \frac{Im(A_{\parallel L}A^*_{\perp L}) + Im(A_{\parallel R}A^*_{\perp R})}{|A_{0L}|^2 + |A_{\parallel L}|^2 + |A_{\perp L}|^2 + |A_{0R}|^2 + |A_{\parallel R}|^2 + |A_{\perp R}|^2}$$

$$S_3 = \frac{1}{2}\frac{|A_{\perp L}|^2 - |A_{\parallel L}|^2 + |A_{\perp R}|^2 - |A_{\parallel R}|^2}{|A_{0L}|^2 + |A_{\perp L}|^2 + |A_{\parallel L}|^2 + |A_{0R}|^2 + |A_{\perp R}|^2 + |A_{\parallel R}|^2}$$

A_{FB} and A_{Im} can both in principal take different values for B^0 and $\overline{B}^0$ decays.

8.5 Angular projections

It is also possible (as described in the previous analysis note [13]) to have sensitivity to these observables by integrating the full differential angular distribution over all but one of the angles. This leads to:

$$\frac{1}{\Gamma}\frac{\mathrm{d}^2\Gamma}{\mathrm{d}\cos\theta_\ell\,\mathrm{d}q^2} = \frac{3}{4}F_L(1-\cos^2\theta_\ell) + \frac{3}{8}(1-F_L)(1+\cos^2\theta_\ell) + A_{FB}\cos\theta_\ell \quad ,$$

$$\frac{1}{\Gamma}\frac{\mathrm{d}^2\Gamma}{\mathrm{d}\cos\theta_K\,\mathrm{d}q^2} = \frac{3}{2}F_L\cos^2\theta_K + \frac{3}{4}(1-F_L)(1-\cos^2\theta_K)$$

and

$$\frac{1}{\Gamma}\frac{\mathrm{d}^2\Gamma}{\mathrm{d}\phi\,\mathrm{d}q^2} = \frac{1}{2\pi}\left[1 + S_3\cos 2\phi + A_{Im}\sin 2\phi\right]$$

The angular distribution in $\cos\theta_K$ depends only on a single parameter F_L, the fraction of longitudinally polarised K^{*0}. The distribution in $\cos\theta_L$ has two free parameters F_L and A_{FB}, the forward-backward asymmetry of the muons in the dimuon rest frame. The angle ϕ depends on F_L, S_3 and A_{Im}.

8.6 Re-parametrisation using A_T^{Re} and A_T^{Im}

It has for a long-time been suggested in the theory literature that the quantity:

$$A_T^2 = \frac{|A_{\perp L}|^2 - |A_{\parallel L}|^2 + |A_{\perp R}|^2 - |A_{\parallel R}|^2}{|A_{\perp L}|^2 + |A_{\parallel L}|^2 + |A_{\perp R}|^2 + |A_{\parallel R}|^2}$$

is a cleaner observable than S_3 because it is free from $|A_{0(L/R)}|^2$ and therefore has a reduced form factor uncertainty. This can be extracted from a fit to the data by replacing S_3 by:

$$S_3 = \frac{1}{2}(1-F_L)A_T^2 \quad .$$

It has also been suggested in Ref.[12] that:

$$A_T^{Re} = 2.\frac{Re(A_{\parallel L}A^*_{\perp L}) - Re(A_{\parallel R}A^*_{\perp R})}{|A_{\parallel L}|^2 + |A_{\perp L}|^2 + |A_{\parallel R}|^2 + |A_{\perp R}|^2}$$

is theoretically a cleaner observable than A_{FB} as it does not depend on $\Gamma = |A_{0L}|^2 + |A_{\parallel L}|^2 + |A_{\perp L}|^2 + |A_{0R}|^2 + |A_{\parallel R}|^2 + |A_{\perp R}|^2$ and instead only contains $A_\parallel$ and $A_\perp$ (reducing hadronic uncertainties). It is also interesting to note that this implies:

$$A_{FB} = \frac{3}{4}F_T A_T^{Re} = \frac{3}{4}(1-F_L)A_T^{Re} \quad .$$

From the expression for the projection of $\cos\theta_\ell$, if $\cos\theta_\ell \to \pm 1$ then:

$$\frac{1}{\Gamma}\frac{\mathrm{d}^2\Gamma}{\mathrm{d}\cos\theta_\ell\,\mathrm{d}q^2} \to \frac{3}{4}(1-F_L) \pm A_{FB} \quad .$$

For $(1\Gamma)(\mathrm{d}^2\Gamma/\mathrm{d}\cos\theta_{\ell\,\mathrm{d}q^2}$ to remain positive for all values of $\cos\theta_l$ then $A_{FB} \leq \frac{3}{4}(1-F_L)$. This requirement is automatically enforced by A_T^{Re} if $-1 < A_T^{Re} < 1$. A similar observable can be found to replace A_{Im}:

$$A_T^{Im} = 2.\frac{Im(A_{\parallel L}A_{\perp L}^*) + Im(A_{\parallel R}A_{\perp R}^*)}{|A_{\parallel L}|^2 + |A_{\perp L}|^2 + |A_{\parallel R}|^2 + |A_{\perp R}|^2}$$

such that:

$$A_{Im} = \frac{1}{2}F_T A_T^{Im} = \frac{1}{2}(1-F_L)A_T^{Im}$$

which simplifies the fit. A constraint still exists between A_T^{Re}, A_T^2 and A_T^{Im} which can not simply be expressed. The effect of such a re-parametrisation can be seen in Fig. 10 where the regions of phase-space in which the fit pdf can go negative are shown for $A_{FB} = 0.1$ and $F_L = 0.8$. These values are similar to the results of the previous analysis [14] in the region $2 < q^2 < 4.3$. The left plot indicates these regions when fitting with A_{Im}, the right plot when fitting with A_T^{Im}. It is clear that the valid phase-space is larger using the A_T^{Im} observable.

8.7 Observable discussion

The physics observables in the angular distribution are all q^2 dependent. In practice what is measured when using a wide bin of q^2 is the rate average of each of the observables over the q^2 bin. So for example,

$$\langle F_{\mathrm{L}}\rangle = \int_{q^2_{\mathrm{min}}}^{q^2_{\mathrm{max}}} \frac{1}{\Gamma}\frac{d\Gamma}{dq^2}F_{\mathrm{L}}(q^2)dq^2$$

$$\langle A_{\mathrm{FB}}\rangle = \int_{q^2_{\mathrm{min}}}^{q^2_{\mathrm{max}}} \frac{1}{\Gamma}\frac{d\Gamma}{dq^2}A_{\mathrm{FB}}(q^2)dq^2$$

The situation is more complicated for terms in the angular expression that contain the product of two q^2-dependent "observables". This includes A_T^{Re}, A_T^{Im}, when re-parameterising the angular distribution and A_T^2. Here, the fit is sensitive to e.g.:

$$\left\langle(1-F_{\mathrm{L}})A_{\mathrm{T}}^2\right\rangle = \int_{q^2_{\mathrm{min}}}^{q^2_{\mathrm{max}}} \frac{1}{\Gamma}\frac{d\Gamma}{dq^2}(1-F_{\mathrm{L}}(q^2)A_{\mathrm{T}}^2(q^2)dq^2$$

which is not the same as the product of the two q^2-averaged values:

$$\left\langle(1-F_{\mathrm{L}})A_{\mathrm{T}}^2\right\rangle \neq \langle(1-F_{\mathrm{L}})\rangle \times \left\langle A_{\mathrm{T}}^2\right\rangle$$

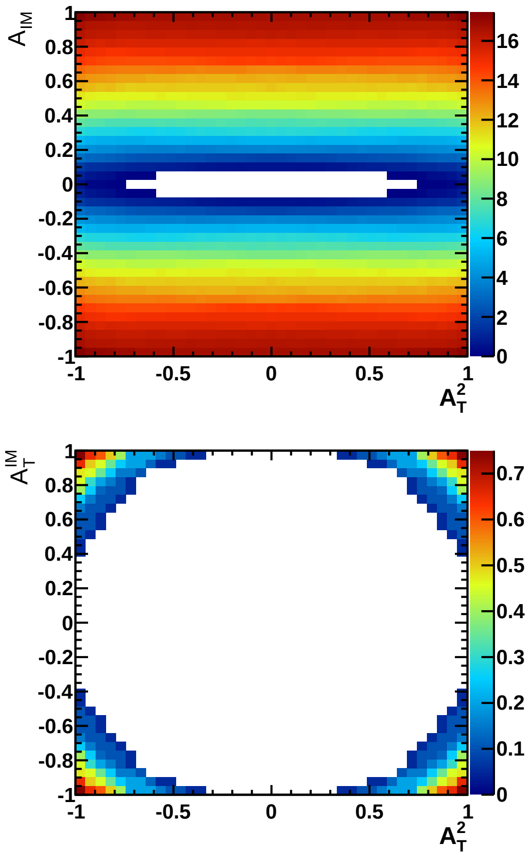


Figure 10: Comparison of the fraction of the pdf that is invalid in regions of phase-space when fitting with the observables A_{Im} (left) and A_T^{Im} (right). The observable A_T^{Im} has a significantly larger valid region, increasing the stability of fits.

unless one of the observables is constant over the q^2-bin. An unfortunate consequence is that the measured quantities, coming from the maximum likelihood fit are not exactly the same as the quantity that is predicted by theorists. They will however tend to be similar unless the q^2-dependence of both of F_{L} and the observable is large.

However the integrated averaged transverse observables which are fitted on data can be compared with well defined quantities that theorists can predict. One has

$$\langle A_{\mathrm{FB}} \rangle = \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma}{dq^2} A_{\mathrm{FB}}(q^2) dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma}{dq^2} dq^2} = \frac{3}{4} \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma}{dq^2} A_{\mathrm{T}}^{Re}(q^2)(1 - F_{\mathrm{L}}(q^2)) dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma}{dq^2} dq^2}$$

$$= \frac{3}{4} \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma}{dq^2} A_{\mathrm{T}}^{Re}(q^2)(1 - F_{\mathrm{L}}(q^2)) dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma(1-F_{\mathrm{L}}(q^2))}{dq^2} dq^2} \times \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma(1-F_{\mathrm{L}}(q^2))}{dq^2} dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma}{dq^2} dq^2}$$

$$= \frac{3}{4} \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma}{dq^2} A_{\mathrm{T}}^{Re}(q^2)(1 - F_{\mathrm{L}}(q^2)) dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma(1-F_{\mathrm{L}}(q^2))}{dq^2} dq^2} \times (1 - \langle F_{\mathrm{L}} \rangle)$$

One can then define

$$\left\langle \tilde{A_{\mathrm{T}}^{Re}} \right\rangle = \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma}{dq^2} A_{\mathrm{T}}^{Re}(q^2)(1 - F_{\mathrm{L}}(q^2)) dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma(1-F_{\mathrm{L}}(q^2))}{dq^2} dq^2} = \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma_T}{dq^2} A_{\mathrm{T}}^{Re}(q^2) dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma_T}{dq^2} dq^2}$$

which can be computed from theoretical models. Similarly one can compare the fitted values of $A_{\mathrm{T}}^{(2)}$ and A_{T}^{Im} with

$$\left\langle \tilde{A_{\mathrm{T}}^{(2)}} \right\rangle = \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma_T}{dq^2} A_{\mathrm{T}}^{(2)}(q^2) dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma_T}{dq^2} dq^2}$$

and

$$\left\langle \tilde{A_{\mathrm{T}}^{Im}} \right\rangle = \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma_T}{dq^2} A_{\mathrm{T}}^{Im}(q^2) dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma_T}{dq^2} dq^2}$$

It has been checked with a very large Monte-Carlo sample that the result of the fit of a given transverse variable in a q^2 bin is actually equal to a high accuracy to the average given above.

9 Measurement of angular observables with likelihood fit

9.1 Background angular model

The background angular model is assumed to be factorisable into three one-dimensional angular distributions. The full angular model is then given by:

$$\begin{aligned} P_{\text{bkg.}}(\cos\theta_l, \cos\theta_K, \phi) &= P_{\text{bkg.}}(\cos\theta_l) P_{\text{bkg.}}(\cos\theta_K) P_{\text{bkg.}}(\phi) \\ &= \left(\sum_{k=0}^{n} c_k^l T_k(\cos\theta_l)\right) \left(\sum_{k=0}^{n} c_k^K T_k(\cos\theta_K)\right) \left(\sum_{k=0}^{n} c_k^\phi T_k(\phi)\right) \end{aligned}$$

where T_k is a k^{th} order Chebychev polynomial of the first kind. The angular distribution is assumed to be independent of the $K^+\pi^-\mu^+\mu^-$ invariant mass for $m_{K\pi\mu^+\mu^-} > 5150\,\text{MeV}/c^2$.

In the likelihood fit for the angular observables, the background shapes in each of the angles are limited to $\mathcal{O}(2)$ (i.e. they are parabolic). Higher order background shapes are investigated as a potential source of systematic uncertainty.

The factorisation assumption is validated using events in the upper $K^+\pi^-\mu^+\mu^-$ mass sideband and a point-to-point dissimilarity test [1] to form an unbinned comparison of the angular model and the data. The probability of the test statistic being smaller than the value observed for the data is 25% (Fig. 11).

9.2 Background distribution in the sidebands

The q^2-distribution of events in the lower (defined as $5150 < m_{K^+\pi^-\mu^+\mu^-} < 5220\,\text{MeV}/c^2$) and upper ($5350 < m_{K^+\pi^-\mu^+\mu^-} < 5800\,\text{MeV}/c^2$) mass sidebands are shown in Fig. 12(a). The χ^2 probability for the normalised distributions of the left and right sidebands to come from the same parent distribution is 30%, i.e. the two sidebands are statistically compatible with each other. This is an important check for the method used for the extraction of the zero-crossing point described in section 21.

The angular fit is done independently for the different bins of q^2, therefore it is not strictly required that the q^2 distribution is the same for the two sidebands. However, it is assumed that the sideband angular distributions describe the combinatorial background in the signal region. Figs. 12 (b), (c) and (d) show the comparison between the angular distributions for the left and the right sideband. The χ^2 probability for the angular distributions of the two sidebands ranges from 16% to 60%. The angular distributions of

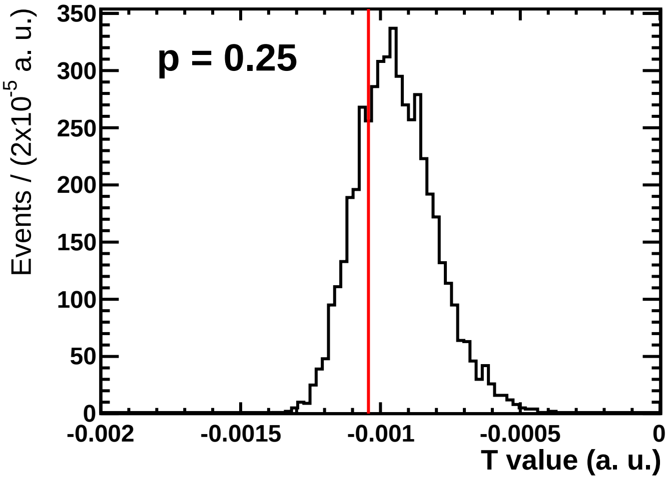
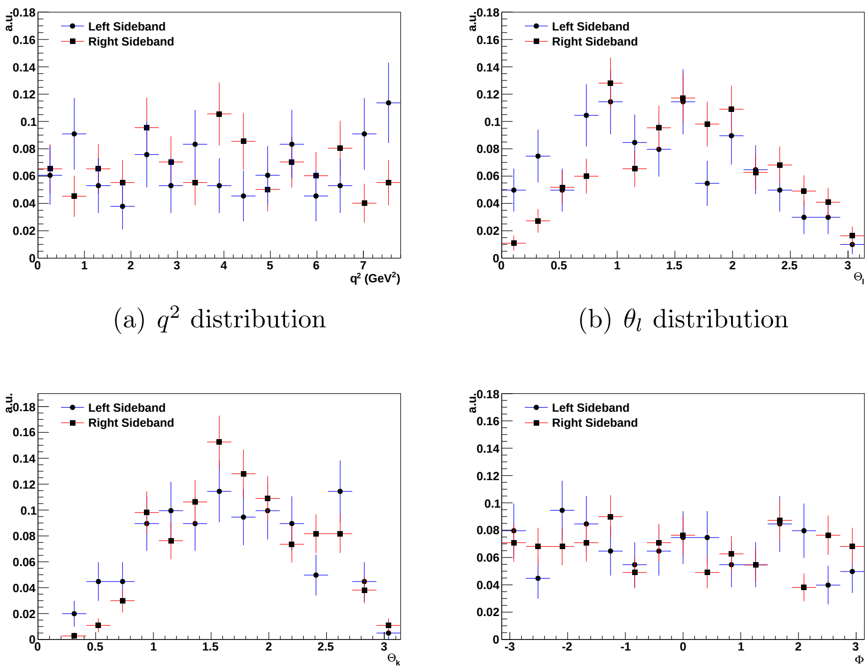


Figure 11: Distribution of the test statistic, T, from a point-to-point dissimilarity test made using the factorised background angular model in the upper mass sideband. The distribution from toy experiments is shown by the curve and the value in data by the vertical line. The probability, $P(T \leq T_{\mathrm{data}}) = 25\%$.

the two sidebands are therefore also statistically compatible with each other. This also demonstrates that there is no anomalous contamination of double semi-leptonic decays in the low-mass sideband (and by extension the signal region).

If the lower mass sideband is extended down to a $K^+ \pi^- \mu^+\mu^-$ invariant mass of $5000\,\mathrm{MeV}/c^2$, there is no longer good agreement between the background angular and q^2 distribution between the upper and lower (left- and right-) mass sidebands. This is expected due to contamination from double semi-leptonic decays and partially reconstructed backgrounds.



(a) q^2 distribution

(b) θ_l distribution

(c) θ_k distribution

(d) ϕ distribution

Figure 12: Comparison between the left and the right sideband for the q^2 and the angular distributions.

9.3 Angular resolution

The signal angular resolution is studied using simulated events. The resolution in θ_K, θ_ℓ and ϕ in physics MC (in the q^2 range $4m_{\mu^2} < q^2 < 19\,\mathrm{GeV}^2/c^4$) is shown in Fig. 13. The resolution is sufficiently good to have a negligible impact on the signal angular fit. No large dependence of the resolution on q^2 is seen.

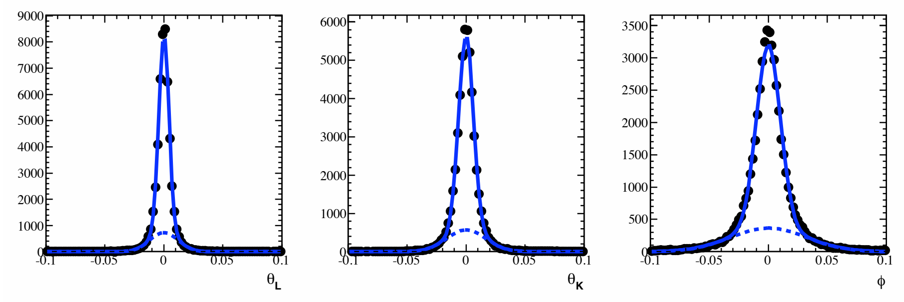


Figure 13: Signal angular resolution in θ_K, θ_ℓ and ϕ as measured using SM-like simulated events.

9.4 $B^0 \leftrightarrow \overline{B}^0$ mis-identification

If a $\overline{B}^0$ decay is mis-identified as a B^0 decay by exchanging the kaon and pion, then $\cos\theta_\ell \to -\cos\theta_\ell$, $\cos\theta_K \to -\cos\theta_K$ and $\phi \to -\phi$. This exchange has dilutes the measured forward-backward asymmetry and A_{Im}, but has no impact on A_T^2 and F_L.

$$A_{FB} \to (1 - 2\omega_{\rm ID})A_{FB}$$

$$A_{Im} \to (1 - 2\omega_{\rm ID})A_{Im}$$

for a $B^0 \leftrightarrow \overline{B}^0$ (equivalently K^{*0} to $\overline{K}^{*0}$) mis-identification pro ability of $\omega_{\rm ID}$. This dilution would be exact if kaon and pion mass were identical. In practice $m_K > m_\pi$ means that the angular distribution in $\cos\theta_K$ is not identical to the distribution of the signal (exchanging $\cos\theta_K \to -\cos\theta_K$). From Sec. 3.4, $\omega_{\rm ID}$ is estimated to be $0.85 \pm 0.02\%$. The mis-identification probability is kept constant in the fit, but will be varied as a source of systematic uncertainty.

9.5 Physical boundaries for angular observables

Tables 10 and 11 below outline the physical ranges of the parameters used in the angular analysis. The table also indicates which variables are at some level intrinsically correlated. For example, $A_T^{Re.}$, $A_{\rm T}^2$ and $A_T^{Im.}$ are all related through $A_{\parallel L,R}$ and $A_{\perp L,R}$. There are three choices of "physics" parameters:

1. Transverse observables ($F_{\rm L}$, $A_{\rm T}^2$, $A_T^{Re.}$ and $A_T^{Im.}$);
2. $F_{\rm L}$, $A_{\rm FB}$, S_3 and S_9 ;
3. $F_{\rm L}$, $A_{\rm FB}$, S_3 and A_9 .

In Table. 10, A_{Im} refers to both S_9 and A_9.

Parameter	Range	Comments
A_{FB}	$-\frac{3}{4} < A_{\text{FB}} < \frac{3}{4}$	Parameter correlated to F_{L}, S_3 and A_{Im}
S_3	$-\frac{1}{2} < S_3 < \frac{1}{2}$	Parameter correlated to F_{L}, A_{FB} and A_{Im}
A_{Im}	$-1 < A_{\text{Im}} < 1$	Parameter correlated to F_{L}, S_3 and A_{FB}
F_{L}	$0 < F_{\text{L}} < 1$	Parameter correlated to A_{FB}, A_{Im} and S_3

Table 10: The "physics" parameters, their allowed ranges and correlations with the other physics parameters.

Parameter	Range	Comments
$A_T^{Re.}$	$-1 < A_T^{Re.} < 1$	Parameter correlated to A_{T}^2 and $A_T^{Im.}$
A_T^{Im}	$-1 < A_T^{Im} < 1$	Parameter correlated to A_{T}^2 and $A_T^{Re.}$
A_{T}^2	$-1 < A_{\text{T}}^2 < 1$	Parameter correlated to $A_T^{Re.}$ and $A_T^{Im.}$
F_{L}	$0 < F_{\text{L}} < 1$	Parameter un-correlated to other parameters

Table 11: The "transverse" parameters, their allowed ranges and correlations with the other physics parameters.

In many cases the physical ranges also correspond to a mathematical boundary. Beyond the physical range the PDF describing the signal can become negative. For example a larger value of A_T^{Re} can make the PDF negative at $\cos\theta_l \sim \pm 1$. When A_{FB}, F_{L}, A_{Im} and $S_3 = \frac{1}{2}A_T^2(1 - F_L)$ are used as the choice of variables, there are mathematical boundaries that require:

$$\begin{aligned} A_{\text{FB}} \leq& \frac{3}{4}(1 - F_{\text{L}}) \quad , \\ A_{\text{Im}} \leq& \frac{1}{2}(1 - F_{\text{L}}) \quad , \\ S_3 \leq& \frac{1}{2}(1 - F_{\text{L}}) \quad . \end{aligned}$$

These constraints can be seen directly in the differential angular distribution and in the expression for A_{FB} in terms of the transversity amplitudes. If $|A_0|^2 \to 1$, then $|A_\parallel|^2$ and $|A_\perp|^2 \to 0$ and $A_{\text{FB}} = 0$. A similar constraint exists between A_{Im} and F_{L}, $A_{\text{Im}} \leq \frac{1}{2}(1 - F_{\text{L}})$. There are also non-trivial boundary effects between A_{FB}, A_{Im} and S_3, that cannot be expressed easily.

9.6 Unbinned maximum likelihood fit for the angular observables

The signal fit parameters are estimated by performing an unbinned maximum likelihood fit to the data, weighting the candidates to account for the detector acceptance. The acceptance weights are defined as the inverse of the efficiency and they are applied in an even-by-event basis. The efficiency for each event is extracted as a function of the three angles and q^2 using phase space MC simulation. This procedure is described in detail in Sec. 11. Multiple candidates are also accounted for by weighting each candidate by the inverse of the number of candidates in each event. In practice, the log-likelihood,

$$
\begin{aligned}
-\log L = -\sum_{i=0}^{N} \alpha\omega_i \log \Big[& f_{\mathrm{sig}} P_{\mathrm{sig.}}(m_{K^+\pi^-\mu^+\mu^-}, \vec{\Omega}_i; \vec{\lambda}_{\mathrm{sig}}) \;\; + \\
& (1 - f_{\mathrm{sig}}) P_{\mathrm{bkg.}}(m_{K^+\pi^-\mu^+\mu^-}, \vec{\Omega}_i, \vec{\lambda}_{\mathrm{bkg}}) \Big]
\end{aligned}
$$

is minimised, where $\vec{\lambda}_{\mathrm{sig}}$ are the physics parameters, f_{sig} is the signal fraction and $\vec{\Omega} = (\cos\theta_l, \cos\theta_K, \phi)$. The weights, ω_i are normalised such that the sum of the weights is the number of candidates, i.e.

$$
\sum_{i=0}^{N} \alpha\omega_i = N \quad .
$$

in each q^2 bin, where α is a scale-factor used to normalise the weights. With this normalisation the weighted "pseudo-likelihood" has a habit of under-covering. This is due to the fact that the correct scaling of the log likelihood is distorted by the weights. Unfortunately the normalisation applied is only a first order correction. Toy Monte Carlo studies showed that the under-coverage is approximately given by $\sum w_i^2 / \sum w_i$, which in our case corresponds to a correction to the error of about 10%.

The full signal PDF is given by:

$$
\begin{aligned}
P_{\mathrm{sig}}(m_{K^+\pi^-\mu^+\mu^-}, \vec{\Omega}_i, \vec{\lambda}_{\mathrm{bkg}}) = & M(m_{K^+\pi^-\mu^+\mu^-} | \sigma_1, \sigma_2, \alpha, n) \times \\
& \left(\int_{q^2_{\mathrm{min}}}^{q^2_{\mathrm{max}}} \frac{1}{\Gamma} \frac{d^4\Gamma}{dq^2\, d\cos\theta_l\, d\cos\theta_K\, d\phi} dq^2 \right)
\end{aligned}
$$

where the signal angular distribution is averaged over the q^2-bin. The background PDF is given by:

$$P_{\text{bkg}}(m_{K^+\pi^-\mu^+\mu^-}, \vec{\Omega}_i, \vec{\lambda}_{\text{bkg}}) = E(m_{K^+\pi^-\mu^+\mu^-}|p_0) \times \left(\sum_{k=0}^{n} c_k^l T_k(\cos\theta_l)\right) \left(\sum_{k=0}^{n} c_k^K T_k(\cos\theta_K)\right) \left(\sum_{k=0}^{n} c_k^\phi T_k(\phi)\right)$$

where the background angular distribution is parametrised as the product of three Chebychev polynomials (of the first kind).

Details of the fit performed in data and of the error computation are given in Sec. 15.

9.7 Free parameters in the likelihood fit

In addition to the 4 physics parameters, there are 8 further free parameters in each of the likelihood fits. The free parameters are summarised in the Table. 12.

Number	Names	Description
2	c_1^l, c_2^l	Parameters describing the background in $\cos\theta_l$
2	c_1^K, c_2^K	Parameters describing the background in $\cos\theta_K$
2	c_1^ϕ, c_2^ϕ	Parameters describing the background in ϕ
1	p_0	Parameter describing the background mass distribution
1	$f_{\text{sig.}}$	Signal fraction

Table 12: Description of the free parameters in the log-likelihood fit for the angular observables.

10 Data-MC corrections

The MC samples used to estimate the contribution from peaking backgrounds and detector / selection acceptance effects have been corrected for data-MC differences. These differences are corrected for in two different ways, depending on whether or not the correction is required before the application of the BDT. If the variable is not present in the BDT, the MC is re-weighted to account for data-MC differences. If the variable is used in the BDT the variable is adjusted (or replaced) before the application of the BDT. Variables that are used in the BDT include the:

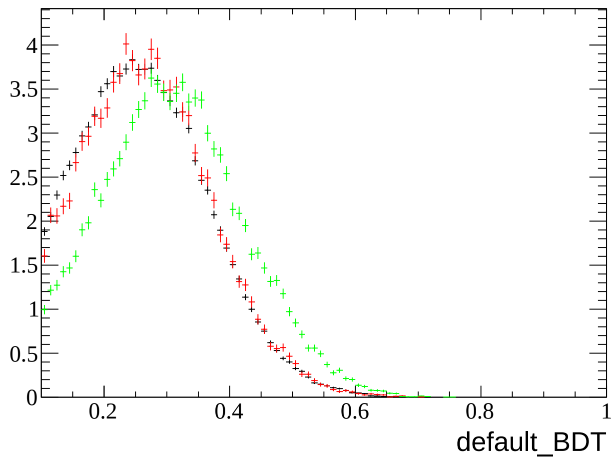
- impact parameter of the B^0 and the four final state particles;
- kaon and pion identification ($\mathrm{DLL}_{K\pi}$) of the K^+ and π^-;
- muon $\mathrm{DLL}_{\mu\pi}$ of the μ^+ and μ^-.

There are differences in the impact parameter resolution between data and the simulation, which have been observed by several analysis. In order to account for these differences, the track states of each of the simulated tracks used to reconstruct the offline selected candidates are smeared using the `Phys/TrackSmearing` tool.

The pion and kaon identification performance of the LHCb detector is studied using the RICH `PIDCalib` tools in data using samples of genuine pions and kaons selected from the decays $D^{*+} \to D^0\pi^+$ where $D^0 \to K^-\pi^+$. In order to properly account for the differences in PID performance, the DLL of pions and kaons in the MC are replaced by sampling from the various DLL distributions of genuine kaon or pions in the data. For each kaon and pion a new value of $\mathrm{DLL}_{K-\pi}$ is assigned according to the momentum and pseudo-rapidity of the particle. This new DLL value is then used in the BDT. For the DLL variables for muons, an analogous procedure is used, but using a tag-and-probe approach with $B^+ \to J/\psi K^+$, where $J/\psi \to \mu^+\mu^-$ in data. The $B^+ \to J/\psi K^+$ sample is obtained from the stripping line `MuIDCalib_JpsiKFromBNoPIDNoMIP`, which does not apply any cut on a probe track.

In addition the MC is re-weighted to account for differences in the relative tracking efficiency between data and MC and for differences in the efficiency of the `IsMuon` requirement (which is applied in the Stripping). Finally the MC samples have been re-weighted to account for differences in the occupancy between data and MC (using the size of the `Rec/Track/Best` container).

The BDT response after the application of the trigger, stripping and offline selection, for $B^0 \to K^{*0} J/\psi$ candidates is shown in Fig. 14. This demonstrates that there is in general an excellent agreement between the MC and data (for the control channel) after the MC tuning procedure, whereas the agreement before the MC tuning is poor (see also Appendix. A.1).



(a) BDT output distribution

Figure 14: BDT response for offline selected candidates $B^0 \to J/\psi\ K^{*0}$ in the data and the MC. The three distributions are Data (Black), data-corrected simulated events (Red) and uncorrected simulated events (Green)

Other data/MC comparisons can be found in the appendix of this note (see Sec. A).

11 Acceptance correction

The reconstruction, trigger and selection each bias the angular and q^2 distributions that are to be measured. For example, for muon candidates to be reconstructed, they must have at least the 3 GeV/c momentum required to traverse the iron muon filter and to leave hits in all the muon stations. This has the effect of warping the $\cos\theta_l$ distribution, removing candidates with $\cos\theta_l$ close to one. Similarly, in $\cos\theta_K$, the impact parameter (IP) requirements made in the trigger algorithms remove events with extreme values of $\cos\theta_K$, as very forward-going hadrons tend to have lower IP. A second effect in $\cos\theta_K$ originates from the low boost of backward-going hadrons at extreme $\cos\theta_K$, given the minimum momentum required to traverse the dipole magnet and tracking stations. The acceptance effect in $\cos\theta_K$ is asymmetric as the kaon tends to be more energetic than the pion after the boosts.

In order to correctly determine the physics parameters that describe the angular distribution, these 'acceptance effects' must be accounted for. In the present analysis this is done by weighting the events that are selected by the inverse of their efficiency in the maximum-likelihood fit to the angular (or q^2-) distribution. The use of event-by-event weights to correct for the acceptance, rather than describing the acceptance in the fit, is driven by the variation of the angular efficiency with q^2. This variation in q^2 can be significant compared to the size of the q^2-bins used in the analysis. Consequently it is not possible to include a single PDF that describes the shape of the acceptance in $\cos\theta_l$, $\cos\theta_K$ and ϕ in a fit to the angular distribution of the daughters.

A factorised approach has been adopted for the angular efficiency. The factorised approach treats the angular efficiency as a function of $\cos\theta_l$, $\cos\theta_K$ and ϕ independently. The efficiency in q^2 does not factorise and is instead binned in $0.5\,\mathrm{GeV}^2/c^4$ q^2-bins, for the region above $6.0\,\mathrm{GeV}^2/c^4$. At low q^2, where the acceptance varies more rapidly, $0.1\,\mathrm{GeV}^2/c^4$ q^2-bins are taken for the region below $1.0\,\mathrm{GeV}^2/c^4$, and $0.2\,\mathrm{GeV}^2/c^4$ q^2-bins elsewhere. This bin size is more than four times narrower than the smallest of the q^2-bins used in the analysis. In each of these small q^2-bins a different angular efficiency is used to calculate the event weights.

After applying the trigger and the full offline selection, approximately two million events remain in the large $B^0 \to K^{*0}\mu^+\mu^-$ phase-space sample for estimating the acceptance correction. These events were generated flat in $\cos\theta_l$, $\cos\theta_K$ and ϕ and have a falling distribution in q^2.

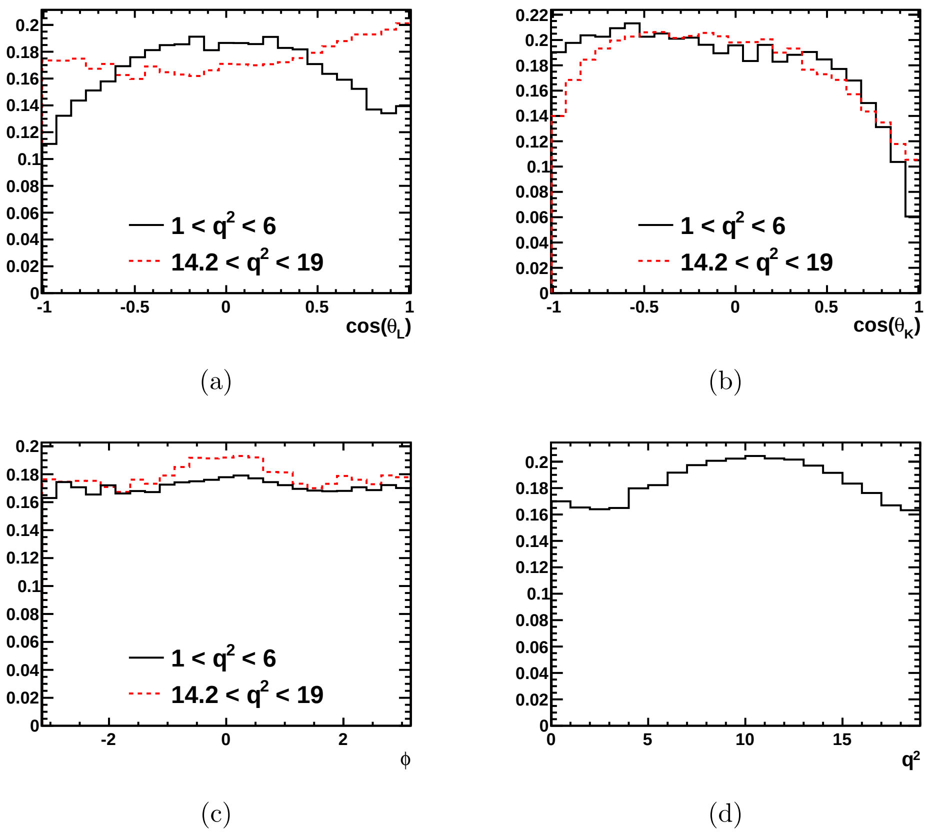


Figure 15: The reconstruction, trigger and offline selection pseudo-efficiencies as a function of the kinematic variables in $B^0 \to K^{*0}\mu^+\mu^-$ SM MC. The variation of the angular efficiencies at low- and high-q^2 is included for reference.

11.1 Exploiting symmetries in the acceptance correction

To maximise the available MC statistics, the efficiency distribution is folded in $\cos\theta_l$ and in ϕ. The $\cos\theta_l$ distribution is assumed to be symmetric about $\cos\theta_l = 0$. For this assumption not to be true there would need to be both a large difference in the efficiency for μ^+ and μ^- (that doesn't cancel when the dipole field is flipped) and a large $\mathcal{CP}$ asymmetry between B^0 and $\overline{B}^0$.

The efficiency in the ϕ angle is assumed to be symmetric with respect to the translation of $\phi \rightarrow \phi + \pi$. The combination of folding the efficiency in ϕ and in $\cos\theta_l$ increases the effective MC statistics by a factor of four.

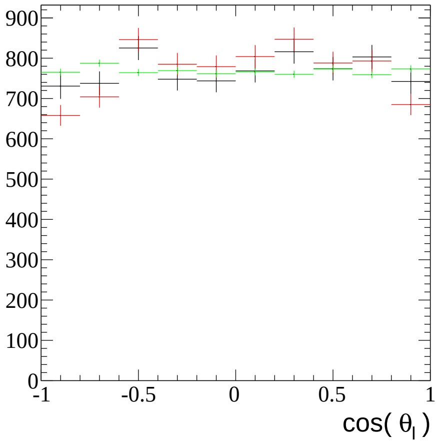
11.2 Testing the acceptance correction

The acceptance correction is verified on MC and later cross-checked using $B^0 \rightarrow K^{*0} J/\psi$ data (Sec. 13). Offline selected phase space MC events are used to verify the performance on MC. The generator level distributions of the phase-space events are flat in $\cos\theta_l$, $\cos\theta_K$ and ϕ and hence provide a good test of the re-weighting. For a given bin in the angular variables, "b", the number of events after the acceptance correction is:

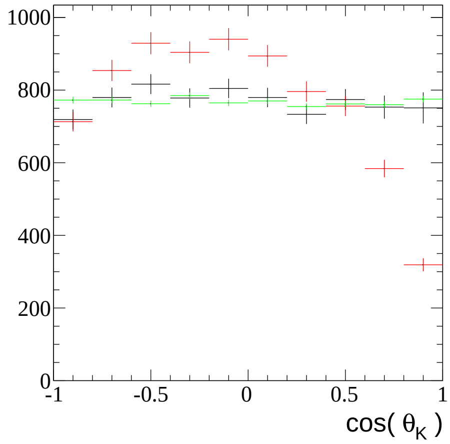
$$N_b = \sum_{i=0}^{N} \frac{1}{\varepsilon_i(\cos\theta_l, \cos\theta_K, \phi, q^2)}$$

If the acceptance correction correctly reproduces the effects of the trigger, reconstruction, stripping and offline selection then, the distribution of N_b across the angular variables should be the same as the generator level distribution.

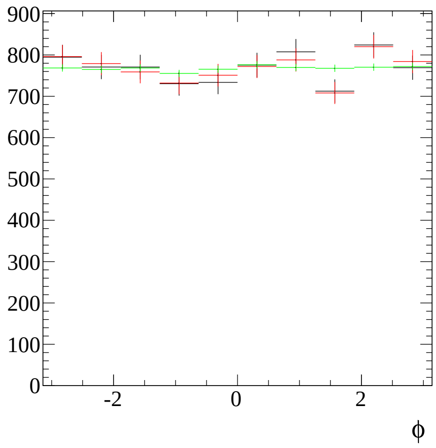
The performance of the factorised acceptance correction on an independent sample of phase space M is shown in Fig. 16. The generator level distributions for $\cos\theta_l$, $\cos\theta_K$, ϕ and q^2 are compared to the distributions after the offline selection, reconstruction, trigger and stripping and to the distribution of candidates weighting for the expected acceptance effect. After the acceptance correction the candidates are flat in $\cos\theta_l$, $\cos\theta_K$ and ϕ and accurately reproduce the generator level distributions.



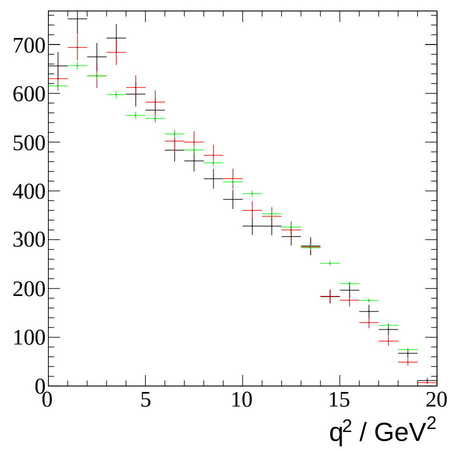
(a) $\cos\theta_l$



(b) $\cos\theta_K$



(c) ϕ



(d) q^2

Figure 16: The effect of the factorised acceptance correction as a function of the angular variables, $\cos\theta_l$, $\cos\theta_K$, ϕ and of q^2. Figs (a,b,c,d) show the original distribution before correction (red), the corrected distribution (black) and the expected distribution (green). The corrected distributions match the expected distributions, with increased corrections both towards extreme $\cos\theta_l$ values and the low q^2 region.

11.3 Systematic uncertainty associated with the acceptance correction

No evidence is seen indicating that the angular efficiency in each of the 0.5 GeV^2/c^4 q^2-bins can not be factorised into three one-dimensional angular efficiencies. It is however very difficult to quantify the level to which these assumptions hold, beyond stating that it appears to hold at the level of $\sim 5-10\%$ (see Appendix B).

Practically, a conservative estimate for the systematic uncertainty on the acceptance correction is estimated by systematically varying the acceptance correction in $\cos\theta_l$, $\cos\theta_K$ and ϕ by 5%, in a way that would introduce the maximum bias in the physics parameters: e.g. by fluctuating the efficiency of events with $\cos\theta_l \sim \pm 1$ up or down by 5% to introduce a bias in A_{FB} or events with $\cos\theta_K \sim 0$ up or down by 5% to bias F_{L}.

12 Validation of the angular analysis with toy-MC

This section details the results of a toy-MC studies with the expected signal and background yield in $1\,\mathrm{fb}^{-1}$ for $0.1 < q^2 < 2.0\,\mathrm{GeV}^2/c^4$. This q^2 range has been chosen for illustrative purposes and similar results are achieved in the other q^2 bins (with the caveats outlined below). Toy datasets were generated with A_{FB}, F_{L}, S_3 and S_9 values as measured in Ref. [8] ($A_{FB} = -0.02$, $F_L = 0.36$, $A_T^2 = -0.16$ and $S_9 = 0.06$). Five hundred datasets were generated.

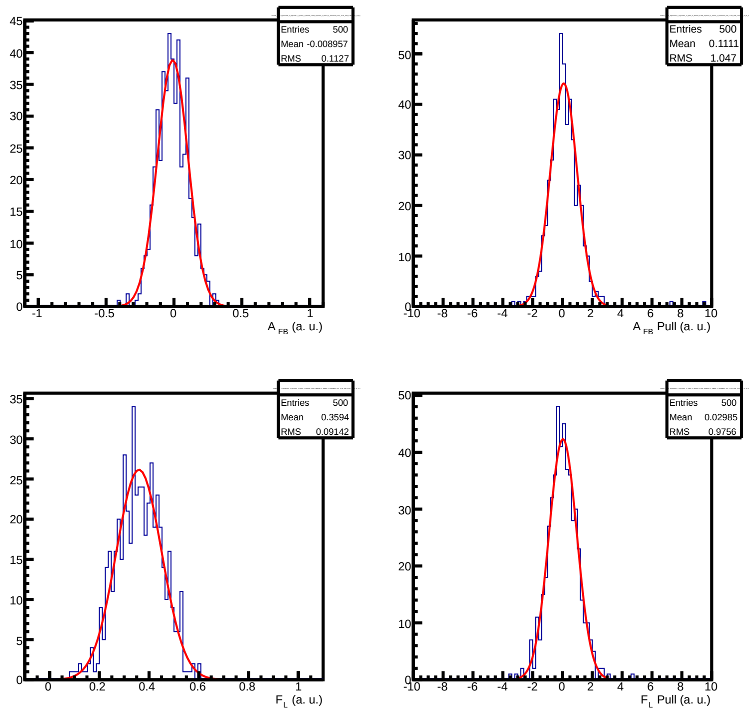
An additional 500 datasets were generated including an S-wave component with parameter values $A_S = -0.2$ and $F_S = 0.08$, which correspond to the values seen in $B^0 \to K^{*0} J/\psi$. In each case, the fit pdf did not contain an S-wave component, effectively constraining $A_S = 0$ and $F_S = 0$. This tests the impact of the S-wave component on the fit.

Signal candidates have been accept-rejected according to the acceptance correction described in Sec. 11 and re-weighted in the subsequent fit. The effect of the weighted data on the error matrix was corrected using a 'sum of weights' correction provided by `RooFit`. Background events were generated flat in the angles but were modelled with a second order polynomial in the fit.

Pulls have been calculated from the difference between the generated value of A_{FB}, F_{L}, S_3 and S_9, and the value returned by the likelihood fit, divided by the parabolic error from the covariance matrix of the likelihood fit.

12.1 MC validation for the observables A_{FB}, F_{L}, S_3 and S_9.

The distribution of fit results for each of the observables A_{FB}, F_{L}, S_3 and S_9 are shown in Fig. 17 (A_{FB} and F_{L}), Fig. 18 (S_3 and S_9). The experimental uncertainty, pull centre and pull width for each observable are summarised in Table. 13.

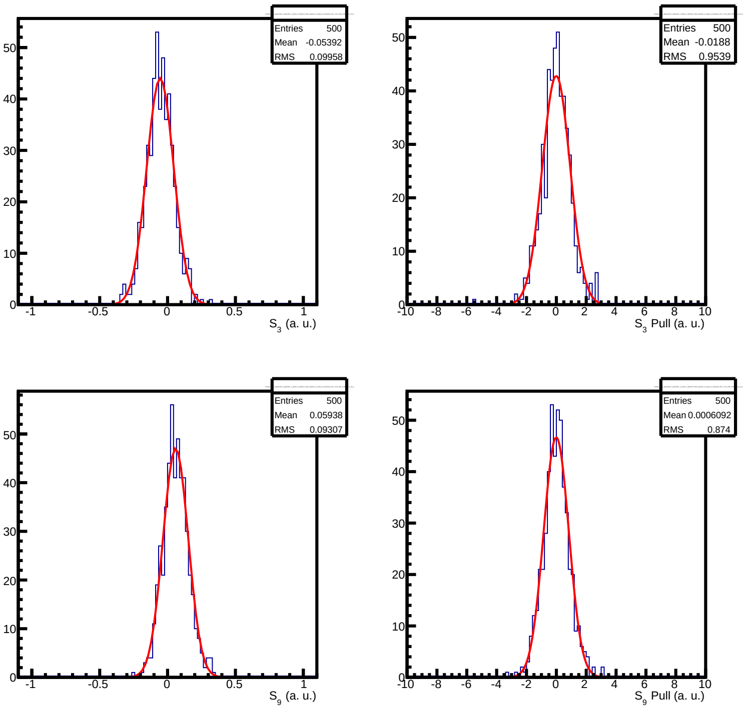


(a)

Figure 17: Distribution of fitted values (left), and pull distribution (right), for the observables A_{FB} (top) and F_{L} (bottom) for 500 toy MC datasets when fitting for A_{FB} and S_9.

Observable	Experimental Uncertainty	Pull Centre	Pull Width
$A_{\rm FB}$	0.113 ± 0.005	0.083 ± 0.041	0.899 ± 0.029
$F_{\rm L}$	0.091 ± 0.004	0.029 ± 0.042	0.935 ± 0.031
S_3	0.100 ± 0.004	-0.010 ± 0.042	0.930 ± 0.031
S_9	0.093 ± 0.004	-0.007 ± 0.038	0.845 ± 0.027

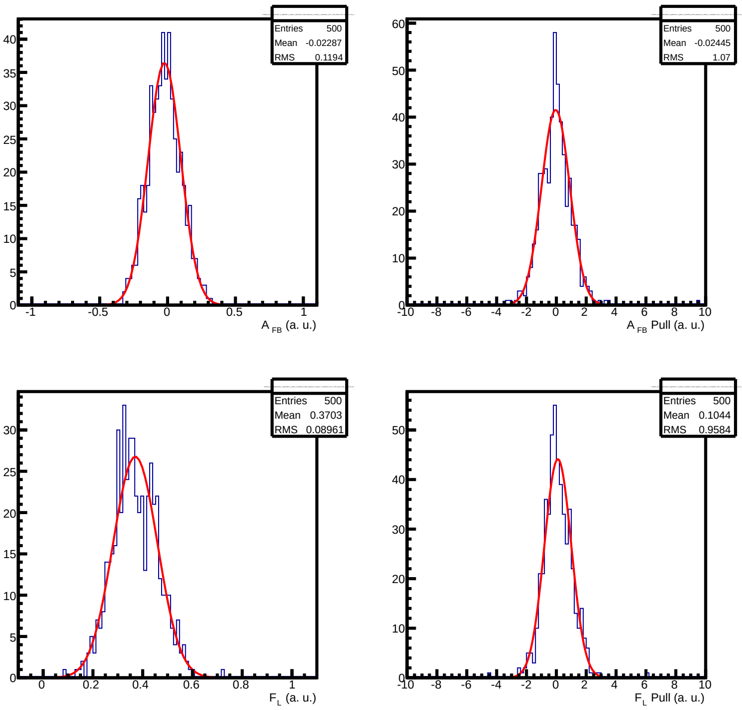
Table 13: Results of fits to 500 toy experiments for the observables $A_{\rm FB}$, $F_{\rm L}$, S_3 and S_9.



(a)

Figure 18: Distribution of fitted values (left), and pull distribution (right), for the observables S_3 (top) and S_9 (bottom) for 500 toy MC datasets when fitting for $A_{\rm FB}$ and S_9.

The distribution of fit results, when generating with $B^0 \to K^{*0} J/\psi$-like swave, for each of the observables are shown in Fig. 19 ($A_{\rm FB}$ and $F_{\rm L}$), Fig. 20 (S_3 and S_9). The experimental uncertainty, pull centre and pull width for each observable are summarised in Table. 14.

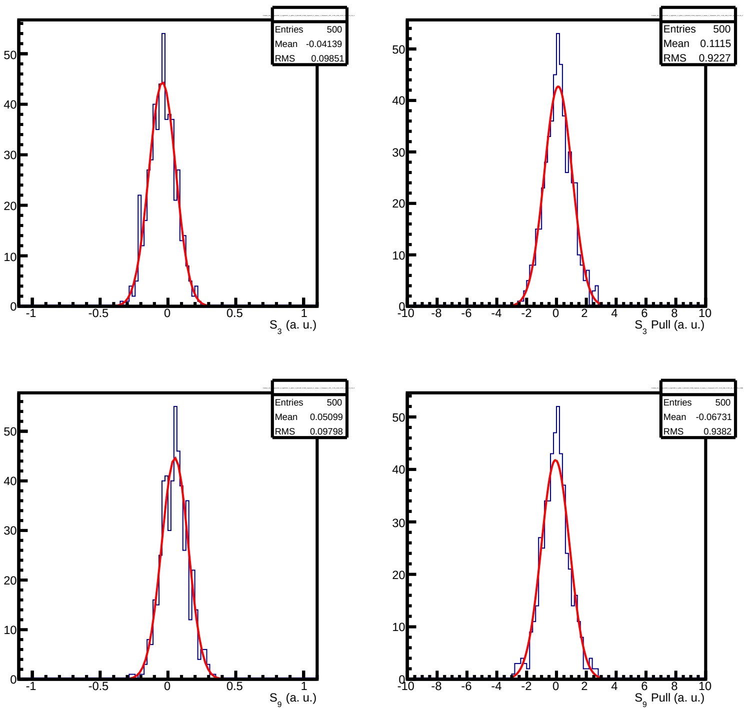


(a)

Figure 19: Distribution of fitted values (left), and pull distribution (right), for the observables A_{FB} (top) and F_{L} (bottom) for 500 toy MC datasets when fitting for A_{FB} and S_9 in the presence of a $B^0 \to K^{*0} J/\psi$-like s-wave.

Observable	Experimental Uncertainty	Pull Centre	Pull Width
A_{FB}	0.120 $\pm$ 0.005	-0.044 $\pm$ 0.043	0.951 $\pm$ 0.032
F_{L}	0.089 $\pm$ 0.004	0.100 $\pm$ 0.041	0.900 $\pm$ 0.030
S_3	0.099 $\pm$ 0.004	0.117 $\pm$ 0.042	0.934 $\pm$ 0.031
S_9	0.098 $\pm$ 0.004	-0.065 $\pm$ 0.043	0.953 $\pm$ 0.032

Table 14: Results of fits to 500 toy experiments including the s-wave component for the observables A_{FB}, F_{L}, S_3 and S_9.



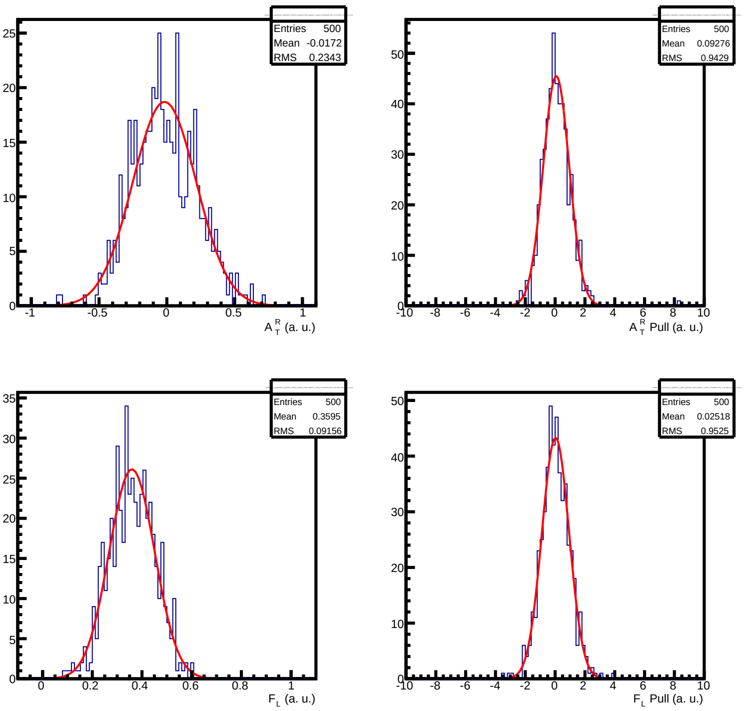
(a)

Figure 20: Distribution of fitted values (left), and pull distribution (right), for the observables S_3 (top) and S_9 (bottom) for 500 toy MC datasets when fitting for A_{FB} and S_9 in the presence of a $B^0 \to K^{*0} J/\psi$-like s-wave.

12.2 MC validation for the transverse observables (A_{T}^{Re}, F_{L}, A_{T}^{2} and A_{T}^{Im})

The study outlined above was repeated, however the fitting scheme was changed to fit for the observables A_{T}^{Re}, F_{L}, A_{T}^{2} and A_{T}^{Im}.

The distribution of fit results for each of the observables are shown in Fig. 21 (A_{T}^{Re} and F_{L}), Fig. 22 (A_{T}^{2} and A_{T}^{Im}). The experimental uncertainty, pull centre and pull width for each observable are summarised in Table. 15.

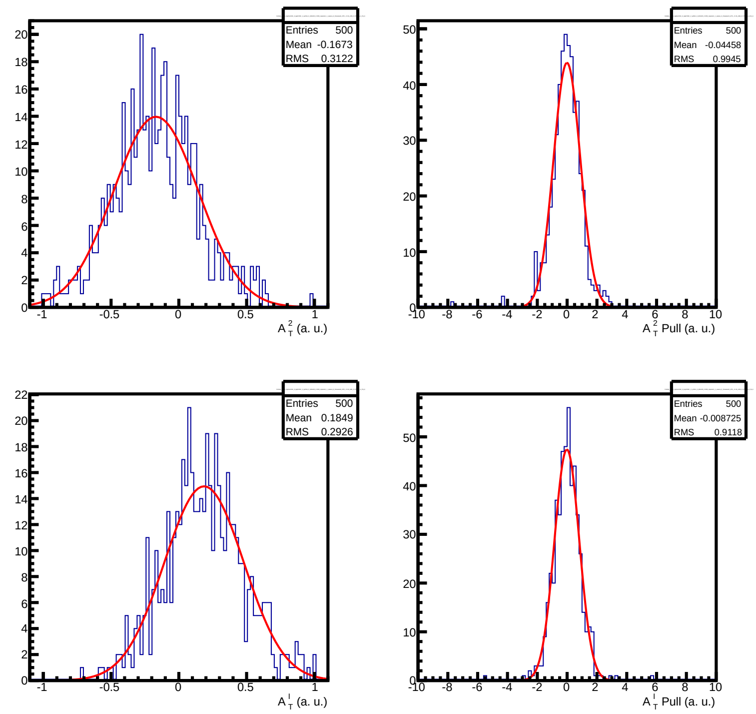


(a)

Figure 21: Distribution of fitted values (left), and pull distribution (right), for the observables A_{T}^{Re} (top) and F_{L} (bottom) for 500 toy MC datasets when fitting for A_{T}^{Re} and A_{T}^{Im}.

Observable	Experimental Uncertainty	Pull Centre	Pull Width
A_{T}^{Re}	0.235 ± 0.010	0.074 ± 0.039	0.876 ± 0.028
F_{L}	0.092 ± 0.004	0.028 ± 0.041	0.913 ± 0.030
A_{T}^{2}	0.315 ± 0.014	-0.010 ± 0.041	0.900 ± 0.030
A_{T}^{Im}	0.294 ± 0.013	-0.015 ± 0.038	0.833 ± 0.027

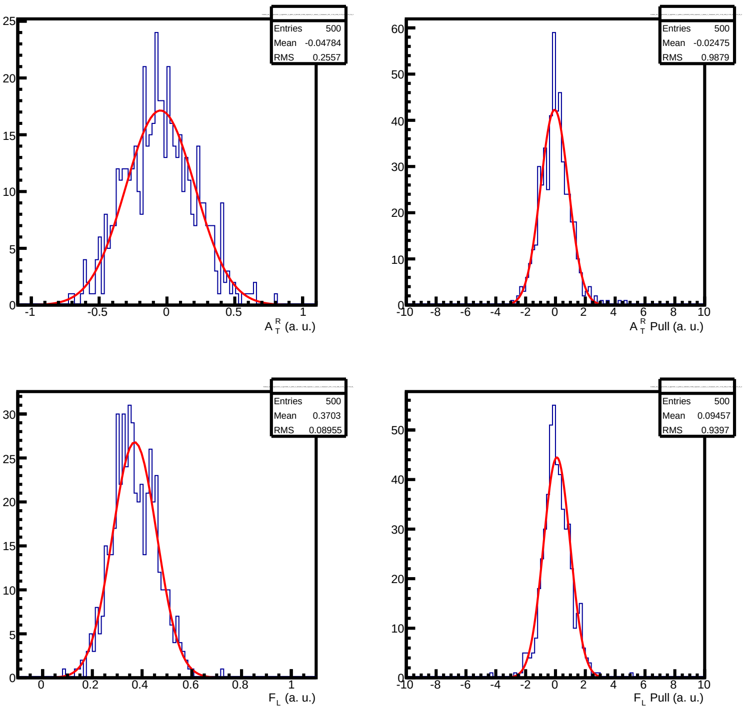
Table 15: Results of fits to 500 toy experiments for the observables A_{T}^{Re}, F_{L}, A_{T}^{2} and A_{T}^{Im}.



(a)

Figure 22: Distribution of fitted values (left), and pull distribution (right), for the observables A_{T}^{2} (top) and A_{T}^{Im} (bottom) for 500 toy MC datasets when fitting for A_{T}^{Re} and A_{T}^{Im}.

The distribution of fit results, when generating with $B^0 \to K^{*0} J/\psi$-like swave, for each of the observables are shown in Fig. 23 (A_{T}^{Re} and F_{L}), Fig. 24 (A_{T}^{2} and A_{T}^{Im}). The experimental uncertainty, pull centre and pull width for each observable are summarised in Table. 16.

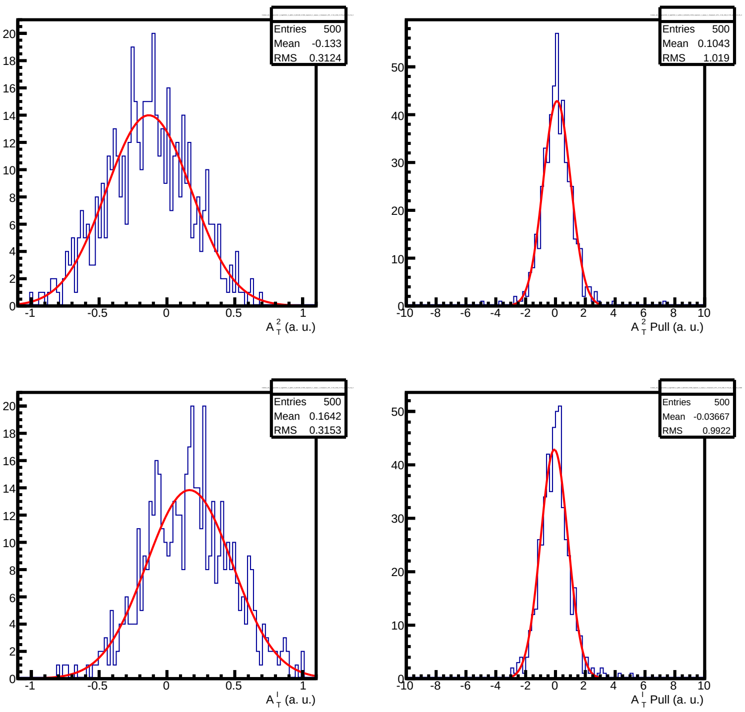


(a)

Figure 23: Distribution of fitted values (left), and pull distribution (right), for the observables A_{T}^{Re} (top) and F_{L} (bottom) for 500 toy MC datasets when fitting for A_{T}^{Re} and A_{T}^{Im} in the presence of a $B^0 \to K^{*0} J/\psi$-like s-wave.

Observable	Experimental Uncertainty	Pull Centre	Pull Width
A_{T}^{Re}	0.256 ± 0.011	-0.059 ± 0.042	0.938 ± 0.031
F_{L}	0.089 ± 0.004	0.092 ± 0.040	0.895 ± 0.029
A_{T}^{2}	0.314 ± 0.014	0.100 ± 0.042	0.925 ± 0.031
A_{T}^{Im}	0.318 ± 0.014	-0.077 ± 0.042	0.923 ± 0.031

Table 16: Results of fits to 500 toy experiments including the s-wave component for the observables A_{T}^{Re}, F_{L}, A_{T}^{2} and A_{T}^{Im}.



(a)

Figure 24: Distribution of fitted values (left), and pull distribution (right), for the observables A_{T}^{2} (top) and A_{T}^{Im} (bottom) for 500 toy MC datasets when fitting for A_{T}^{Re} and A_{T}^{Im} in the presence of a $B^0 \to K^{*0} J/\psi$-like s-wave.

13 Validation of the angular analysis with $B^0 \to K^{*0} J/\psi$

The full fitting strategy for $B^0 \to K^{*0}\mu^+\mu^-$ has been validated using $B^0 \to K^{*0} J/\psi$ candidates. The angular distribution of these candidates can be well described by the same angular distributions (in one, two or three dimensions) that were discussed for $B^0 \to K^{*0}\mu^+\mu^-$. The only differences arise from having $A_{\mathrm{FB}} = 0$ and a single set of amplitudes (with no differentiation between left- and right- handedness). These differences have no impact on the form of the angular distribution.

A fit to the full statistics of the $B^0 \to K^{*0} J/\psi$ sample is described in Sec. 13.2. A more appropriate comparison to $B^0 \to K^{*0}\mu^+\mu^-$ is made by then splitting the large $B^0 \to K^{*0} J/\psi$ sample in the data into small 100 event sub-samples, which loosely corresponds to the expected statistics in the least occupied q^2 bin.

13.1 Comparison with results from full angular analysis at LHCb and BaBar

The $B^0 \to K^{*0} J/\psi$ transversity amplitudes from a full angular analysis at LHCb and BaBar can be found in Tables 17 and 18 respectively. Ignoring the S-wave contribution this gives values of: F_{L} of 0.57 and 0.56 respectively; A_T^2 of -0.14 and 0.05 respectively and S_9 of -0.07 and -0.08 respectively.

	Including S-wave	No S-wave
$\lvert A_{\parallel}\rvert^2$	0.252 ± 0.020	0.253 ± 0.020
$\lvert A_{\perp}\rvert^2$	0.178 ± 0.022	0.191 ± 0.019
$\delta_{\parallel} - \delta_0$	-2.87 ± 0.11	-2.82 ± 0.12
$\delta_{\perp} - \delta_0$	3.02 ± 0.10	3.07 ± 0.09

Table 17: $B^0 \to K^{*0} J/\psi$ transversity amplitudes from a full angular analysis with $36\,\mathrm{pb}^{-1}$ of integrated luminosity at LHCb (from Ref. [20]).

	No S-wave
$\vert A_\parallel\vert^2$	$0.211 \pm 0.010 \pm 0.006$
$\vert A_\perp\vert^2$	$0.233 \pm 0.010 \pm 0.005$
$\delta_\parallel - \delta_0$	$-2.93 \pm 0.08 \pm 0.04$
$\delta_\perp - \delta_0$	$2.91 \pm 0.05 \pm 0.03$

Table 18: $B^0 \to K^{*0} J/\psi$ transversity amplitudes from a full angular analysis performed by BaBar (from Ref. [21]).

13.2 Fitting the full $B^0 \to J/\psi\, K^{*0}$ sample

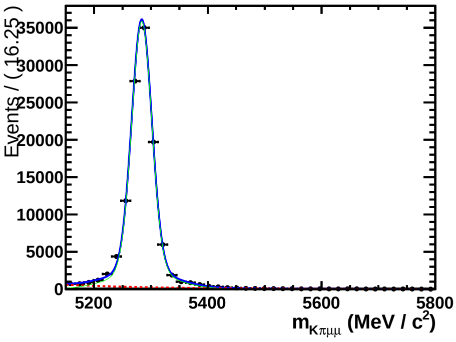
The full sample of $B^0 \to J/\psi\ K^{*0}$ events were fitted, with and without an S-wave component, to extract the observables A_T^R, F_L, A_T^2 and A_T^I (and A_S and F_S). A comparison with the results from the BaBar collaborations full angular analysis of $B^0 \to J/\psi\ K^{*0}$ provides a powerful validation of the fitting procedure. The fit results are summarised in Table. 19. The values obtained in the present study are in good agreement with those from BaBar, with $A_{\rm FB} \sim 0$. Note, the errors are not comparable on $A_{\rm T}^2$ because of the use of a partial angular analysis compared to the full angular analysis by BaBar.

Observable	Present result (w/ S-wave)	Present result (w/o S-wave)	BaBar value (w/o S-wave)
A_T^{Re}	0.009 ± 0.007	0.009 ± 0.007	N/A
F_L	0.561 ± 0.002	0.552 ± 0.002	0.56 ± 0.03
A_T^2	0.042 ± 0.015	0.029 ± 0.013	0.05 ± 0.03
A_T^{Im}	-0.362 ± 0.016	-0.313 ± 0.014	-0.34 ± 0.05
A_S	-0.174 ± 0.003	N/A	N/A
F_S	0.078 ± 0.006	N/A	N/A

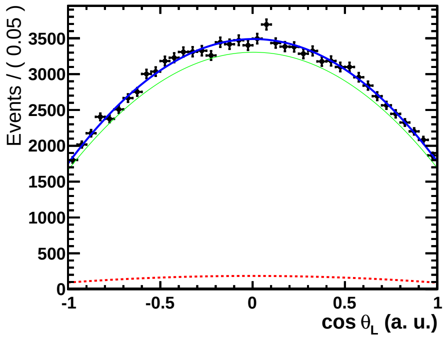
Table 19: Comparison of $B^0 \to J/\psi\ K^{*0}$ fit results from the present study, with and without the S-wave component, with the BaBar result from Ref. [21].

Note, there is no first principle reason to expect $B^0 \to K^{*0} J/\psi$ to have $A_{\rm T}^2 = 0$. It is non-zero in QCD factorisation [22].

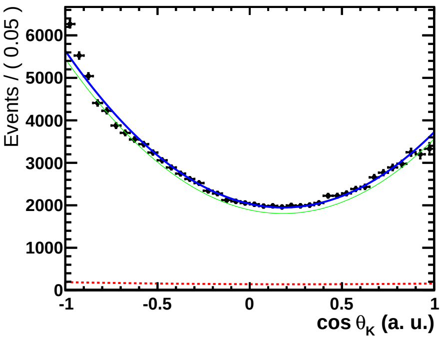
The one-dimensional projections of the $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distributions with the fitted PDF are shown in Fig. 25. The sinusoidal variation of ϕ results from a non-zero value of S_9 (and A_T^{Im}). No asymmetry is seen in $\cos\theta_l$, but a significant asymmetry is visible in $\cos\theta_K$. This asymmetry results from interference of the $K^{*0}(892)$ with a broad $K^+\pi^-$ S-wave.



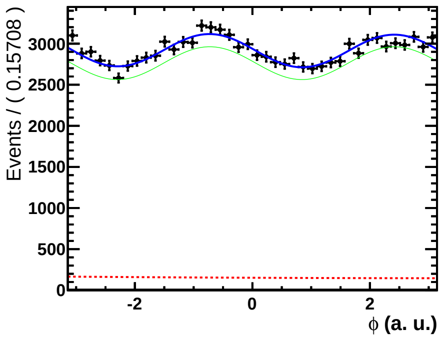
(a) $m_{K\pi\mu\mu}$



(b) $\cos\theta_l$



(c) $\cos\theta_K$



(d) ϕ

Figure 25: 1D projections of the four fitted quantities for the full $B^0 \to J/\psi$ K^{*0} dataset; (a) mass, (b) $\cos(\theta_L)$, (c) $\cos(\theta_K)$ and (d) ϕ. The fitted pdf (blue), the signal-only pdf (green) and background-only pdf (red dash) are overlaid.

The disagreement at $\cos\theta_K \sim -1$ in Fig. 25 is not understood. The disagreement in the shape is at the level of $\pm 5\%$ and is covered as a systematic uncertainty. No such disagreement is seen in $\cos\theta_l$ and ϕ.

13.3 Validation using 100 event sub-samples

A further check of the fitting procedure was performed by splitting the full $B^0 \to J/\psi\, K^{*0}$ dataset into sub-samples. For this study, 1159 sub-samples of 100 events were used, corresponding roughly to the expected statistics in the least occupied q^2 bin ($2 < q^2 < 4.3\,\mathrm{GeV}^2/c^4$). By fitting each sub-sample individually, the experimental precision and pull distributions in each observable could be analysed in the data. Due to the low level of background in each sub-sample (we expect around 5 background events in the upper

Observable	Experimental Uncertainty	Pull Centre	Pull Width
A_T^{Re}	0.249 ± 0.006	0.017 ± 0.034	0.978 ± 0.024
F_L	0.097 ± 0.002	-0.206 ± 0.041	1.160 ± 0.029
A_T^2	0.495 ± 0.017	-0.015 ± 0.032	0.903 ± 0.022
A_T^{Im}	0.480 ± 0.017	0.207 ± 0.028	0.811 ± 0.020

Table 20: Results of 1159 fits to 100 event sub-samples of the $B^0 \to J/\psi\, K^{*0}$ dataset neglecting the S-wave component.

B^0 mass sideband) the polynomial used to model the angular shape of the background events was reduced from second to first order. The pull value for each sub-sample was calculated using the central value obtained from an equivalent fit to the full $B^0 \to J/\psi\ K^{*0}$ dataset. Fits with results at a physical boundary are removed, as their errors can not be trusted.

The results of this study, when the S-wave terms are neglected is summarised in Table. 20. The pull distribution of A_{T}^2 and A_T^{Im} are biased. This bias occurs because the experimental uncertainty on the observables is large compared to the parameter range.

14 Summary of validation studies

The validation studies with toy-MC and $B^0 \to K^{*0} J/\psi$ highlight some of the difficulties of this analysis:

- The impact of the boundaries described in Sec. 9.5 is clearly evident.In the toy studies the boundaries show up as a non-Gaussian distribution for the results of the toys - which in turn results in pull distributions that have a width larger or smaller than one.

- In some cases the allowed range of the parameters is small compared to the uncertainty on the fits (e.g. A_T^2 for large F_{L}).

This may make it look like the fit performance on toy-MC is poor. It is clear that it is not always suitable to trust the covariance matrix returned by `MINUIT` as an estimate of the errors. This is particularly true for any parameter that is close to a boundary.

15 Angular analysis fit results

This section details the result of the angular fits in the six-plus-one q^2-bins. Results of fits for both sets of observables, $\{A_{\mathrm{FB}}$, F_{L}, S_3, S_9 and $A_9\}$ and $\{A_{\mathrm{T}}^{Re}$, F_{L}, A_{T}^{Im} and A_{T}^{2} $\}$, are detailed.

The central values for the two sets of observables are shown in Table. 21 and Table. 22 respectively.

$q^2(\mathrm{GeV}^2/c^4)$	A_{FB}	F_{L}	S_3	S_9	A_9
$0.10 < q^2 < 2.00$	-0.02	0.37	-0.04	0.05	0.12
$2.00 < q^2 < 4.30$	-0.20	0.74	-0.04	-0.03	0.06
$4.30 < q^2 < 8.68$	0.16	0.57	0.08	0.01	-0.13
$10.09 < q^2 < 12.86$	0.28	0.48	-0.16	-0.01	-0.00
$14.18 < q^2 < 16.00$	0.51	0.33	0.03	0.00	-0.06
$16.00 < q^2 < 19.00$	0.30	0.37	-0.22	0.06	-0.00
$1.00 < q^2 < 6.00$	-0.17	0.65	0.03	0.07	0.03

Table 21: Angular analysis central values for the observables A_{FB}, F_{L}, S_3, S_9 and A_9.

$q^2(\mathrm{GeV}^2/c^4)$	A_{T}^{Re}	F_{L}	A_{T}^{2}	A_{T}^{Im}
$0.10 < q^2 < 2.00$	-0.05	0.37	-0.14	0.16
$2.00 < q^2 < 4.30$	-1.00	0.74	-0.29	-0.23
$4.30 < q^2 < 8.68$	0.50	0.57	0.36	0.05
$10.09 < q^2 < 12.86$	0.71	0.48	-0.60	-0.06
$14.18 < q^2 < 16.00$	1.00	0.33	0.07	0.02
$16.00 < q^2 < 19.00$	0.64	0.37	-0.71	0.18
$1.00 < q^2 < 6.00$	-0.66	0.65	0.17	0.41

Table 22: Angular analysis central values for the observables A_{T}^{Re}, F_{L}, A_{T}^{Im} and A_{T}^{2}.

15.1 Error estimation

The estimation of parameter errors is complicated by the presence of mathematical boundaries in the fit. This is described in Sec. 9. To negate the boundary effects two different methods are pursued when estimating the statistical uncertainties on the angular observables: Feldman-Cousins and MINOS-like $\Delta LL = \pm\frac{1}{2}$ from the profile-likelihood (in the allowed parameter range).

15.1.1 Feldman-Cousins estimate of the confidence interval

The Feldman-Cousins technique for determining confidence intervals is described in Ref. [23]. The application of Feldman-Cousins to estimate the 68% confidence interval is described below, using F_{L} as an example. The same process is applied for all four observables in the six-plus-one q^2 bins.

First a fit is performed to estimate the best-fit values for all of the parameters, including F_{L} and the nuisance parameters, λ. The nuisance parameters include the other angular observables, A_{FB}, A_{Im} and S_3. This set of fit-parameters will be denoted $\hat{F_{\mathrm{L}}}$ and $\hat{\lambda}$. Next a scan is performed over the full range of F_{L} ($0 < F_{\mathrm{L}} < 1$). For each value of F_{L}, the likelihood ratio:

$$R^i = \frac{L(\vec{x}|F_{\mathrm{L}}^i, \hat{\hat{\lambda}}^i)}{L(\vec{x}|\hat{F}_{\mathrm{L}}, \hat{\lambda})}$$

is calculated, where $\hat{\hat{\lambda}}^i$ is used to represent the best-fit value for the nuisance parameters with F_{L} fixed to be F_{L}^i.

At every point in the parameter space 500 toys are generated from F_{L}^i and $\hat{\hat{\lambda}}^i$, and the likelihood ratio is calculated for each toy. A confidence interval is then determined from the fraction of toys that have $R^i_{\mathrm{toy}} > R^i_{\mathrm{data}}$.

Toy-data sets are accept-rejected and then re-weighted to account for the angular acceptance. Without simulating the q^2-dependence it is not possible to fully reproduce the acceptance effect seen in data. Instead, the acceptance distribution is assumed to be that of the average q^2-value in the q^2-bin. The toy-data sets are generated with the maximum likelihood estimate values obtained from the fit to the data with the parameter of interest fixed. When fitting a penalty term has been included in the log-likelihood to penalise combinations of parameters that are outside the mathematically allowed region of parameter space.

15.1.2 Potential problems with FC near boundaries

Problems have been seen with the Feldman-Cousins intervals if parameters are near a mathematical boundary. This is true in several regions of parameter space, most notably in the $2 < q^2 < 4.3\,\mathrm{GeV}^2/c^4$ q^2-bin. Whilst FC deals well with having the parameter of interest near a boundary, the fits to the toy-MC can have significant problems if one of other parameters is near the boundary. In cases like this, the minimisation of `MINUIT` has trouble converging to the correct minimum.

If the `MINUIT` convergence fails, or the minima exists outside of a valid region of phase-space (i.e. where either the signal or background angular pdfs go negative), an alternative sequential minimisation is performed.

15.1.3 Falling back on sequential minimisation

The sequential minimisation is simply a sequence of `MINUIT` fits where the initial parameters of each fit in the sequence are set to the final values of the previous fit. The initial parameters for the first fit in the sequence are set to sensible values. At the start of each of the fits in the sequence, the partial derivatives of the likelihood are computed to estimate sensible step sizes for each of the floating parameters. The sequence is ended once the change in likelihood value between two fits is less than 10^{-6}, or the sequence is 20 fits long.

In some cases it is possible, due to boundary effects and/or parameter correlations, that the sequential fit will fail to converge or converge to a local minima. To protect against this, the sequential minimisation is performed multiple times with a Gaussian fluctuation of the initial signal parameters (the parameter values are constrained to the valid region of the phase-space). The sequential minimisation that yields the best likelihood value is chosen as the best fit result for the signal parameters.

15.2 Candidate distributions

The distribution of events in mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ in the six q^2-bins is given in Figs. 26-32. The distribution of events in the signal mass window and upper mass sideband is shown in Figs. 33-39.

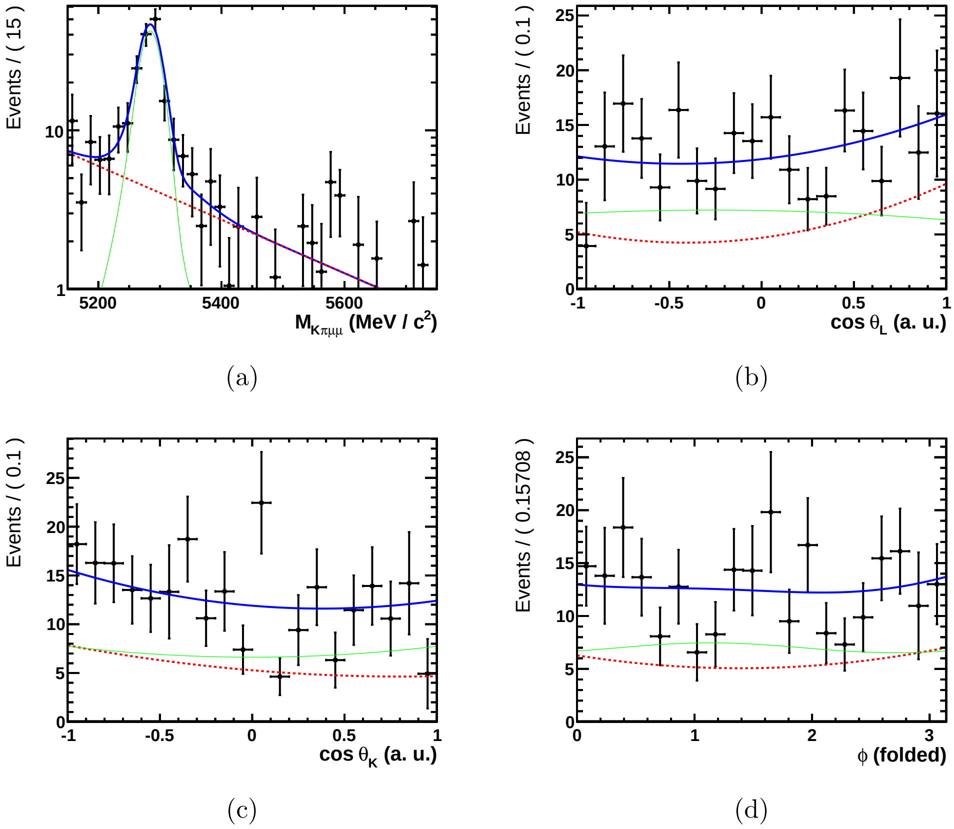


Figure 26: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$ in the full mass range. The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.

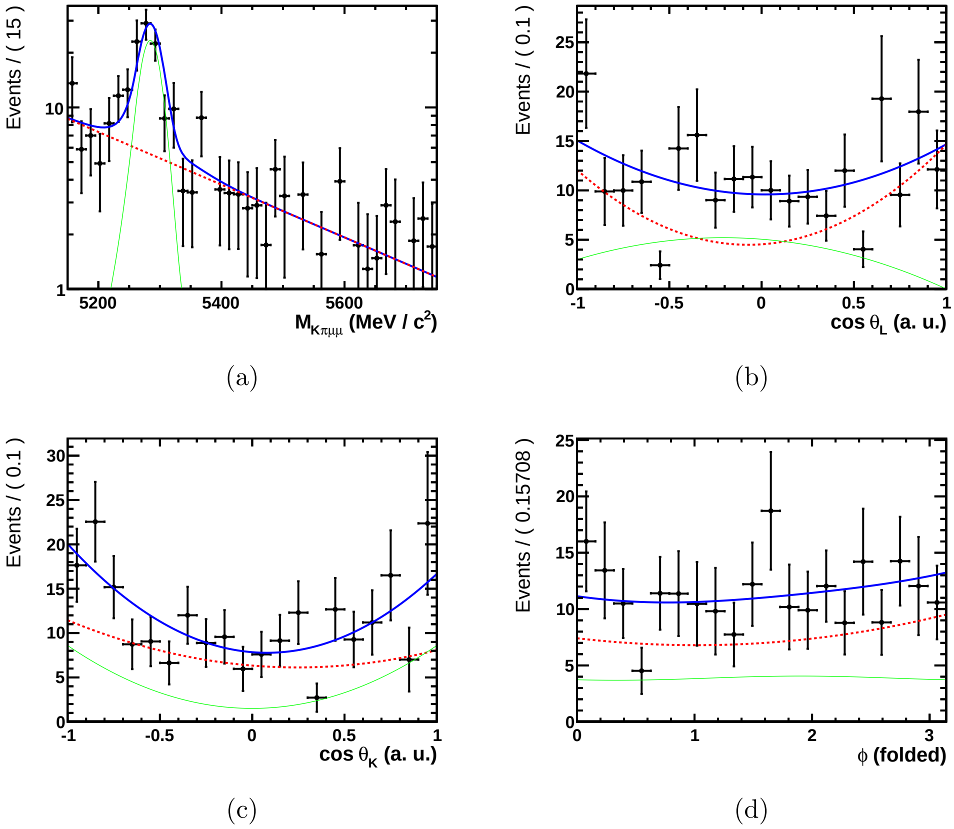


Figure 27: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $2 < q^2 < 4.3\,\mathrm{GeV}^2/c^4$ in the full mass range. The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.

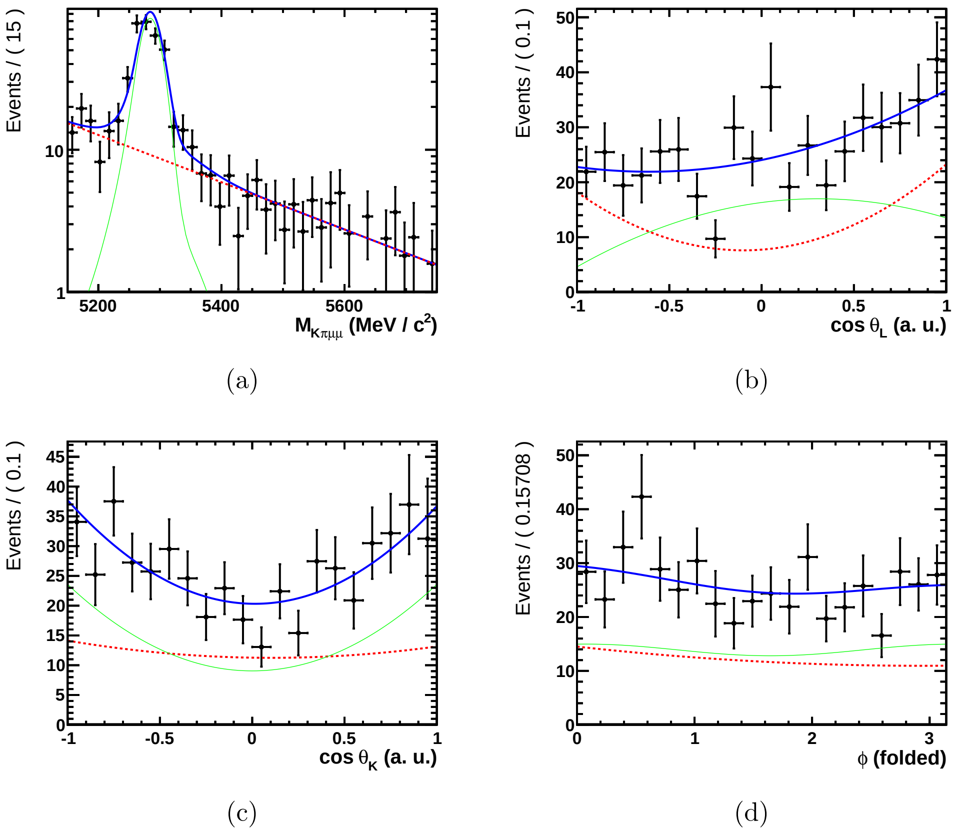


Figure 28: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $4.3 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$ in the full mass range. The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.

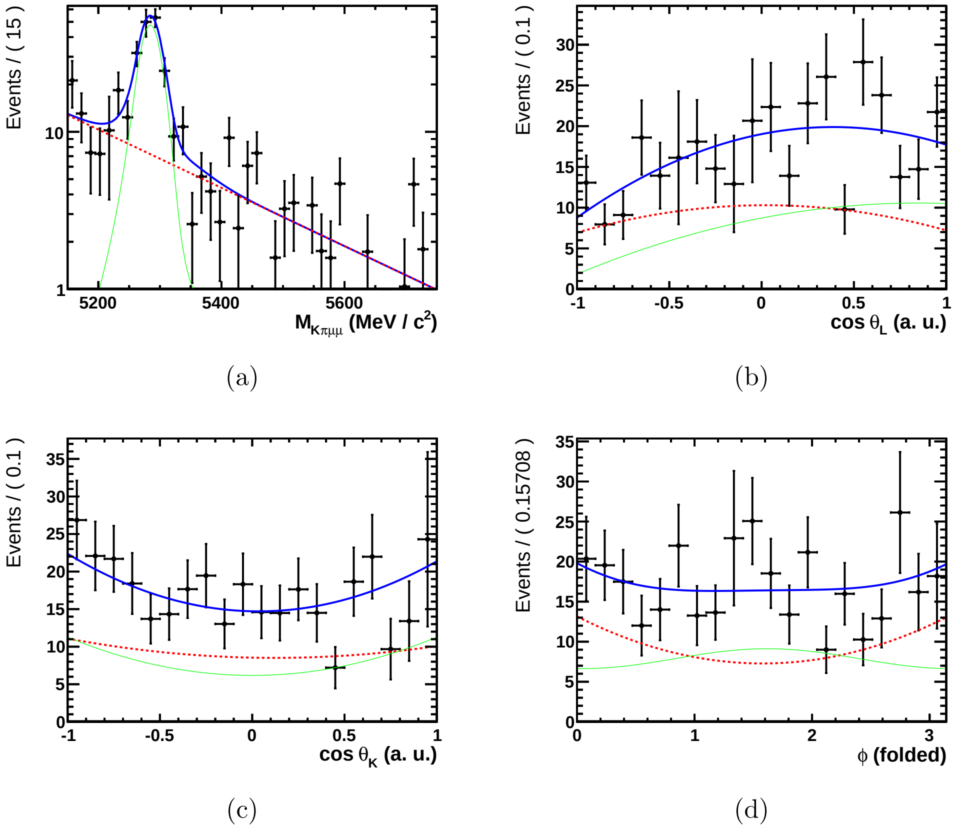


Figure 29: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$ in the full mass range. The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.

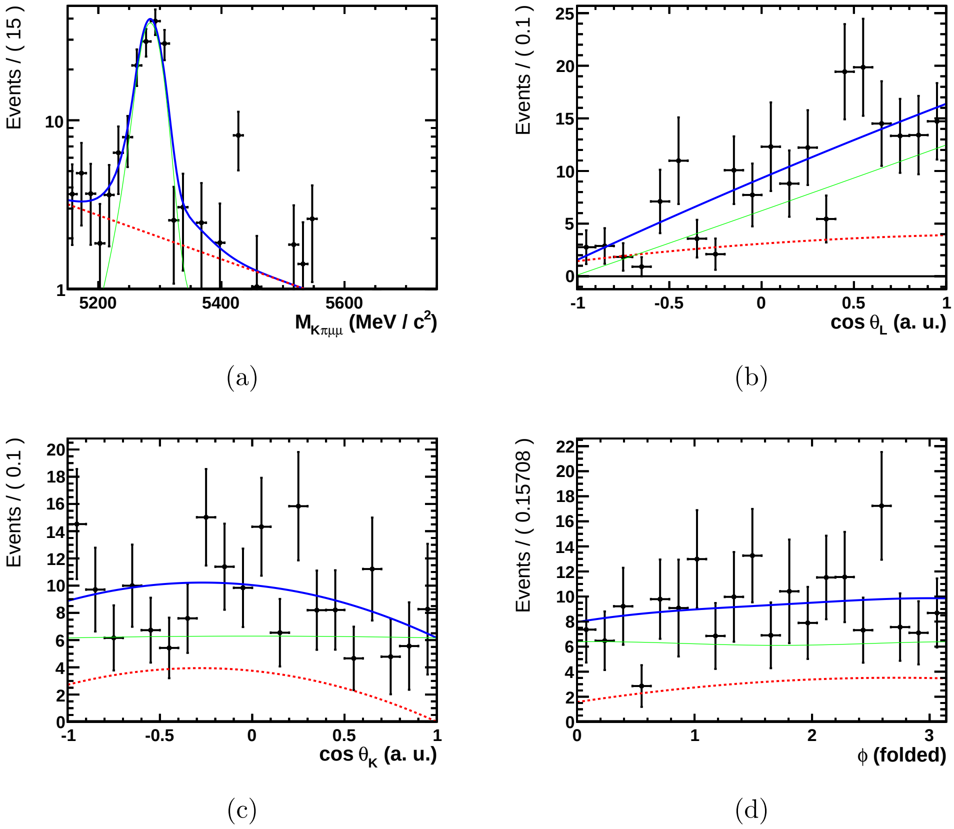


Figure 30: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $14.18 < q^2 < 16\,\mathrm{GeV}^2/c^4$ in the full mass range. The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.

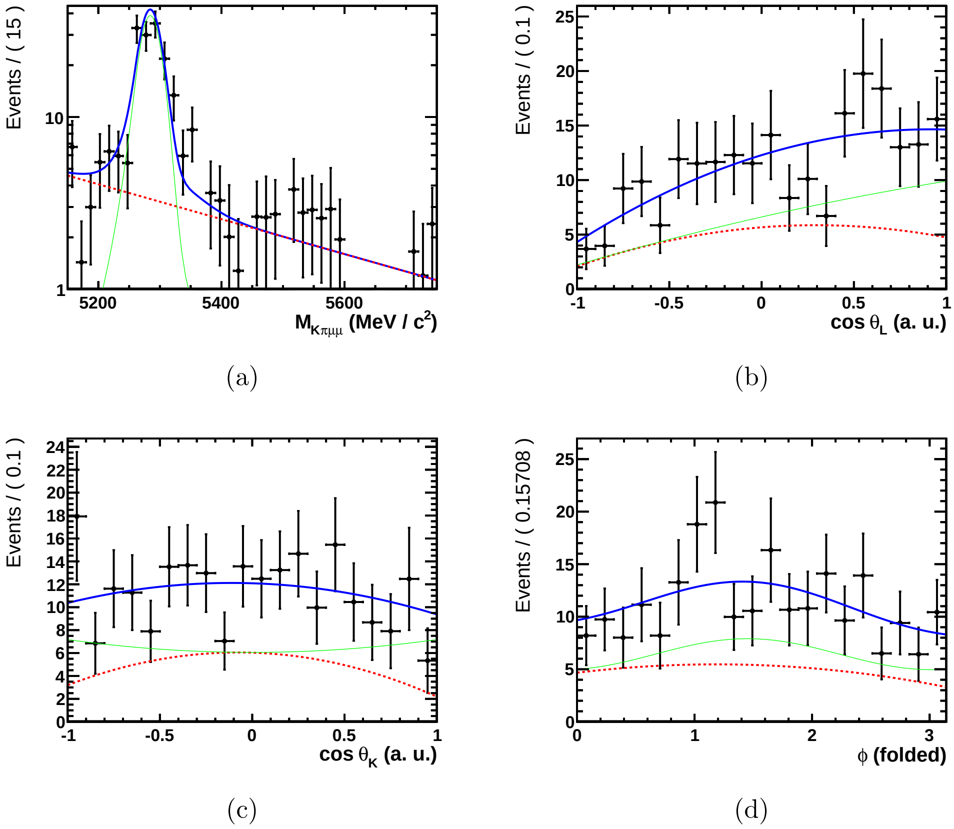


Figure 31: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $16 < q^2 < 19\,\mathrm{GeV}^2/c^4$ in the full mass range. The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.

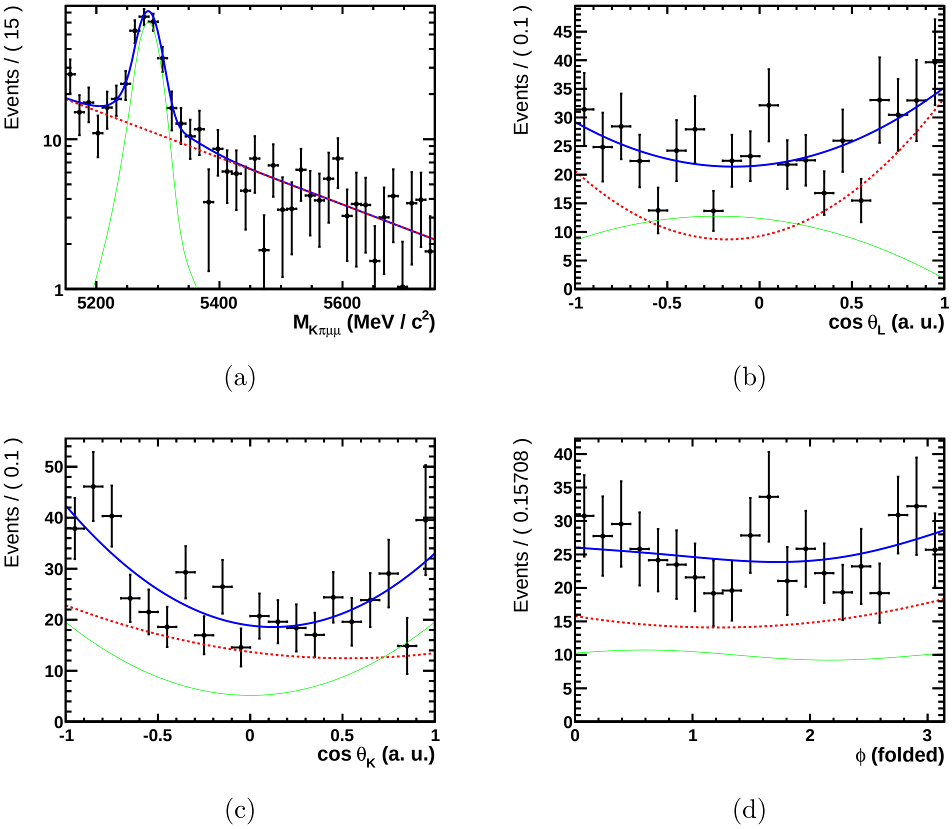
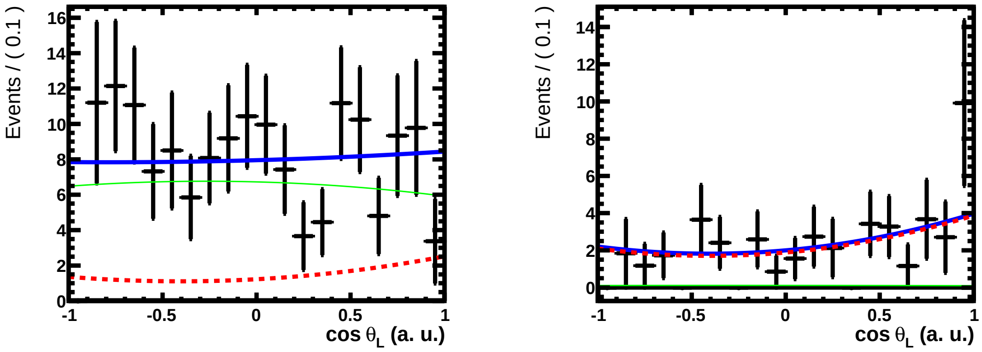
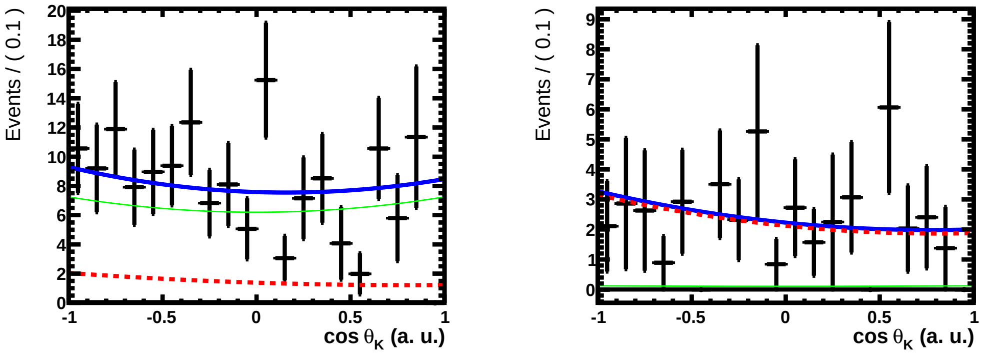


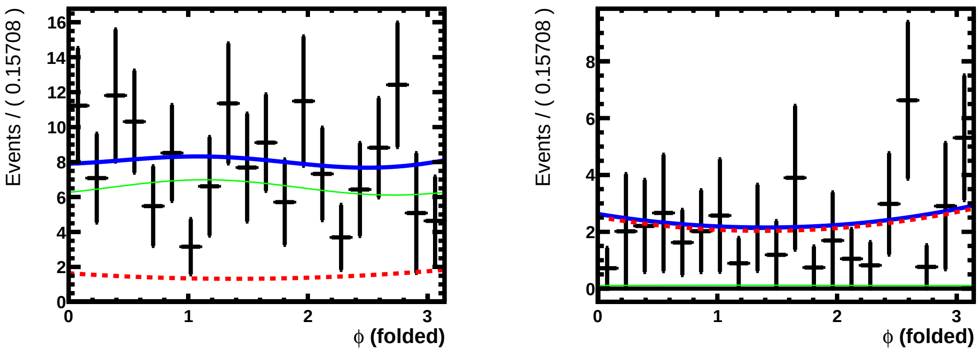
Figure 32: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $1 < q^2 < 6\,\mathrm{GeV}^2/c^4$ in the full mass range. The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.



(a)

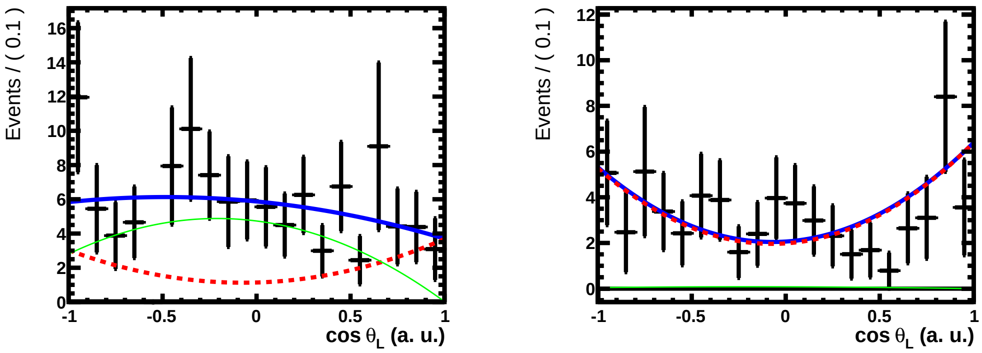


(b)

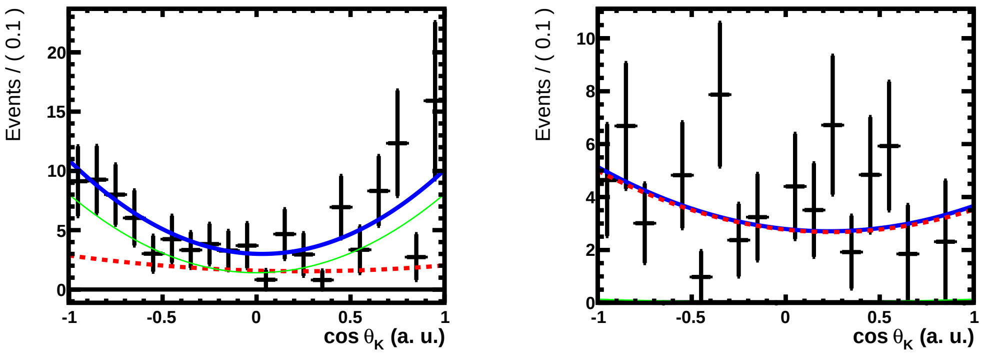


(c)

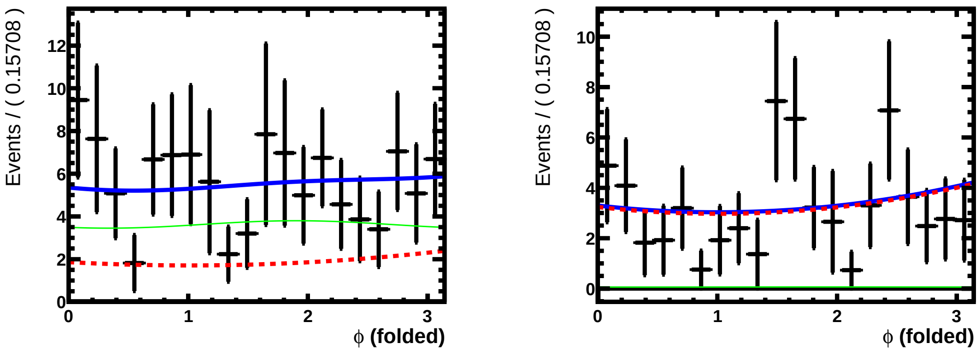
Figure 33: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$ in the signal mass window (left) and upper mass sideband (right). The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.



(a)

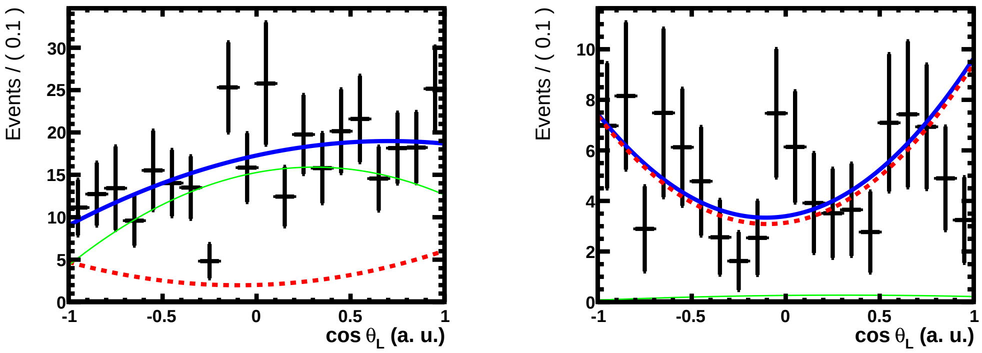


(b)

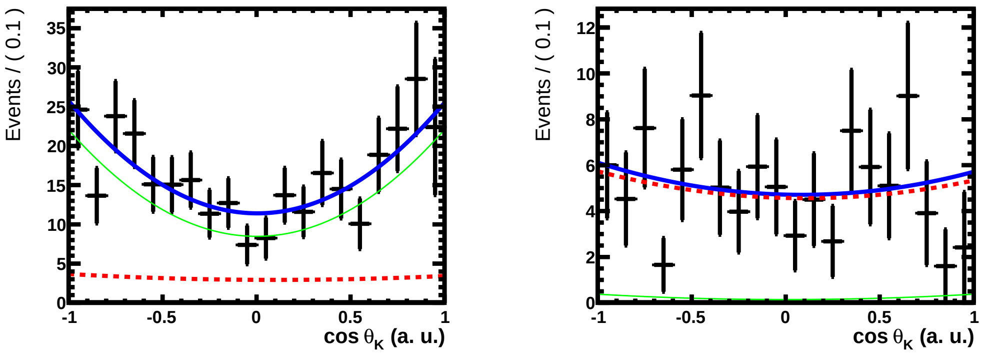


(c)

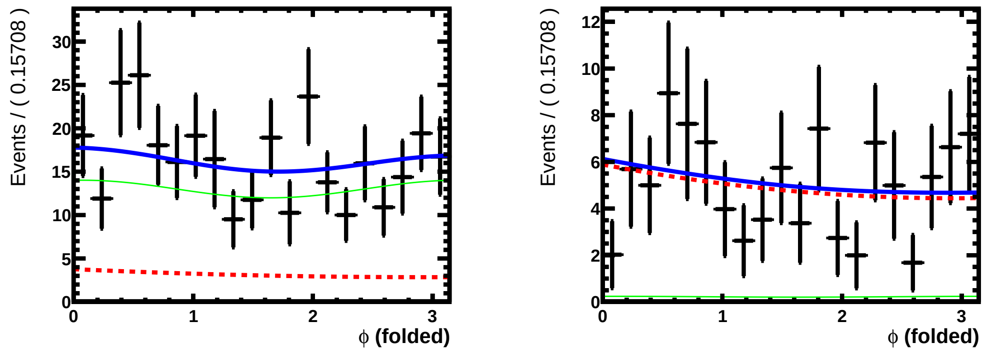
Figure 34: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $2 < q^2 < 4.3\,\mathrm{GeV}^2/c^4$ in the signal mass window (left) and upper mass sideband (right). The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.



(a)

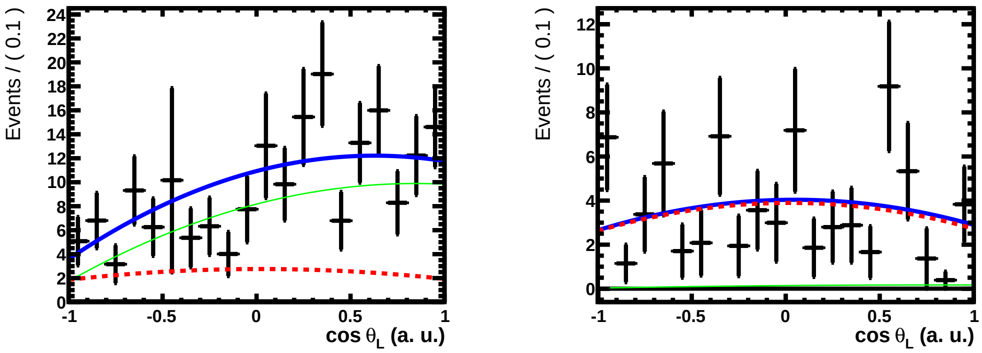


(b)

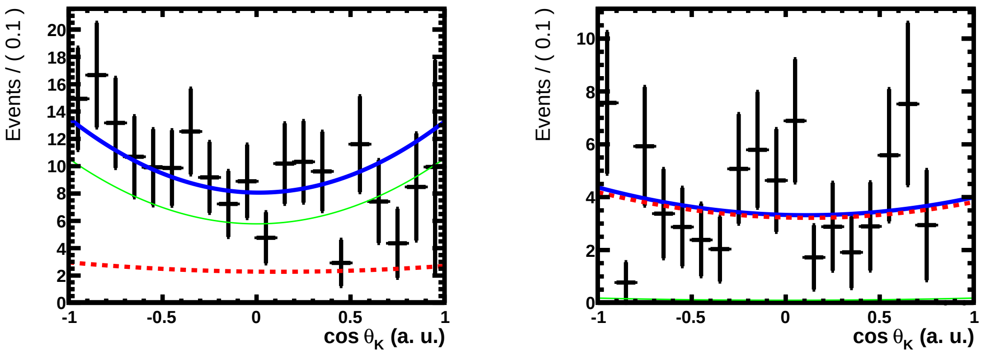


(c)

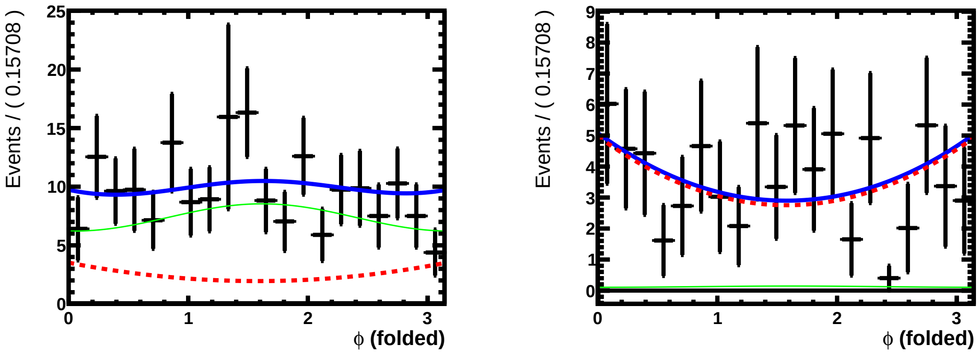
Figure 35: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $4.3 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$ in the signal mass window (left) and upper mass sideband (right). The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.



(a)

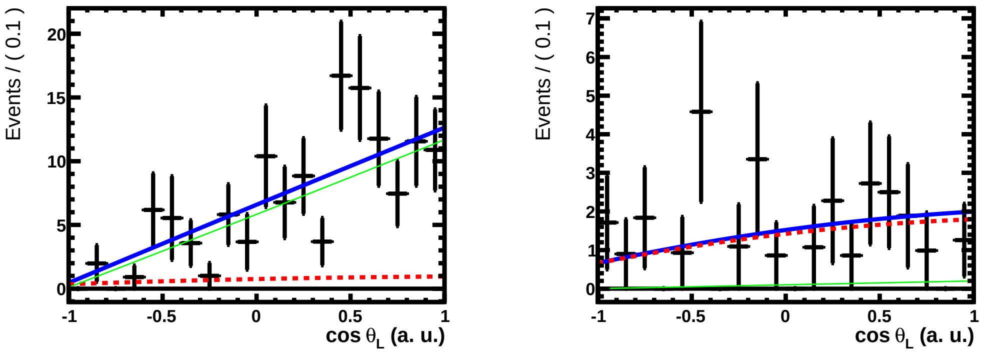


(b)

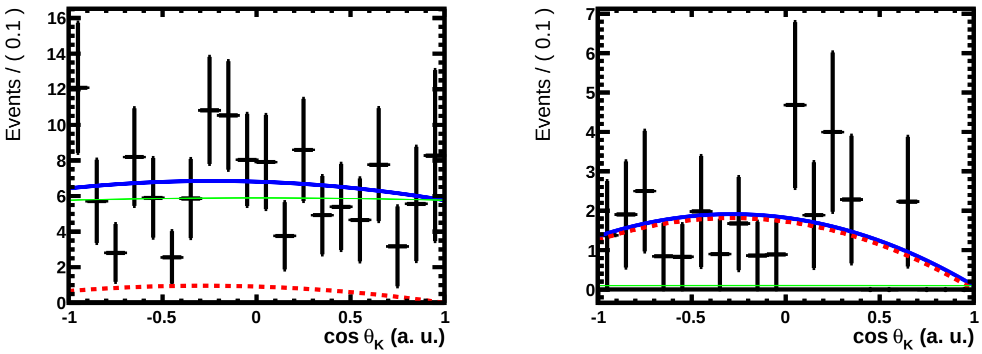


(c)

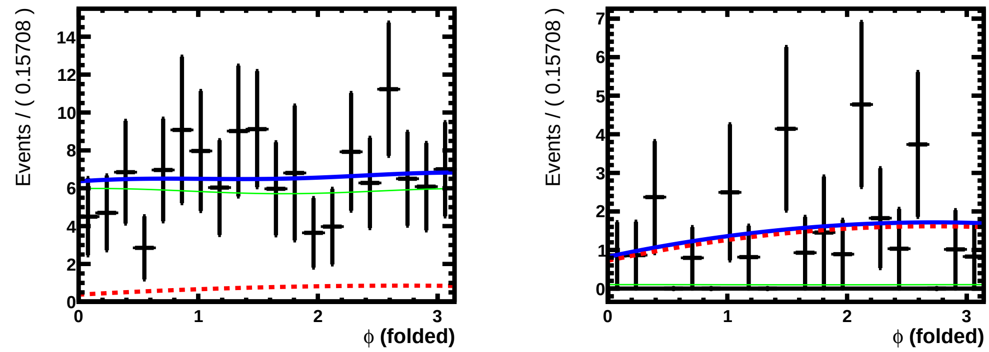
Figure 36: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$ in the signal mass window (left) and upper mass sideband (right). The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.



(a)

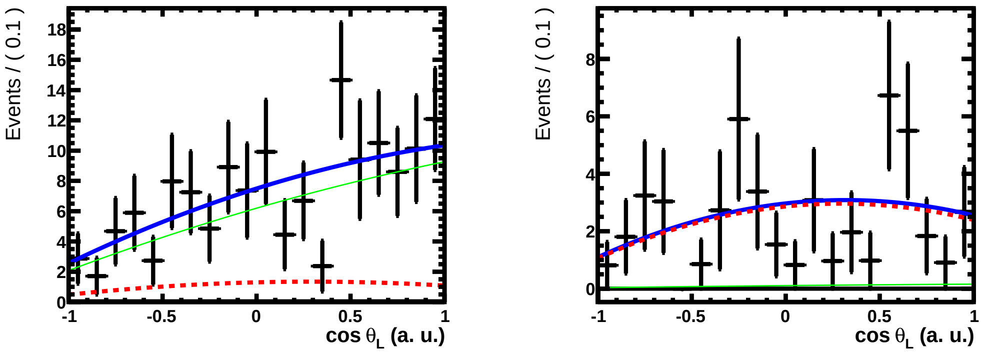


(b)

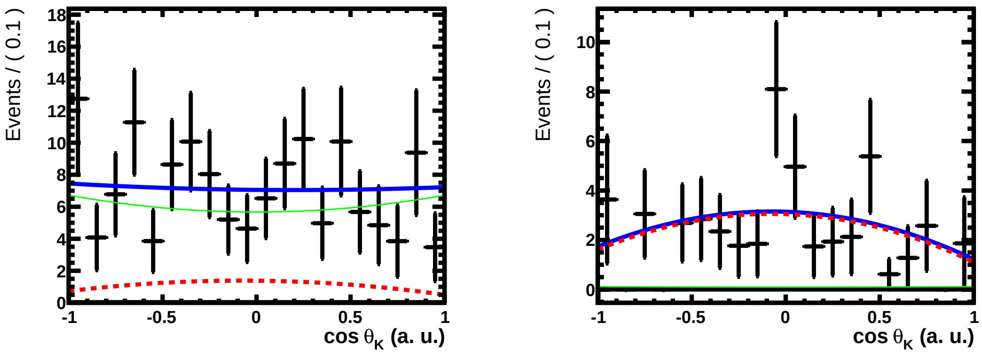


(c)

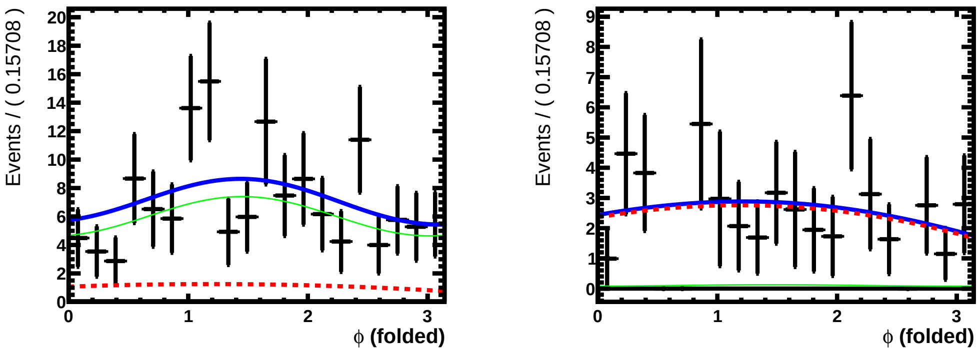
Figure 37: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $14.18 < q^2 < 16\,\mathrm{GeV}^2/c^4$ in the signal mass window (left) and upper mass sideband (right). The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.



(a)



(b)



(c)

Figure 38: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $16 < q^2 < 19\,\mathrm{GeV}^2/c^4$ in the signal mass window (left) and upper mass sideband (right). The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.

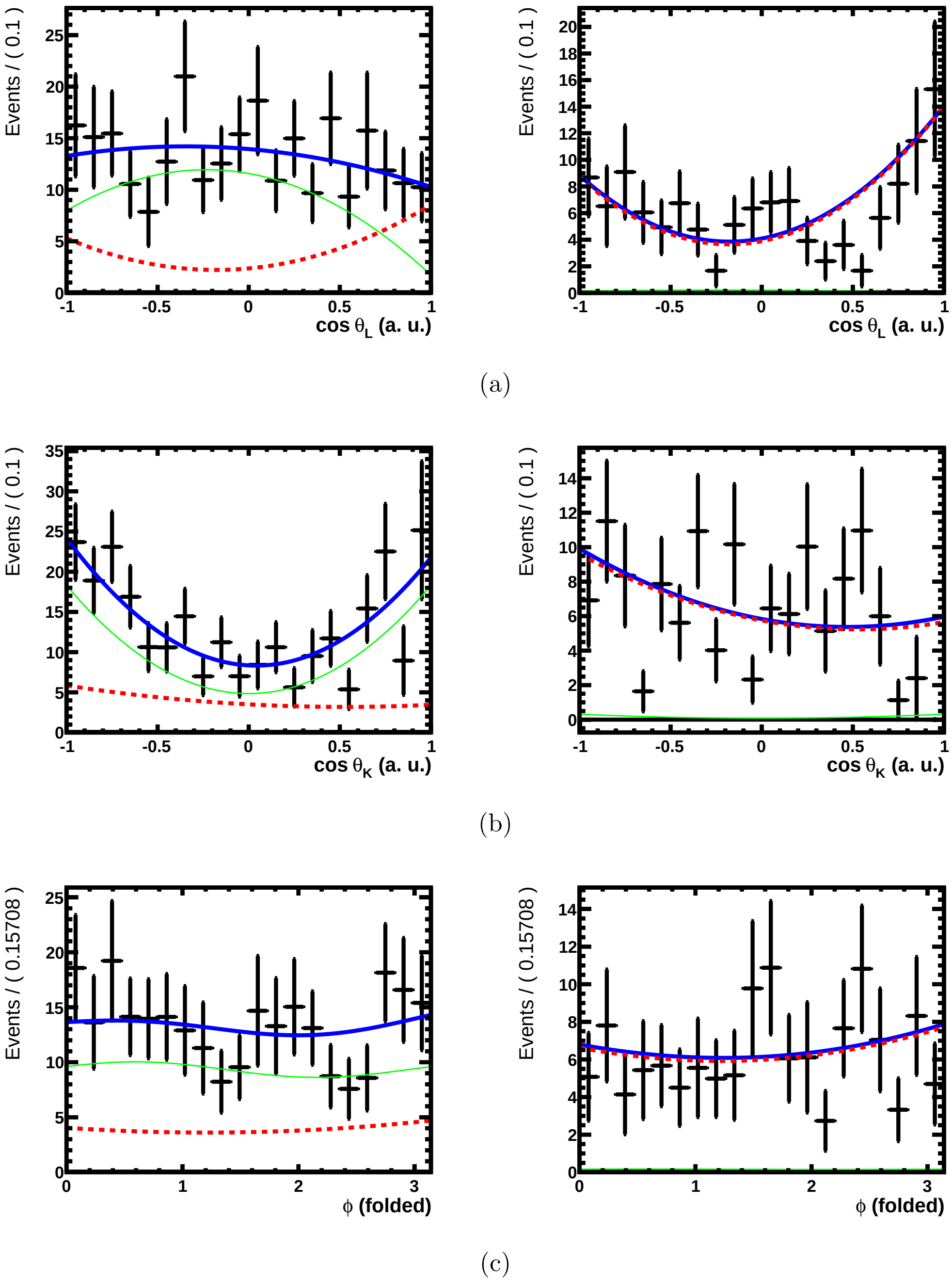


Figure 39: The $K^+\pi^-\mu^+\mu^-$ invariant mass, $\cos\theta_l$, $\cos\theta_K$ and ϕ distribution of candidates with $1 < q^2 < 6\,\mathrm{GeV}^2/c^4$ in the signal mass window (left) and upper mass sideband (right). The blue-line is a fit to the data. The green-line is the signal component and the red-dashed line is the background component.

15.3 Comparison of interval estimates

A comparison of the confidence and credible intervals on A_{FB}, F_{L}, §3, §9 and A_9 is given in Tables. 23 - 27 . In general there is good agreement between the result obtained using the Feldman-Cousins technique and by integrating a 68% credible interval of the profile-likelihood. Differences arise close to the mathematical boundary, due to the different treatment of the boundary effect in the two techniques. In several bins it was not possible to obtain MINOS error estimates directly from `MINUIT` for the lower or upper part of the interval. Most notably in the second and fifth q^2 bin where A_{FB} is very close to the edge of the mathematically defined parameter space.

The confidence and credible intervals can be seen in the plots contained in the webspace area at this location
(`http://www.hep.ph.ic.ac.uk/~cp309/FCandMINOS\_Results/results/`)

q^2 range	FC	MINOS
$0.1 < q^2 < 2.0$	$[-0.14, 0.10]$	$[-0.10, 0.07]$
$2.0 < q^2 < 4.3$	$[-0.28, -0.12]$	$[-0.27, -0.13]$
$4.3 < q^2 < 8.68$	$[0.11, 0.22]$	$[0.11, 0.22]$
$10.09 < q^2 < 12.86$	$[0.22, 0.35]$	$[0.22, 0.35]$
$14.18 < q^2 < 16.$	$[0.46, 0.58]$	$[0.46, 0.56]$
$16. < q^2 < 19.$	$[0.22, 0.38]$	$[0.22, 0.38]$
$1.0 < q^2 < 6.0$	$[-0.23, -0.11]$	$[-0.23, -0.10]$

Table 23: 68% intervals on A_{FB} in the six-plus-one q^2 bins from Feldman-Cousins and MINOS, when fitting for A_{FB}, F_{L}, S_3 and S_9. For more details please see the description in the text.

q^2 range	FC	MINOS
$0.1 < q^2 < 2.0$	$[0.28, 0.47]$	$[0.30, 0.45]$
$2.0 < q^2 < 4.3$	$[0.65, 0.84]$	$[0.65, 0.84]$
$4.3 < q^2 < 8.68$	$[0.50, 0.64]$	$[0.51, 0.63]$
$10.09 < q^2 < 12.86$	$[0.39, 0.56]$	$[0.41, 0.55]$
$14.18 < q^2 < 16.$	$[0.26, 0.41]$	$[0.27, 0.40]$
$16. < q^2 < 19.$	$[0.30, 0.46]$	$[0.30, 0.45]$
$1.0 < q^2 < 6.0$	$[0.58, 0.73]$	$[0.59, 0.73]$

Table 24: 68% intervals on F_{L} in the six-plus-one q^2 bins from Feldman-Cousins and MINOS, when fitting for A_{FB}, F_{L}, S_3 and S_9. For more details please see the description in the text.

q^2 range	FC	MINOS
$0.1 < q^2 < 2.0$	$[-0.14, 0.06]$	$[-0.15, 0.07]$
$2.0 < q^2 < 4.3$	$[-0.10, 0.06]$	$[-0.11, 0.07]$
$4.3 < q^2 < 8.68$	$[0.02, 0.15]$	$[0.01, 0.15]$
$10.09 < q^2 < 12.86$	$[-0.23, -0.05]$	$[-0.23, -0.04]$
$14.18 < q^2 < 16.$	$[-0.07, 0.12]$	$[-0.07, 0.11]$
$16. < q^2 < 19.$	$[-0.31, -0.12]$	$[-0.30, -0.11]$
$1.0 < q^2 < 6.0$	$[-0.04, 0.10]$	$[-0.05, 0.11]$

Table 25: 68% intervals on S_3 in the six-plus-one q^2 bins from Feldman-Cousins and MINOS, when fitting for A_{FB}, F_{L}, S_3 and S_9. For more details please see the description in the text.

q^2 range	FC	MINOS
$0.1 < q^2 < 2.0$	$[-0.04, 0.15]$	$[-0.05, 0.16]$
$2.0 < q^2 < 4.3$	$[-0.07, 0.08]$	$[-0.08, 0.10]$
$4.3 < q^2 < 8.68$	$[-0.05, 0.09]$	$[-0.06, 0.08]$
$10.09 < q^2 < 12.86$	$[-0.12, 0.09]$	$[-0.13, 0.10]$
$14.18 < q^2 < 16.$	$[-0.08, 0.09]$	$[-0.08, 0.10]$
$16. < q^2 < 19.$	$[-0.04, 0.17]$	$[-0.05, 0.17]$
$1.0 < q^2 < 6.0$	$[-0.01, 0.16]$	$[-0.01, 0.16]$

Table 26: 68% intervals on S_9 in the six-plus-one q^2 bins from Feldman-Cousins and MINOS, when fitting for A_{FB}, F_{L}, S_3 and S_9. For more details please see the description in the text.

q^2 range	FC	MINOS
$0.1 < q^2 < 2.0$	$[0.03, 0.21]$	$[0.02, 0.22]$
$2.0 < q^2 < 4.3$	$[-0.02, 0.18]$	$[-0.04, 0.18]$
$4.3 < q^2 < 8.68$	$[-0.20, -0.06]$	$[-0.20, -0.06]$
$10.09 < q^2 < 12.86$	$[-0.11, 0.11]$	$[-0.12, 0.11]$
$14.18 < q^2 < 16.$	$[-0.14, 0.05]$	$[-0.14, 0.04]$
$16. < q^2 < 19.$	$[-0.10, 0.10]$	$[-0.10, 0.11]$
$1.0 < q^2 < 6.0$	$[-0.05, 0.11]$	$[-0.06, 0.11]$

Table 27: 68% intervals on A_9 in the six-plus-one q^2 bins from Feldman-Cousins and MINOS, when fitting for A_{FB}, F_{L}, S_3 and A_9. For more details please see the description in the text.

A comparison of the confidence and credible intervals on A_{T}^{Re}, F_{L}, A_{T}^{2} and A_{T}^{Im} is given in Tables. 28 - 31.

q^2 range	FC	MINOS
$0.1 < q^2 < 2.0$	$[-0.29, 0.21]$	$[-0.22, 0.14]$
$2.0 < q^2 < 4.3$	$[-1.00, -0.87]$	$[-1.00, -0.80]$
$4.3 < q^2 < 8.68$	$[0.36, 0.66]$	$[0.35, 0.66]$
$10.09 < q^2 < 12.86$	$[0.56, 0.86]$	$[0.56, 0.87]$
$14.18 < q^2 < 16.$	$[0.95, 1.00]$	$[0.93, 1.00]$
$16. < q^2 < 19.$	$[0.49, 0.79]$	$[0.49, 0.80]$
$1.0 < q^2 < 6.0$	$[-0.88, -0.42]$	$[-0.91, -0.40]$

Table 28: 68% intervals on A_{T}^{Re} in the six-plus-one q^2 bins from Feldman-Cousins and MINOS, when fitting for A_{T}^{Re}, F_{L}, A_{T}^{2} and A_{T}^{Im}. For more details please see the description in the text.

q^2 range	FC	MINOS
$0.1 < q^2 < 2.0$	$[0.27, 0.48]$	$[0.30, 0.45]$
$2.0 < q^2 < 4.3$	$[0.63, 0.84]$	$[0.65, 0.84]$
$4.3 < q^2 < 8.68$	$[0.50, 0.64]$	$[0.51, 0.63]$
$10.09 < q^2 < 12.86$	$[0.40, 0.56]$	$[0.41, 0.55]$
$14.18 < q^2 < 16.$	$[0.26, 0.41]$	$[0.27, 0.40]$
$16. < q^2 < 19.$	$[0.29, 0.46]$	$[0.30, 0.45]$
$1.0 < q^2 < 6.0$	$[0.58, 0.74]$	$[0.59, 0.73]$

Table 29: 68% intervals on F_{L} in the six-plus-one q^2 bins from Feldman-Cousins and MINOS, when fitting for A_{T}^{Re}, F_{L}, A_{T}^{2} and A_{T}^{Im}. For more details please see the description in the text.

q^2 range	FC	MINOS
$0.1 < q^2 < 2.0$	$[-0.44, 0.20]$	$[-0.48, 0.21]$
$2.0 < q^2 < 4.3$	$[-0.75, 0.36]$	$[-0.88, 0.45]$
$4.3 < q^2 < 8.68$	$[0.05, 0.66]$	$[0.03, 0.67]$
$10.09 < q^2 < 12.86$	$[-0.87, -0.18]$	$[-0.87, -0.17]$
$14.18 < q^2 < 16.$	$[-0.21, 0.33]$	$[-0.21, 0.34]$
$16. < q^2 < 19.$	$[-0.97, -0.36]$	$[-0.96, -0.37]$
$1.0 < q^2 < 6.0$	$[-0.24, 0.56]$	$[-0.31, 0.64]$

Table 30: 68% intervals on A_{T}^{2} in the six-plus-one q^2 bins from Feldman-Cousins and MINOS, when fitting for A_{T}^{Re}, F_{L}, A_{T}^{2} and A_{T}^{Im}. For more details please see the description in the text.

q^2 range	FC	MINOS
$0.1 < q^2 < 2.0$	$[-0.12, 0.47]$	$[-0.17, 0.51]$
$2.0 < q^2 < 4.3$	$[-0.50, 0.54]$	$[-0.59, 0.72]$
$4.3 < q^2 < 8.68$	$[-0.26, 0.36]$	$[-0.28, 0.39]$
$10.09 < q^2 < 12.86$	$[-0.47, 0.37]$	$[-0.51, 0.39]$
$14.18 < q^2 < 16.$	$[-0.25, 0.29]$	$[-0.26, 0.30]$
$16. < q^2 < 19.$	$[-0.14, 0.53]$	$[-0.16, 0.53]$
$1.0 < q^2 < 6.0$	$[-0.04, 0.83]$	$[-0.07, 0.87]$

Table 31: 68% intervals on A_{T}^{Im} in the six-plus-one q^2 bins from Feldman-Cousins and MINOS, when fitting for A_{T}^{Re}, F_{L}, A_{T}^{2} and A_{T}^{Im}. For more details please see the description in the text.

15.4 Feldman Cousins CL at the SM point

As a measure of the consistency of the angular fit results and the SM prediction, the Feldman Cousins CL for the SM point was calculated. In contrast to the one dimensional FC confidence intervals, this CL is calculated varying all four angular observables simultaneously.

One thousand toy datasets were generated at the SM-predicted central values of the angular observables $\{A_{\mathrm{FB}},\ F_{\mathrm{L}},\ S_3,\ S_9\}$ in each q^2 bin. The standard angular fit is performed on each toy dataset and the value of the likelihood, R_0 is recorded. Another angular fit is performed with the angular observables fixed to their SM-predicted values and the value of the likelihood for this fit (R_1) is recorded. The likelihood ratio $R_{\mathrm{toy}} = R_0/R_1$ is then calculated. The same procedure is performed on fits to candidates from the data to obtain the likelihood ratio R_{data}. The p-value is then calculated by integrating the distribution of 1000 R_{toy} values from R_{data} to infinity. The same procedure is repeated for the set of angular observables $\{A_{\mathrm{T}}^{Re},\ F_{\mathrm{L}},\ A_{\mathrm{T}}^{2},\ A_{\mathrm{T}}^{Im}\}$.

The resulting CLs are summarised in Tab. 32. No results are presented for the $10.09 < q^2 < 12.86$ bin as no SM prediction is available in this q^2 region.

Differences can arise between the two sets of CL-values for two reasons: small differences can arise due to limited number of pseudo-experiments that are generated; larger differences can arise in the second q^2 bin due to the influence of the boundaries on the toy experiments.

$q^2(\mathrm{GeV}^2/c^4)$	CL for $\{A_{\mathrm{FB}},\ F_{\mathrm{L}},\ S_3,\ S_9\}$	CL for $\{A_{\mathrm{T}}^{Re},\ F_{\mathrm{L}},\ A_{\mathrm{T}}^{2},\ A_{\mathrm{T}}^{Im}\}$	p-value
$0.10 < q^2 < 2.00$	0.16	0.18	0.21
$2.00 < q^2 < 4.30$	0.50	0.57	0.32
$4.30 < q^2 < 8.68$	0.68	0.71	0.65
$10.09 < q^2 < 12.86$	NA	NA	NA
$14.18 < q^2 < 16.00$	0.39	0.38	0.79
$16.00 < q^2 < 19.00$	0.28	0.28	0.05
$1.00 < q^2 < 6.00$	0.67	0.72	0.48

Table 32: Angular analysis CLs at the SM point and p-value for the set of observables $\{A_{\mathrm{FB}},\ F_{\mathrm{L}},\ S_3,\ S_9\}$ and $\{A_{\mathrm{T}}^{Re},\ F_{\mathrm{L}},\ A_{\mathrm{T}}^{2},\ A_{\mathrm{T}}^{Im}\}$ in each analysis q^2 bin.

15.5 Extracting the p-value for the SM point

The p-value of the SM point (including the background description) has also been estimated using an unbinned goodness of fit test (point-to-point dissimilarity test [1]). The test is performed only considering the angular phase-space defined by $\cos\theta_l$, $\cos\theta_K$ and ϕ. A weighting function of the form $\Psi = e^{-x^2/2\sigma^2}$ is used, where σ is defined such that Ψ covers 5% of the angular phase-space. The results of this test are summarised in Table. 32. In all cases the results indicate that the fit model at the SM point is a reasonable description of the data. The test was repeated with Ψ covering 10% of the phase-space, with no change in the conclusion.

Note, for the results in Table. 32 there is also a reasonably large uncertainty on what is meant by the SM point, coming from theoretical uncertainties and differences between different theory predictions.

16 Introducing a $K^+\pi^-$ system S-wave

The inclusion of a spin-0 $K^+\pi^-$ component to the $K^+\pi^-$ system, that can interfere with the $K^{*0}(892)$, is motivated by the analysis of the angular and mass distribution of $B^0 \to K^{*0} J/\psi$ decays (see for example Ref. [21]). The impact of the S-wave is evaluated and treated as systematic uncertainty on the differential branching fraction and angular observables. The size of this systematics is evaluated from the signal data. A 68% CL upper limit for the S-wave in the region $1 - 6\mathrm{GeV}^2$ is estimated. This value is conservatively used as a systematic uncertainty. More details can be found in the following sections.

16.1 Impact on the angular distributions: formalism

When taking into account this new spin-0 component, the longitudinal amplitude is replaced in the angular expression by the sum of two terms: the usual one, $A_{0L/R}$ which corresponds to the longitudinal polarisation amplitude of the K^{*0} (which has a Breit Wigner dependence as function of the $K^+\pi^-$ mass) and a second amplitude , A^0_{0L}, corresponding to the S-wave contribution. This new amplitude at first approximation can be assumed to be constant over the $\pm 100\,\mathrm{MeV}/c^2$ interval around the K^{*0} mass used in this analysis.

Explicitly, this corresponds to the transformation:

$$A_{0,L/R} \cos\theta_K \to \frac{1}{\sqrt{3}} A^0_{0,L/R} + A_{0,L/R} \cos\theta_K$$

The immediate impact of the additional left- and right-handed S-wave amplitudes is to modify Γ such that[2]:

$$\Gamma = |A^0_0|^2 + |A_0|^2 + |A_\parallel|^2 + |A_\perp|^2 = |A^0_0|^2 + \Gamma'$$

where A^0_0 is the amplitude for the S-wave component. This will modify the standard observables, leading to:

[2]The discussion of the S-wave is largely based on Ref. [24]

$$
\begin{aligned}
A_{\mathrm{FB}} &= \frac{3}{4}\frac{Re(A_{\parallel L}A^*_{\perp L}) - Re(A_{\parallel R}A^*_{\perp R})}{\Gamma'} \\
&= \frac{3}{4}\frac{Re(A_{\parallel L}A^*_{\perp L}) - Re(A_{\parallel R}A^*_{\perp R})}{\Gamma(1-F_S)}
\end{aligned}
$$

$$
F_{\mathrm{L}} = \frac{|A_0|^2}{\Gamma'} = \frac{|A_0|^2}{\Gamma(1-F_S)}
$$

$$
\begin{aligned}
A_{\mathrm{Im}} &= \frac{Im(A_{\parallel L}A^*_{\perp L}) + Im(A_{\parallel R}A^*_{\perp R})}{\Gamma'} \\
&= \frac{Im(A_{\parallel L}A^*_{\perp L}) + Im(A_{\parallel R}A^*_{\perp R})}{\Gamma(1-F_S)} \\
S_3 &= \frac{1}{2}\frac{|A_{\perp L}|^2 - |A_{\parallel L}|^2 + |A_{\perp R}|^2 - |A_{\parallel R}|^2}{\Gamma'} \\
&= \frac{1}{2}\frac{|A_{\perp L}|^2 - |A_{\parallel L}|^2 + |A_{\perp R}|^2 - |A_{\parallel R}|^2}{\Gamma(1-F_s)}
\end{aligned}
$$

where A_{FB}, A_{Im}, S_3 and F_L remain defined w.r.t. the K^{*0} and

$$
F_S = |A_0^0|^2/\Gamma
$$

is the fractional contribution of the S-wave amplitude and is expected to be small. There is also a new forward-backward asymmetry, A_S that appears in the kaon angle. This comes from interference between the S-wave amplitude and the longitudinal K^{*0} amplitude,

$$
A_S = \frac{1}{\Gamma}\sqrt{3}\left[|A_{0,L}||A^0_{0,L}|\cos\delta_L + |A_{0,R}||A^0_{0,R}|\cos\delta_R\right] \ .
$$

Interference terms between $A^0_{0,L/R}$ and $A_{\perp,L/R}$ or $A_{\parallel,L/R}$ are removed by the $\hat{\phi}$ transformation. Accounting for the S-wave amplitude, the 'folded' angular distribution can be written:

$$\frac{1}{\Gamma}\frac{d^4\Gamma}{dq^2\, d\cos\theta_K\, d\cos\theta_l\, d\hat{\phi}} = \frac{9}{16\pi}\Bigg[\ \frac{2}{3}F_{\rm S}(1-\cos^2\theta_l) + \frac{4}{3}A_{\rm S}\cos\theta_K(1-\cos^2\theta_l)\ +$$
$$2(1-F_{\rm S})F_{\rm L}\cos^2\theta_K(1-\cos^2\theta_l)\ +$$
$$\frac{1}{2}(1-F_{\rm S})(1-F_{\rm L})(1-\cos^2\theta_K)(1+\cos^2\theta_l)\ +$$
$$(1-F_{\rm S})S_3(1-\cos^2\theta_K)(1-\cos^2\theta_l)\cos 2\hat{\phi}\ +$$
$$\frac{4}{3}(1-F_{\rm S})A_{\rm FB}(1-\cos^2\theta_K)\cos\theta_l\ +$$
$$(1-F_{\rm S})A_{\rm Im}(1-\cos^2\theta_K)(1-\cos^2\theta_l)\sin 2\hat{\phi}\ \Bigg]\ .$$

The one dimensional projections of the angular distribution are given by :

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\theta_l} = \frac{3}{4}\left[F_S + (1-F_S)F_L\right]\left[1-\cos^2\theta_\ell\right]\ +$$
$$\frac{3}{8}\left[(1-F_S)(1-F_L)\right]\left[1+\cos^2\theta_\ell\right] + (1-F_S)A_{FB}\cos\theta_\ell$$

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\theta_K} = \frac{F_S}{2} + A_S\cos\theta_K$$
$$\frac{3}{2}(1-F_S)F_L\cos^2\theta_K + \frac{3}{4}\left[(1-F_S)(1-F_L)\right]\left[1-\cos^2\theta_K\right]\ .$$

16.2 Exploiting the phase change across the Breit-Wigner to estimate the S-wave

The size of the interference term, A_S, depends on the relative strong phase difference between A_0 and A_0^0 and on F_S and F_L. Ignoring for the moment the left- and right-handedness of the amplitudes, the maximum possible size of A_S is bounded by the size of F_S and F_L:

$$|A_S| \leq \sqrt{3}(F_S(1-F_S)F_L)^{1/2}\ .$$

For a non-relativistic Breit-Wigner distribution, A_0 can be split into real and imaginary parts:

$$Re(A_0(m_{K^+\pi^-})) = \frac{a}{1+a^2} \quad\text{and}\quad Im(A_0(m_{K^+\pi^-})) = \frac{i}{1+a^2}$$

where

$$a = \frac{m_{K^+\pi^-} - m_{K^{*0}}}{\Gamma/2} \quad ,$$

and $m_{K^{*0}}$ is the pole mass of the K^{*0} Breit-Wigner. In terms of $Re(A_0^0)$, $Im(A_0^0)$, $Re(A_0)$ and $Im(A_0^0)$, A_S becomes:

$$A_S(a) \propto Re(A_0^0)Re(A_0) + Im(A_0^0)Im(A_0)$$

There is also a phase change of A_0 between the left- and right-hand side of the Breit-Wigner. If $Re(A_0^0)$ and $Im(A_0^0)$ are assumed to be constant across the $\pm 100\,\text{MeV}/c^2$ mass window used in the analysis, then the phase change of the Breit-Wigner, of A_0, can be exploited to measure the size of F_S from the asymmetry in $\cos\theta_K$ for events above and below the K^{*0} pole mass.

If the average values of A_S in the $100\,\text{MeV}/c^2$ window above and below the pole mass are A_+ and A_-, then $A_+ \pm A_-$ can be used to isolate $Re(A_0)$ and $Im(A_0)$ parts of the Breit-Wigner. Further it can be shown that:

$$< F_{\rm S} >= \frac{[(A_+ + A_-)^2/4 + (A_+ - A_-)^2/(4 \times 1.23)] \times 3.24/(3F_{\rm L})}{1 - [(A_+ + A_-)^2/4 + (A_+ - A_-)^2/(4 \times 1.23)] \times 3.24/(3F_{\rm L})} \quad (4)$$

where the numerical term are obtained, after integration, for $\frac{\Gamma}{2} = 26\,\text{MeV}/c^2$. The measurement of F_S that comes from A_+ and A_- is statistically more precise than simply fitting directly for F_S and A_S as independent variables because the measurement is based on a sizable interference term, rather than a measurement of a small extra amplitude – in simpler terms A_S can be more precisely determined that F_S.

The procedure has been validated with a large statistics sample of $B^0 \to K^{*0} J/\psi$ events comparing the calculated $< F_{\rm S} >$ to the fitted $F_{\rm S}$, as shown in Section F.

Given the good results obtained for the $B^0 \to J/\psi K^{*0}$ decay, the procedure can been applied to $B^0 \to K^{*0}\mu^+\mu^-$. This has been done for two different q^2 ranges: 1-19 GeV^2/c^4 and also 1-6 GeV^2/c^4. Unfortunately, in the latter case the statistics is too low for the fit (with the S-wave parameters) to converge successfully. To reduce the number of parameters we integrate over the ϕ angle. This does not change the sensitivity to the S-wave parameters (the sensitivity to which comes mainly through $\cos\theta_K$) and removes two angular observables, simplifying the fit.

The result of the fit in the q^2 region from 1 to 19 GeV^2/c^4 and 1 to 6 GeV^2/c^4, excluding the J/ψ and $\psi(2S)$, is given on Table 33 and Figures 79,80, 81, 82. The values of F_S have been computed assuming Gaussian distributed errors on F_L and $A_S^{\pm}$. The same results are obtained by doing a profile likelihood scan.

If the S-wave contribution is fixed to 0, the $F_{\rm L}$ value is 0.52 ± 0.03 and 0.68 ± 0.06 for the 1 to 19 GeV^2/c^4 and 1 to 6 GeV^2/c^4 regions respectively. Consistent with the nominal fit results.

In the high K^{*0} energy approximation $F_{\rm S}$ is expected to have the same q^2 dependence as $F_{\rm L}$ (driven by the q^2 dependence of the transverse amplitudes). This implies that taking the 68% CL upper limit in the region 1-6 GeV^2/c^4 as a systematic is a conservative estimate for every bin, since $F_{\rm L}$ is largest in this region.

		$1 < q^2 < 19\,\mathrm{GeV}^2/c^4$	$1 < q^2 < 6\,\mathrm{GeV}^2/c^4$
Fitted parameters	A_T^{Re}	0.619 ± 0.088	-0.490 ± 0.293
	F_L	0.523 ± 0.031	0.700 ± 0.066
	A_S^+	-0.025 ± 0.051	0.003 ± 0.109
	A_S^-	-0.162 ± 0.058	-0.228 ± 0.119
Using eq 4	$< F_{\rm S} >$	0.025 ± 0.018	0.038 ± 0.043
		(< 0.04 at 68% CL)	(< 0.07 at 68% CL)

Table 33: Fit results for F_S and $A_S^{\pm}$ in the q^2 region from 1 to 19 GeV^2/c^4 and 1 to 6 GeV^2/c^4 when including the S-wave.

17 Correction for the threshold terms

In the angular fit we neglect lepton masses. This assumption holds everywhere apart the first q^2 bin. When muon masses are not neglected, terms with additional q^2-dependence appear. The effect of neglicting these terms is corrected for a posteriori as discussed in the next sections. This correction roughly corresponds to a 10-20% factor for all observables, apart for F_L for which this effect is negligible.

17.1 Procedure to correct for the threshold terms

Since we do not have yet enough data to perform a complete parametrisation as a function of the dimuon invariant mass squared, the only way the dependence on q^2 is taken into account in the analysis is by performing the fit separately in wide bins of q^2. In each of these bins, the resulting "physics" parameters represent an average over that q^2 bin.

If we revisit the full PDF for the angular distribution then a q^2-dependence arises from three separate places:

1. the q^2 dependence of the form factors;
2. an explicit dependence on q^2 that accompanies $\mathcal{C}_7$ and $\mathcal{C}_7'$;
3. threshold terms that depend on $x = 4m_\mu^2/q^2$ in the angular distribution.

One and two can be associated with the q^2 dependence of the amplitudes, or equally of the observables. The third type of q^2 dependence has until now been completely neglected. These threshold terms are negligible at high q^2 where $q^2 \gg m_\mu^2$ and $x \to 0$, but may become significant as $q^2 \to 0$, in particular in the $0 < q^2 < 2\,\mathrm{GeV}^2/c^4$ bin. If we revisit the angular distribution, the impact of the threshold terms is to modify I_1 through I_9 as:

$$\begin{aligned}
I_1^s &= \frac{3}{4}\left[1 - \frac{x}{3}\right](|A_\parallel|^2 + |A_\perp|^2) + \frac{x}{2}(|A_\parallel|^2 + |A_\perp|^2)\\
I_1^c &= [1 + x]\,|A_0|^2\\
I_2^s &= \frac{1}{4}\,[1 - x]\,(|A_\parallel|^2 + |A_\perp|^2)\\
I_2^c &= -\,[1 - x]\,|A_0|^2\\
I_3 &= \frac{1}{2}\,[1 - x]\,(|A_\perp|^2 - |A_\parallel|^2)\\
I_6 &= 2\left[\sqrt{1 - x}\right] Re(A_{\parallel L}A_{\perp L}^* - A_{\parallel R}A_{\perp R}^*)\\
I_9 &= [1 - x]\,Im(A_{\parallel L}A_{\perp L}^* + A_{\parallel R}A_{\perp R}^*)
\end{aligned} \tag{5}$$

As $x \to 1$, the angular distribution actually becomes isotropic in $\cos\theta_\ell$, $\cos\theta_K$ and ϕ and we lose all sensitivity to the observables.

These new terms create a problem for the q^2 averaging (see Sec. 8.7). Unfortunately, as a result of neglecting the threshold terms, in the fit to the data in the $0 < q^2 < 2\,\mathrm{GeV}^2/c^4$ bin, the measured values of the physics parameters will be a biased estimate of the pure physics quantities predicted by theory. A procedure to estimate this bias is described below.

17.2 Correction procedure

Integrating the full angular expression over $\cos\theta_l$, $\cos\theta_K$ and ϕ, yields:

$$\Gamma = \left[1 + \frac{x}{2}\right]\left(|A_\parallel|^2 + |A_\perp|^2 + |A_0|^2\right) \ .$$

The individual terms in the angular distribution can also be updated to include a dependence on x, e.g.

$$\frac{I_3}{\Gamma} = \frac{\frac{1}{2}(1-x)((|A_\perp|^2 - |A_\parallel|^2)}{(1+\frac{x}{2})(|A_\parallel|^2 + |A_\perp|^2 + |A_0|^2)} = \frac{(1-x)}{(1+\frac{x}{2})}\frac{1}{2}A_T^2(1-F_\mathrm{L}) = \beta(q^2)A_T^2(q^2)(1-F_\mathrm{L}(q^2)).$$

When averaging over the $0 < q^2 < 2\,\mathrm{GeV}^2/c^4$ bin, there are now three q^2 dependent terms to worry about. As a reminder, in the simpler case when ignoring the threshold terms there are two q^2 dependent terms F_L and A_T^2. In this case the fit is sensitive to a rate average of $A_T^2(q^2)$, where you sum over narrow q^2 bins , q_i^2, weighting A_T^2 by $N(q_i^2)(1 - F_L(q_i^2))$. Now that there are three q^2 dependent terms some assumption needs to be made on the q^2 dependence of the observables in order to unfold the effect of the $x-$dependence from the measured observables.

The only physics parameter that is not biased by the threshold effect is F_L. F_L is essentially determined by the $cos(\theta_K)$ distribution which take the form:

$$\frac{4}{3}(1+\frac{x}{2})[2(F_L)cos^2(\theta_K) + (1-F_L)sin^2(\theta_K)] \qquad (6)$$

This expression is obtained from the full angular distribution neglecting the ϕ depending terms and integrating over $cos(\theta_\ell)$. While Eq. 6 depends on x, it does not not change the shape of the distribution, only the amplitude. So, the threshold term has no impact on F_L.

To correct the other physics parameters for the threshold effects and obtain the true average, one needs to model the q^2 dependence of the physics parameters in the bin. A first approximation is to take A_T^2 and A_T^{Im} as constant and A_T^{Re} as rising linearly, since it must be 0 at $q^2 = 0$.

To do the weighting, one also needs to model the q^2 variation of the transverse width. This can be achieved by using the experimental distribution of the events as function of q^2 weighted by the term $(1 - F_L)$, modeling a plausible variation of $(1 - F_L)$ as function of q^2, as for example:

$$F_L(q_i^2) = \frac{aq_i^2}{1 + aq_i^2} \tag{7}$$

This parameterisation of F_{L} is "physics" inspired. F_{L} changes rapidly in q^2 at low q^2 but must become zero as $q^2 \to 0$ (the photon is transversely polarised). It is also expected (in all models) to rise smoothly across the $0 < q^2 < 2\,\mathrm{GeV}^2/c^4$ bin.

17.2.1 Correction factors

To first approximation, by neglecting the threshold terms we have underestimated the size of the angular observables in the $0 < q^2 < 2\,\mathrm{GeV}^2/c^4$ bin. The multiplicative correction factors needed to correct our measurement take the form of Eq. 8 for A_T^2 and A_T^{Im} and Eq. 9 for A_T^{Re}. They can be directly evaluated on data assuming a shape for F_L as in equation 7.

For a pure signal sample,

$$Corr(A_T^2) = Corr(A_T^{Im}) = \frac{\sum_{i=1}^{N}(1 - F_L(q_i^2))}{\sum_{i=1}^{N}(\frac{1-x_i}{1+\frac{x_i}{2}})(1 - F_L(q_i^2))} \tag{8}$$

$$Corr(A_T^{Re}) = \frac{\sum_{i=1}^{N}(1 - F_L(q_i^2))}{\sum_{i=1}^{N}(\frac{\sqrt{1-x_i}}{1+\frac{x_i}{2}})(1 - F_L(q_i^2))} \quad . \tag{9}$$

The result of the fit neglecting the threshold terms in the bin $0 < q^2 < 2\,\mathrm{GeV}^2/c^4$ has to be multiplied by these corrections to take into account the impact of the mass of the muon, as follows (similar relations hold for A_T^{Im} and A_T^{Re}):

$$A_T^2(0.1 - 2) = A_T^2(0.1 - 2)_{fromfit} \times Corr(A_T^2) \tag{10}$$

For the errors we multiply by the corrections on the errors (similar relations hold for A_T^{Im} and A_T^{Re}):

$$err(A_T^2(0.1 - 2)) = err(A_T^2(0.1 - 2))_{fromfit} \times Corr(err(A_T^2)) \tag{11}$$

It can be demonstrated that the corrections for S_3, A_{FB} and A_{Im} are the same as those for A_T^2, A_T^{Re} and A_T^{Im} respectively, since, according to section 8.7, the following relations hold:

$$< A_{Im} >= \frac{1}{2} < \tilde{A_T^{Im}} > (1- < F_L >) \tag{12}$$

$$< S_3 >= \frac{1}{2} < \tilde{A_T^{(2)}} > (1- < F_L >) \tag{13}$$

$$< A_{FB} >= \frac{3}{4} < \tilde{A_T^{Re}} > (1- < F_L >) \tag{14}$$

This correction procedure has been validated using the MC, as discussed in Appendix E.

17.3 Results of the evaluation of the corrections on data.

To evaluate the values of the corrections on data where we do not have a pure sample of signal events, Eq. 8 and 9 need to be modified introducing W_i as follows:

$$Corr(A_T^2) = Corr(A_T^{Im}) = \frac{\sum_{i=1}^{N}(1 - F_L(q_i^2))W_i}{\sum_{i=1}^{N}(\frac{1-x}{1+\frac{x_i}{2}})(1 - F_L(q_i^2))W_i} \tag{15}$$

$$Corr(A_T^{Re}) = \frac{\sum_{i=1}^{N}(1 - F_L(q_i^2))W_i}{\sum_{i=1}^{N}(\frac{\sqrt{1-x}}{1+\frac{x_i}{2}})(1 - F_L(q_i^2))W_i} \tag{16}$$

where W_i is a weight for the event i, which is the product of the weight taking into account the acceptance effects and a $s\mathcal{P}lot$ weight that comes from a fit to the $K^+\pi^-\mu^+\mu^-$ invariant mass distribution and is used to subtract the background.

The results are shown on table 34 for three possible values of the parameter a (in Eq. 7). The linear approximation of A_{FB} and A_T^{Re} is used to estimate the size of the correction for these observables.

To determine the parameter a on data, the mean value of F_L has been calculated using the following expression:

$$< F_L >= \frac{\sum_{i=1}^{N} F_L(q_i^2)W_i}{\sum_{i=1}^{N} W_i} = \frac{\sum_{i=1}^{N}(\frac{aq_i^2}{1+aq_i^2})W_i}{\sum_{i=1}^{N} W_i} \tag{17}$$

scanning the values of a between 0.2 and 1.3. The resulting curve is shown on Fig. 40, and the intersection with the measured value of $F_L = 0.36 \pm 0.10$ gives the measured value of a of $a = 0.67^{+0.54}_{-0.30}$.

	$a = 0.37$	$a = 67$	$a = 1.21$
Correction on A_T^2, S_3, A_T^{Im}, A_{Im}	1.18	1.20	1.22
Correction on $err(A_T^2)$, $err(S_3)$, $err(A_T^{Im})$, $err(A_{Im})$	1.16	1.18	1.20
Correction on A_T^{Re}, A_{FB}	1.12	1.13	1.14
Correction on A_T^{Re}, A_{FB} (linear approx)	1.06	1.06	1.07
Correction on $err(A_T^{Re})$, $err(A_{FB})$	1.11	1.12	1.14

Table 34: Values of the corrections evaluated with formulae 15 and 16 using 254 candidates in the range (0.1-2) GeV^2/c^4,assuming a behaviour for F_L as in Eq. 7. Three different values of the parameter a of $F_L(q^2)$, defined in Eq. 7, have been considered.

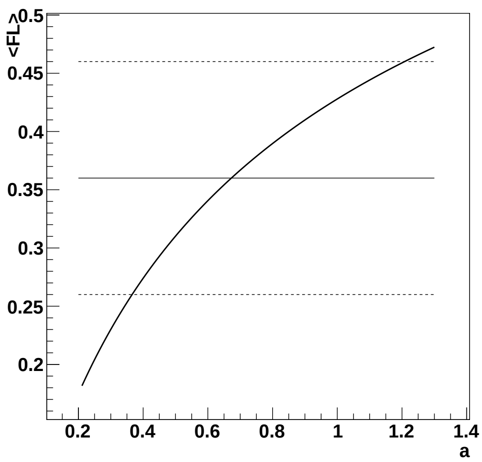


Figure 40: The curve represent the values of $< F_L >$ as function of a as calculated on data using Eq. 17 . The horizontal lines represent the measured value of F_L and its error. The intersection with the curve gives the measurement of $a = 0.67^{+0.54}_{-0.30}$.

As a cross-check we also computed the correction assuming a linear behaviour for F_L as function of q^2 (see E.2) , obtaining similar results.

18 Systematic uncertainties on and cross checks of the angular observables

Sources of systematic uncertainty are considered if they introduce an angular or q^2-dependent bias in the acceptance correction or can significantly change the estimated $B^0 \to K^{*0}\mu^+\mu^-$ signal yield. This includes data-MC corrections that vary with the momentum or p_{T} of the kaon, pion or muons.

Common sources of systematic uncertainty for all of the analyses presented in this note are:

- the statistical uncertainty on the acceptance correction coming from limited MC statistics;
- the uncertainty on the acceptance coming from the factorisation assumptions;
- the uncertainty on the acceptance coming from data-MC corrections;
- the uncertainty on the acceptance correction coming from differences in trigger efficiency between data and MC;
- the uncertainty on the line-shape of the $K^+\ \pi^-\ \mu^+\mu^-$ invariant mass.

For the differential branching fraction analysis, the contributions from:

- $B^0_s \to \phi\mu^+\mu^-$ with $K \to \pi$ mis-id;
- $B^0_s \to \overline{K}^{*0}\mu^+\mu^-$.

are explored. For the angular analysis and zero-crossing point extraction the impact of:

- $B^0 \leftrightarrow \overline{B}^0$ mis-id.

is considered. The letter in the subsection headings is a key that can be used when refering to the tables that appear later in this section and in Sec. 7.5.

18.1 Statistical uncertainty on the acceptance correction [A]

The statistical uncertainty on the factorised acceptance correction is small for most of the q^2 range. At high-q^2 it can become more significant due to limited MC statistics. In the $16 < q^2 < 19\,\mathrm{GeV}^2/c^4$ bin, where the uncertainty is largest, the statistical uncertainty on the acceptance corrections is 1-2%.

18.2 Acceptance correction binning [B]

One potential source of systematic bias is in the choice of q^2 binning for the acceptance correction - particularly in regions where the efficiency changes rapidly in q^2. To estimate the maximum possible size of this effect, the fit is repeated using the acceptance correction in ϕ, $\cos\theta_l$ and $\cos\theta_K$ for the neighbouring q^2 bins.

18.3 Systematic biases on the acceptance correction and the break down of factorisation [C]

To account for possible systematic biases in the acceptance correction, that are not accounted for else-where, an additional systematic uncertainty of 10% is applied to the acceptance correction. This is used as a "catch-all" for any effect in the acceptance correction that has not been fully understood in the studies in this note. To maximise any potential bias coming from this change in the acceptance correction this 10% variation is applied in a coherent way, e.g.

$$w_i \to w_i(1 \pm 0.1 \times |\cos\theta_{l;i}|)$$

or

$$w_i \to w_i(1 \pm 0.1 \times |\cos\theta_{K;i}|)$$

Variations are also tried in which $\cos\theta_l$ and $\cos\theta_K$ efficiencies are varied simultaneously. A non-factorisable variation of efficiency where:

$$w_i \to w_i(1 \pm 0.1 \times \sin(\pi.\cos\theta_{l;i})\sin(\pi.\cos\theta_{K;i})$$

is also considered.

No additional variation is applied to the ϕ angle as the ϕ-acceptance is thought to be a predominantly geometrical effect and is less effected by traditional data-MC differences.

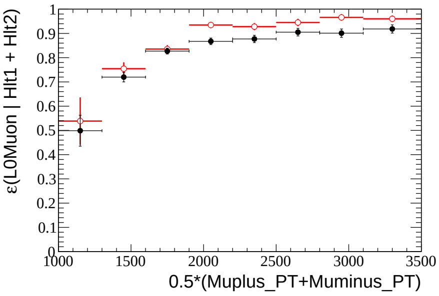
These 10% variations are conservative and could be relaxed if better agreement were to be achieved for $B^0 \to K^{*0} J/\psi$ decay or larger MC statistics were available.

18.4 Trigger efficiency [D]

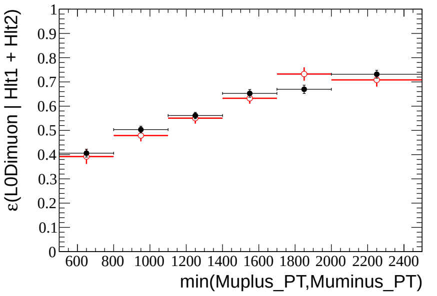
The trigger efficiency in data can estimated using the `Tis-Tos` technique on $B^0 \to K^{*0} J/\psi$ and compared to `MC11a` MC that has been selection with

Stripping 17 and Triggered with TCK 0x40760037. Fig. 41 shows the variation of the trigger efficiency in data and MC as a function of the kinematic properties of the muon system. For the L0Muon trigger the efficiency is compared as a function of the average p_{T} of the μ^+ and μ^-. Whilst there is a clear systematic difference seen in the efficiency, it appears to be independent of the muon p_{T} to $\mathcal{O}(1\%)$. A similar behaviour is exhibited by Hlt 1 and Hlt 2. At L0 and Hlt,1, the muon kinematics are the dominant contribution in determining the trigger efficiency.

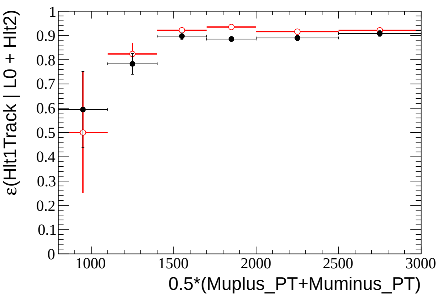
A similar study was completed in Ref. [14] for $B^0 \to K^{*0} J/\psi$ in MC10. In keeping with the previous analysis the effect of trigger is estimated by varying the efficiency of soft muons ($p \leq 10\,\mathrm{GeV}/c$) by 3% in the acceptance correction. Remaining differences will be caught by the vairation of the acceptance correction described above.



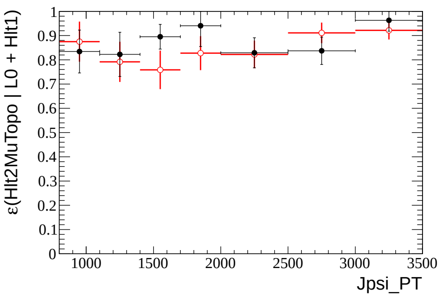
(a) L0Muon



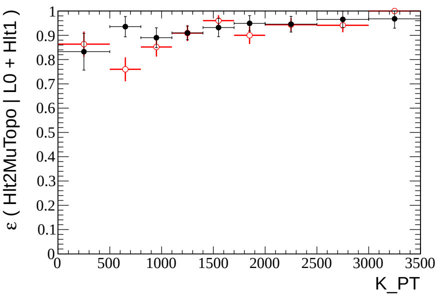
(b) L0Dimuon



(c) Hlt1Track



(d) Hlt2MuTopo



(e) Hlt2MuTopo

Figure 41: Trigger efficiency for $B^0 \to K^{*0}J/\psi$ candidates in data (solid marker) and truth-matched $B^0 \to K^{*0}J/\psi$ candidates in MC11a estimated using the Tis-Tos technique.

18.5 Data-MC corrections

18.5.1 `IsMuon` efficiency [E]

An estimate for the systematic associated with the IsMuon performance is made by fluctuating the efficiency of the two muons in the MC within the uncertainty on data-MC correction. For a conservative estimate, the efficiency of tracks with momentum $\leq 10\,\mathrm{GeV}/c$ is fluctuated downwards (upwards) and with momentum $> 10\,\mathrm{GeV}/c$ upwards (downwards) within their uncertainty. The uncertainty is typically $2-10\%$ and varies with momentum and η. The regions with the largest uncertainty are also the least polpulated by signal candidates in the data.

18.5.2 Tracking efficiency [F]

An estimate for the systematic associated with the tracking performance is estimated by fluctuating the efficiency for each of the four tracks in MC within the uncertainty on data-MC correction. For a conservative estimate, the efficiency of tracks with momentum $\leq 10\,\mathrm{GeV}/c$ is fluctuated downwards (upwards) and with momentum $> 10\,\mathrm{GeV}/c$ upwards (downwards) within their uncertainty. The uncertainty is typically $2-10\%$ and varies with momentum and η. Again, the regions with the largest uncertainty are also the least polpulated.

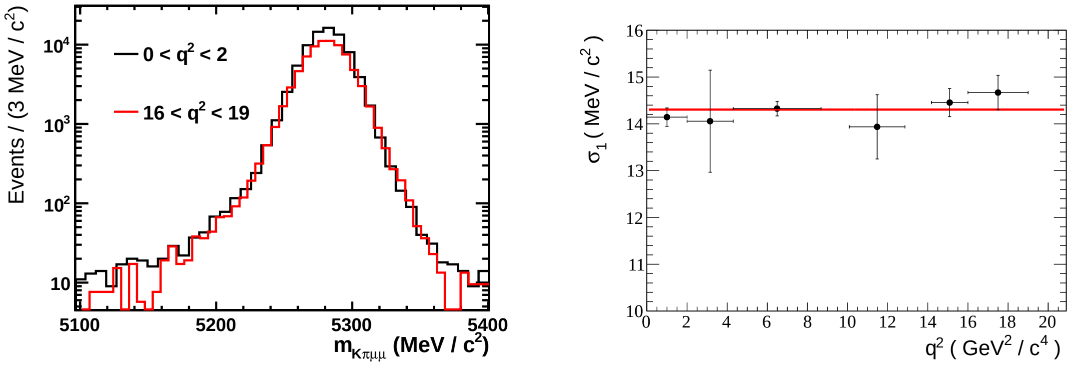
18.5.3 PID performance [G]

The PID distributions used for the MC are sampled from a D^{*+} calibration sample in bins of (p, η) and occupancy. There are two possible sources of uncertainty associated with this calibration sample: a statistical uncertainty associated with the number of $K^{\pm}/\pi^{\pm}$ candidates in each of the bins and a systematic uncertainty associated with the choice of binning.

A systematic uncertainty on the $\mathrm{DLL}_{K\pi}$ and $\mathrm{DLL}_{\mu\pi}$ corrections is estimated on the binning scheme, by assigning 50% of events within 10% of the bin width to the lower (higher) bin edge a DLL from the lower (higher) bin of the calibration sample.

18.5.4 IP smearing [H]

A conservative estimate of the systematic uncertainty on the IP smearing is made by producing an acceptance correction without IP smearing.



(a) (b)

Figure 42: The $K^+\pi^-\mu^+\mu^-$ invariant mass distribution of MC $B^0 \to K^{*0}\mu^+\mu^-$ candidates at high- and low-q^2 (a) and the q^2 dependence of the Gaussian width of the double Crystal Ball shapes used to model the invariant mass distribution (b).

18.5.5 BDT input variable re-weighting [I]

The variable B^0 p_T is re-weighted when applying the BDT to the sample of generated events used to defined the acceptance correction. The re-weighting of this variable was removed and a new acceptance correction produced. The same procedure was also performed for the variable B^0 p.

18.6 Signal mass model [J]

In the fits to the $K^+\pi^-\mu^+\mu^-$ invariant mass, the signal line-shape is assumed to be the same for the signal and control channel and to be independent of q^2. This has been cross checked for simulated $B^0 \to K^{*0}\mu^+\mu^-$ events in the q^2 bins used in this analysis. The Gaussian width of the double Crystal Ball shapes used to model the invariant mass distribution of these fits can be seen in Figure. 42 (b). A straight-line fit to this data yields a gradient of about 5%. This 5% is assigned as a systematic uncertainty by varying the width of the signal dsitribution by $\pm 5\%$ in the likelihood fits.

18.7 Background angular model [K]

In the angular fit, the background shape in each angle is modelled by a 2$^{\text{nd}}$ order polynomial. The systematic uncertainty associated with this choice of parameterisation is estimated by fitting using 0$^{\text{th}}$, 1$^{\text{st}}$ and 3$^{\text{rd}}$ order polynomials. Zeroth- and first-order background models are not expected to accurately describe background shape. Consider that a zeroth order polynomial is unable to model an asymmetric distribution of background events in $\cos\theta_l$. This will result in the mis-measurement of A_{FB} (A_{T}^{Re}). Similarly, a first order polynomial is unable to model any higher-order variations in the ϕ distribution of background events, to which the measurement of S_3, S_9 and A_9 (A_{T}^2, A_{T}^{Im}) are sensitive. The results of this study are in Appendix. H.

The sensitivity of the fit results to statistical fluctuations in the background is examined using pseudo-experiments. Problems could arise due to the small number of background candidates and the event weighting procedure. This could lead to events in an unlikely region of phase-space obtaining large weights and dramatically changing the background shape. To explore these effects, 10000 toy datasets are generated with the background flat in the angles (a 0^{th} order polynomial). These datasets are fitted with 1^{st} and 3^{rd} order polynomials, and compared to “nominal” fits performed using 2^{nd} order polynomials. The results of this study are summarised in Tables. 36-40 and Tables. 41-44. These biases are small.

18.8 $K^{*0} \leftrightarrow \overline{K}^{*0}$ mis-id [L]

The systematic bias coming from $K^{*0} \leftrightarrow \overline{K}^{*0}$ is negligible (below the 1% level) and will only impact A_{T}^{Re} and A_{T}^{Im} (A_{FB}, S_9 and A_9).

18.9 Peaking backgrounds [M]

The uncertainty on the peaking backgrounds from $B_s^0 \to \phi\mu^+\mu^-$ ($\pm 0.5\%$) and $B_s^0 \to K^{*0}\mu^+\mu^-$ are considered for the differential branching fraction. $B_s^0 \to K^{*0}\mu^+\mu^-$ has not yet been seen. For the analysis it is assumed that the ratio of this decay mode to $B^0 \to K^{*0}\mu^+\mu^-$ is a simple ratio of the CKM elements and f_s/f_d, i.e. it is approximately 1%. An uncertainty of $\pm 1\%$ is assumed on this number.

Peaking backgrounds are not accounted for directly in the angular fits. It is difficult to satisfactorily account for this contribution due to the unknown angular distribution of $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$. Instead a conservative estimate is assumed in which these backgrounds have the same shape as the signal angular distributions, and maximal or minimal values of the physics parameters (e.g. $A_{FB} = \pm 1$ and $F_L = 0$, 1). This leads to a systematic uncertainty at the level of 2% for $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$. These variations are not included in the tables below.

18.10 Multiple candidates [N]

The fits for the angular observables have been repeated removing all events that contain multiple candidates (1%). This has a negligible impact on the final result (this variation is not shown in tables below).

18.11 Removal of soft-tracks [O]

The fits to the angular observables have also been repeated by removing events with tracks with momenta less than $5\,\mathrm{GeV}/c$ (and recomputing the acceptance correction). This variation is prompted by Fig. 57 in Appendix A. The number of $B^0 \to K^{*0}\mu^+\mu^-$ candidates removed by this requirement is small in the data. These candidates tend to sit at the extremes of $\cos\theta_K$ and typically have large weights. The effect of removing these candidates is indicated in Tables. 41-44.

18.12 Uncertainty on the S-wave component [P]

The fits are performed assuming the absence of an S-wave component. The systematic uncertainty introduced by this assumption was estimated by incorporating an s-wave into the pdf with the properties extracted in Sec. 16. This corresponds to the parameters $A_S = -0.11$ and $F_S = 0.07$.

18.13 Estimation of the systematic uncertainty on the angular observables

Systematic uncertainties on the angular observables have been estimated in two ways:

1. In an ad-hoc way, by systematically varying the acceptance correction and repeating the fit to the data with weights from this new acceptance correction;

2. Using toy pseudo-experiments.

Results from the first approach are included in Appendix. H. The second approach is described below.

In the toy approach, the typical size of the systematic bias is estimated by generating toys with the nominal acceptance effect and the signal and background parameters fixed to their best fit values to the data. In the FC toys, each candidate is the weighted by the same acceptance function that is used to accept-reject events. Here, the toys are instead weighted according to the acceptance effect after the systematic effect of interest has been varied; i.e. the acceptance used to weight the toys is not the same as the one that has been used to accept-reject them.

Ten thousand toy datasets were generated for each systematic variation described above, with the measured central values in Tables 21 and 22. The standard angular fit was then performed on each generated dataset to obtain the distribution of fitted values for each angular observable and each systematic variation.

The size of the systematic uncertainty on each physics parameter is calculated as the difference between the value of the physic parameter used to generate the toys and the mean value of the parameter from the angular fits to the toys. The standard error on the mean is used as a measure of the statistical uncertainty arising from the limited number of generated datasets, for each observable in each q^2 bin.

This procedure is not used to estimate the systematic uncertainty related to peaking backgrounds, which is described in section 18.9, or that related to multiple candidates, which is described in section 18.10.

19 Calculating the overall systematic contribution

The combined systematic uncertainty on each observable is then calculated from:

- the largest of the $\cos\theta_l$ [up,down], $\cos\theta_K$ [up,down], and non-factorisable $\cos\theta_l \cos\theta_K$ [up,down] variations;
- the systematic variation of the muon identification efficiency;
- the systematic variation of the tracking efficiency;
- the systematic variation of the trigger efficiency;
- the systematic variation between the IP smeared and the non-IP smeared simulated events;
- the systematic variation of the signal mass resolution;
- the systematic variation of the PID, by varying the PID binning;
- the systematic variation achieved when using the neighbouring q^2 bin for the acceptance;
- the introduction of a 7% S-wave;
- the possible bias from peaking backgrounds.

These contributions were added in quadrature ignoring correlations.

For completeness, the variations that do not represent reasonable changes in the analysis procedure and instead constitute cross checks are listed below:

- Cut on the hadron momentum;
- Tightening of the peaking background vetoes;
- Reweighting (or not) the B momentum and the B p_{T}.
- Removal of events containing multiple candidates.
- Variation of the background angular fit to 0^{th}, 1^{st} or 3^{rd} order (see Appendix H).

These variations do not have any significant impact on the final result.

The values in the tables of systematic uncertainties, shown in section 19.0.1, are calculated in the following way. Toy datasets are produced by generating events and performing an accept-reject procedure to replicate the acceptance effect. The systematic studies are performed by re-weighting the events according to a systematically varied acceptance correction and performing the angular fit. The results of these fits are compared to the "nominal" fit result, when using the same acceptance correction that is used to accept-reject the events.

Ten thousand datasets are generated using the same acceptance correction that is used to accept-reject the events. These datasets are fitted, obtaining a distribution of fitted values for each observable. The mean of these distributions are shown in the first row of the tables, the row labelled "nominal". Ten thousand datasets are then generated for each systematic variation, now using a systematically varied acceptance correction. The same fit is then performed on each of these datasets to obtain a systematically varied distribution of fitted values. The mean of each systematically varied distribution is extracted. The difference of the two means is then the systematic uncertainty that corresponds to each systematic variation, and is shown in the tables.

The standard error on each 'nominal' value is also calculated. If the standard error is larger than a given systematic uncertainty obtained from the above procedure, then the standard error is taken as that systematic uncertainty.

19.0.1 Glossary of contributions

Systematic	Description
CTK[...]	Variation of the acceptance in $\cos\theta_K$ by $\pm 10\%$
CTL[...]	Variation of the acceptance in $\cos\theta_l$ by $\pm 10\%$
CTL,CTK[...]	Variation of the acceptance in $\cos\theta_l$ and $\cos\theta_K$ by $\pm 10\%$ in a factorisable way
Nonfac	Variation of the acceptance in $\cos\theta_l$ and $\cos\theta_K$ by 10% in a non-factorisable way
HadronP	Removal of tracks with $p < 5\,\text{GeV}/c$ in the analysis and acceptance correction
Muon ID	Variation of `IsMuon` relative efficiency within its measured uncertainty
KPI[...]	Variation of the $K^{*0} \to \overline{K}^{*0}$ mis-ID
NoMultCan	Removal of events containing multiple candidates
Sigma[...]	Variation of the mass resolution of $B^0 \to K^{*0}\mu^+\mu^-$ with respect to $B^0 \to K^{*0}J/\psi$ by 5%
Tracking	Variation of the relative tracking efficiency between data and MC within in the measured uncertainty
Trigger	Simulation of trigger efficiency with a 3% efficiency loss for soft-muons
S-wave	See text
PID Correction [Up,Down]	Assignment of DLL value for neighbouring bin for events close to bin edge
Bin [Up, Down]	Change q^2 binning scheme take acceptance for bin above or bin below.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
fileNominal	-0.020 +/- 0.001	-0.185 +/- 0.001	0.171 +/- 0.000	0.277 +/- 0.001	0.486 +/- 0.001	0.291 +/- 0.001	-0.151 +/- 0.001
AC CTL Up [C]	-0.002	-0.003	0.002	0.004	0.006	0.002	-0.002
AC CTL Down [C]	0.001	0.004	-0.002	-0.004	-0.005	-0.004	0.004
AC CTK Up [C]	0.001	0.006	-0.005	-0.007	-0.010	-0.006	0.006
AC CTK Down [C]	0.000	-0.007	0.005	0.008	0.012	0.007	-0.005
AC CTL Up CTK Down [C]	0.001	0.003	-0.004	-0.003	-0.004	-0.004	0.002
AC CTL Down CTK Up [C]	-0.000	0.010	-0.008	-0.011	-0.013	-0.012	0.010
AC CTL Up CTK Down [C]	-0.003	-0.010	0.009	0.014	0.016	0.011	-0.008
AC CTL Down CTK Down [C]	-0.002	-0.002	0.003	0.004	0.006	0.002	-0.002
AC Non-factorisable Up [C]	0.001	0.001	-0.000	0.001	0.000	-0.000	0.002
AC Non-factorisable Down [C]	0.002	-0.000	0.001	-0.000	-0.000	0.001	0.000
AC q^2 binning +1 [B]	0.001	-0.002	0.002	0.002	-0.003	-0.007	-0.001
AC q^2 binning -1 [B]	-0.001	0.004	-0.002	0.005	0.007	0.005	0.001
1st Order Background Model [K]	-0.001	0.002	0.000	0.001	0.002	-0.000	0.000
3rd Order Background Model [K]	-0.000	-0.001	-0.000	0.000	-0.001	0.000	0.002
K^{*0} mis-ID Down [L]	-0.000	0.002	-0.001	-0.000	0.000	0.000	0.001
K^{*0} mis-ID Up [L]	-0.001	-0.001	0.001	0.001	0.002	0.000	0.000
Signal mass width Down [J]	-0.002	-0.000	0.001	0.002	-0.000	-0.000	0.001
Signal mass width Up [J]	-0.001	0.001	0.001	0.001	0.001	0.000	0.001
S-wave component [P]	-0.000	0.007	-0.008	-0.011	-0.014	-0.012	0.007
IP Smearing [H]	0.002	0.002	-0.001	0.001	0.005	0.001	-0.002
B^0 p_T re-weighting [I]	-0.001	-0.000	0.002	0.003	0.002	0.001	-0.001
B^0 p re-weighting [I]	-0.000	-0.002	0.002	0.006	0.005	0.003	-0.001
IsMuon efficiency Up [E]	0.001	0.002	-0.001	0.000	0.001	0.000	0.002
IsMuon efficiency Down [E]	-0.001	-0.006	0.002	0.001	0.002	-0.001	-0.002
Removal of soft tracks [O]	0.001	-0.003	0.003	0.007	0.007	0.004	-0.002
PID performance +5% [G]	0.001	0.001	-0.000	0.001	0.000	-0.003	0.001
PID performance -5% [G]	-0.001	0.000	0.000	-0.002	0.000	-0.003	0.002
PID performance +10% [G]	-0.001	0.002	-0.001	-0.001	0.000	-0.002	0.003
PID performance -10% [G]	-0.000	-0.002	-0.001	0.001	-0.003	-0.002	0.002
PID performance +30% [G]	-0.001	-0.000	-0.001	0.000	0.001	-0.000	0.004
PID performance -30% [G]	-0.000	-0.001	-0.000	0.001	-0.003	-0.003	0.003
Tracking efficiency Up [F]	0.002	-0.000	0.001	-0.000	0.001	0.001	0.001
Tracking efficiency Down [F]	0.000	-0.001	0.001	0.000	-0.000	-0.000	0.001
Trigger efficiency Up [D]	0.001	-0.002	0.001	0.003	0.004	-0.000	-0.001
Trigger efficiency Down [D]	0.000	0.000	0.002	0.002	0.002	0.001	0.000

Table 36: Variation of $A_{\rm FB}$ when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
fileNominal	0.362 +/- 0.001	0.669 +/- 0.001	0.551 +/- 0.001	0.462 +/- 0.001	0.330 +/- 0.001	0.374 +/- 0.001	0.633 +/- 0.001
AC CTL Up [C]	-0.010	-0.008	-0.008	-0.007	-0.010	-0.007	-0.006
AC CTL Down [C]	0.012	0.004	0.010	0.008	0.007	0.008	0.012
AC CTK Up [C]	0.017	0.010	0.018	0.018	0.014	0.015	0.018
AC CTK Down [C]	-0.019	-0.014	-0.021	-0.017	-0.018	-0.019	-0.016
AC CTL Up CTK Down [C]	0.008	0.005	0.012	0.011	0.006	0.010	0.009
AC CTL Down CTK Up [C]	0.028	0.017	0.027	0.025	0.019	0.025	0.027
AC CTL Up CTK Down [C]	-0.030	-0.022	-0.029	-0.028	-0.025	-0.027	-0.025
AC CTL Down CTK Down [C]	-0.011	-0.008	-0.010	-0.010	-0.010	-0.008	-0.007
AC Non-factorisable Up [C]	-0.001	-0.001	-0.000	0.000	-0.001	0.000	0.002
AC Non-factorisable Down [C]	0.001	-0.001	-0.001	0.000	-0.000	-0.002	0.001
AC q^2 binning +1 [B]	0.004	-0.003	-0.007	-0.001	0.004	0.014	-0.001
AC q^2 binning -1 [B]	-0.005	0.004	0.008	-0.012	-0.013	-0.011	0.007
1st Order Background Model [K]	0.001	0.005	-0.001	0.002	-0.002	-0.001	0.002
3rd Order Background Model [K]	-0.001	-0.005	-0.000	-0.001	-0.000	-0.002	0.001
K^{*0} mis-ID Down [L]	-0.000	-0.000	-0.000	-0.001	-0.003	-0.001	0.002
K^{*0} mis-ID Up [L]	-0.001	-0.001	0.000	-0.000	-0.002	0.000	0.001
Signal mass width Down [J]	0.001	-0.003	-0.000	-0.003	0.000	-0.000	0.001
Signal mass width Up [J]	-0.000	-0.002	-0.001	-0.001	-0.002	0.000	0.001
S-wave component [P]	0.004	-0.009	-0.008	-0.001	0.003	0.003	-0.009
IP Smearing [H]	0.017	-0.000	0.003	0.000	-0.007	-0.007	-0.007
B^0 p_T re-weighting [I]	-0.002	-0.003	-0.004	-0.004	-0.004	-0.002	-0.002
B^0 p re-weighting [I]	0.003	-0.005	-0.010	-0.011	-0.008	-0.005	-0.004
IsMuon efficiency Up [E]	0.007	0.002	0.002	0.000	-0.001	-0.002	0.007
IsMuon efficiency Down [E]	-0.013	-0.011	-0.005	-0.000	-0.002	0.002	-0.005
Removal of soft tracks [O]	0.004	-0.006	-0.012	-0.015	-0.012	-0.015	-0.003
PID performance +5% [G]	0.005	0.001	0.003	-0.000	0.001	0.007	0.006
PID performance -5% [G]	0.003	-0.000	0.002	0.005	-0.001	0.006	0.007
PID performance +10% [G]	0.011	-0.000	0.004	0.001	-0.001	0.005	0.010
PID performance -10% [G]	-0.004	-0.006	0.002	-0.000	0.004	0.004	0.006
PID performance +30% [G]	0.011	-0.001	0.004	0.002	-0.002	0.003	0.010
PID performance -30% [G]	-0.002	-0.004	0.002	0.000	0.004	0.004	0.006
Tracking efficiency Up [F]	-0.001	-0.002	-0.002	-0.001	-0.002	-0.001	0.002
Tracking efficiency Down [F]	-0.001	-0.002	0.001	0.002	-0.000	-0.000	0.002
Trigger efficiency Up [D]	-0.002	-0.003	-0.005	-0.005	-0.006	-0.003	-0.003
Trigger efficiency Down [D]	0.003	-0.002	-0.005	-0.006	-0.004	-0.004	-0.000

Table 37: Variation of F_{L} when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
fileNominal	-0.050 +/- 0.001	-0.027 +/- 0.001	0.072 +/- 0.001	-0.142 +/- 0.001	0.029 +/- 0.001	-0.203 +/- 0.001	0.023 +/- 0.001
AC CTL Up [C]	-0.002	-0.001	0.001	-0.001	-0.003	-0.002	0.003
AC CTL Down [C]	-0.001	-0.002	0.000	0.000	-0.001	-0.001	0.001
AC CTK Up [C]	0.000	-0.000	-0.001	0.003	-0.004	0.003	0.001
AC CTK Down [C]	-0.002	0.000	0.002	-0.003	-0.002	-0.005	0.003
AC CTL Up CTK Down [C]	0.000	-0.000	-0.001	0.001	-0.002	-0.001	-0.000
AC CTL Down CTK Up [C]	0.000	0.000	-0.002	0.004	-0.003	0.004	0.002
AC CTL Up CTK Down [C]	-0.002	-0.002	0.003	-0.005	-0.000	-0.005	0.002
AC CTL Down CTK Down [C]	-0.001	-0.002	0.001	-0.004	-0.001	-0.005	0.003
AC Non-factorisable Up [C]	-0.001	-0.003	0.002	-0.001	-0.001	-0.002	0.002
AC Non-factorisable Down [C]	-0.000	-0.002	0.001	-0.001	-0.000	-0.001	0.000
AC q^2 binning +1 [B]	-0.002	0.000	-0.001	-0.001	0.004	0.002	0.001
AC q^2 binning -1 [B]	0.004	-0.002	-0.003	-0.008	0.006	0.000	0.001
1st Order Background Model [K]	-0.001	0.000	0.000	-0.001	-0.003	-0.001	0.002
3rd Order Background Model [K]	-0.001	-0.001	0.001	-0.001	-0.002	-0.001	0.001
K^{*0} mis-ID Down [L]	0.000	-0.000	0.001	-0.001	-0.001	-0.001	0.001
K^{*0} mis-ID Up [L]	-0.000	-0.001	0.001	-0.002	-0.001	-0.001	0.002
Signal mass width Down [J]	-0.000	-0.001	0.001	-0.002	-0.002	-0.001	0.003
Signal mass width Up [J]	-0.001	-0.002	0.001	-0.001	-0.002	0.000	0.002
S-wave component [P]	0.005	0.002	-0.006	0.011	-0.005	0.016	-0.001
IP Smearing [H]	0.003	-0.002	0.000	0.003	0.003	-0.010	0.005
B^0 p_T re-weighting [I]	0.001	0.000	0.001	-0.003	-0.000	-0.000	0.002
B^0 p re-weighting [I]	-0.002	0.000	0.004	-0.002	-0.001	0.001	0.002
IsMuon efficiency Up [E]	0.000	-0.000	0.001	-0.001	-0.001	-0.000	0.002
IsMuon efficiency Down [E]	-0.003	-0.002	0.001	-0.001	-0.004	-0.001	-0.000
Removal of soft tracks [O]	-0.003	0.000	0.001	-0.003	-0.001	-0.006	-0.002
PID performance +5% [G]	0.001	-0.000	-0.002	-0.001	-0.004	0.000	0.001
PID performance -5% [G]	0.000	-0.000	-0.001	0.000	-0.001	0.001	0.001
PID performance +10% [G]	0.004	0.001	-0.002	-0.001	-0.003	0.002	-0.001
PID performance -10% [G]	0.000	0.001	-0.001	-0.000	-0.006	-0.003	0.002
PID performance +30% [G]	0.005	0.000	-0.002	0.000	-0.004	0.002	0.001
PID performance -30% [G]	-0.001	0.000	-0.002	-0.001	-0.006	-0.003	0.001
Tracking efficiency Up [F]	-0.001	-0.001	0.001	-0.001	-0.002	-0.001	0.002
Tracking efficiency Down [F]	-0.001	-0.001	0.000	0.000	-0.001	-0.001	0.001
Trigger efficiency Up [D]	-0.003	0.001	0.003	0.000	-0.002	-0.002	0.000
Trigger efficiency Down [D]	-0.003	-0.001	0.002	-0.001	-0.001	-0.000	0.002

Table 38: Variation of S_3 when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
fileNominal	0.056 +/- 0.001	-0.018 +/- 0.001	0.010 +/- 0.001	-0.019 +/- 0.001	0.002 +/- 0.001	0.057 +/- 0.001	0.065 +/- 0.001
AC CTL Up [C]	0.002	0.000	-0.000	-0.000	0.000	0.000	0.000
AC CTL Down [C]	0.000	0.001	-0.000	-0.000	-0.000	0.002	-0.002
AC CTK Up [C]	-0.003	0.001	0.000	-0.000	0.000	-0.001	-0.002
AC CTK Down [C]	0.000	0.000	0.001	-0.002	0.002	0.001	0.001
AC CTL Up CTK Down [C]	-0.002	0.001	-0.001	0.000	-0.000	-0.002	-0.002
AC CTL Down CTK Up [C]	-0.002	-0.000	-0.001	0.001	0.001	-0.002	-0.003
AC CTL Up CTK Down [C]	0.000	0.000	-0.001	-0.000	-0.001	0.001	0.003
AC CTL Down CTK Down [C]	0.000	0.002	-0.000	-0.001	-0.000	0.001	0.003
AC Non-factorisable Up [C]	0.001	-0.001	0.000	0.000	-0.001	-0.000	-0.001
AC Non-factorisable Down [C]	-0.002	0.000	0.000	-0.000	-0.002	-0.000	0.000
AC q^2 binning +1 [B]	-0.001	0.000	0.000	-0.000	-0.001	-0.002	-0.000
AC q^2 binning -1 [B]	0.002	0.001	-0.001	0.001	-0.002	-0.000	-0.000
1st Order Background Model [K]	-0.001	0.001	-0.001	0.000	0.001	-0.000	-0.001
3rd Order Background Model [K]	-0.001	0.001	-0.001	-0.001	-0.001	-0.002	-0.000
K^{*0} mis-ID Down [L]	-0.000	-0.000	-0.000	-0.002	-0.001	-0.001	-0.000
K^{*0} mis-ID Up [L]	-0.001	0.000	-0.001	-0.001	-0.002	0.001	-0.001
Signal mass width Down [J]	0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Signal mass width Up [J]	-0.000	0.000	-0.001	-0.000	0.001	-0.001	-0.002
S-wave component [P]	-0.006	-0.001	-0.002	0.000	-0.001	-0.005	-0.006
IP Smearing [H]	-0.001	0.001	-0.001	-0.000	-0.001	0.000	0.000
B^0 p_T re-weighting [I]	-0.000	-0.001	-0.000	0.001	-0.000	-0.002	-0.001
B^0 p re-weighting [I]	-0.001	-0.001	-0.001	0.000	-0.000	0.001	0.002
IsMuon efficiency Up [E]	-0.000	0.001	0.001	0.001	-0.000	0.001	-0.000
IsMuon efficiency Down [E]	-0.000	-0.000	-0.001	-0.001	-0.001	-0.001	0.000
Removal of soft tracks [O]	-0.001	0.000	0.000	-0.001	-0.001	-0.000	-0.001
PID performance +5% [G]	-0.001	0.002	-0.000	0.000	-0.002	-0.001	-0.002
PID performance -5% [G]	-0.001	0.000	0.001	0.000	-0.000	-0.001	-0.002
PID performance +10% [G]	-0.000	-0.001	-0.001	-0.001	0.001	-0.001	-0.000
PID performance -10% [G]	0.000	-0.002	0.000	-0.000	-0.000	-0.002	0.001
PID performance +30% [G]	-0.000	0.001	-0.000	-0.001	-0.001	-0.000	-0.001
PID performance -30% [G]	0.000	-0.001	0.000	0.001	0.000	-0.002	-0.001
Tracking efficiency Up [F]	-0.001	-0.000	-0.001	0.000	-0.001	0.000	-0.001
Tracking efficiency Down [F]	-0.002	-0.000	-0.001	-0.000	-0.000	-0.001	-0.001
Trigger efficiency Up [D]	-0.001	-0.000	-0.001	0.000	-0.000	0.000	0.001
Trigger efficiency Down [D]	-0.000	-0.000	-0.001	0.002	0.000	0.001	0.000

Table 39: Variation of S_9 when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
fileNominal	0.119 +/- 0.001	0.048 +/- 0.001	-0.133 +/- 0.001	0.002 +/- 0.001	-0.061 +/- 0.001	0.001 +/- 0.001	0.021 +/- 0.001
AC CTL Up [C]	0.002	-0.001	-0.001	-0.002	-0.001	-0.000	0.001
AC CTL Down [C]	0.001	-0.001	0.000	-0.001	0.000	-0.000	-0.001
AC CTK Up [C]	-0.001	-0.001	0.003	-0.001	0.001	0.002	0.001
AC CTK Down [C]	0.004	0.003	-0.005	-0.000	0.000	0.002	0.002
AC CTL Up CTK Down [C]	0.000	-0.000	0.003	0.000	0.000	-0.001	-0.000
AC CTL Down CTK Up [C]	0.000	-0.002	0.005	-0.000	0.001	0.000	-0.000
AC CTL Up CTK Down [C]	0.003	0.002	-0.005	-0.001	-0.002	-0.001	0.002
AC CTL Down CTK Down [C]	0.004	0.002	-0.004	0.001	-0.000	0.001	0.001
AC Non-factorisable Up [C]	0.002	0.000	-0.001	0.001	0.000	0.003	0.000
AC Non-factorisable Down [C]	0.001	0.001	-0.001	-0.001	0.000	0.002	0.001
AC q^2 binning +1 [B]	-0.002	-0.001	-0.003	-0.000	0.001	0.001	0.001
AC q^2 binning -1 [B]	0.003	0.000	0.001	0.000	0.000	0.001	0.001
1st Order Background Model [K]	0.003	-0.001	-0.000	-0.000	0.000	0.001	-0.001
3rd Order Background Model [K]	0.002	-0.001	0.000	-0.001	0.002	-0.001	0.001
K^{*0} mis-ID Down [L]	0.001	-0.001	-0.000	-0.001	0.001	0.000	0.001
K^{*0} mis-ID Up [L]	0.001	-0.000	-0.001	0.000	0.001	0.001	0.000
Signal mass width Down [J]	0.001	0.002	-0.001	0.003	-0.000	0.003	0.001
Signal mass width Up [J]	0.001	0.000	-0.001	-0.000	0.001	0.000	0.001
S-wave component [P]	-0.008	-0.001	0.013	-0.001	0.007	0.002	-0.003
IP Smearing [H]	-0.001	0.001	-0.001	-0.000	-0.001	0.001	0.000
B^0 p_T re-weighting [I]	0.000	0.000	-0.002	-0.001	0.000	0.002	0.000
B^0 p re-weighting [I]	0.003	0.001	-0.003	-0.002	-0.001	-0.000	-0.000
IsMuon efficiency Up [E]	0.002	-0.001	0.001	-0.001	-0.001	0.000	0.000
IsMuon efficiency Down [E]	0.002	0.002	-0.001	-0.002	0.001	0.001	0.002
Removal of soft tracks [O]	0.003	-0.000	-0.002	0.000	-0.000	0.002	0.000
PID performance +5% [G]	0.001	0.000	0.000	-0.001	0.002	0.001	0.000
PID performance -5% [G]	-0.000	-0.001	0.001	-0.001	0.001	0.001	0.001
PID performance +10% [G]	0.002	-0.000	0.000	-0.002	-0.001	0.002	0.001
PID performance -10% [G]	0.003	0.000	0.001	-0.002	0.001	0.001	-0.000
PID performance +30% [G]	0.001	0.001	-0.000	0.000	-0.000	0.002	0.000
PID performance -30% [G]	0.001	0.001	0.001	0.001	0.001	0.000	0.000
Tracking efficiency Up [F]	0.000	0.002	-0.001	0.000	0.001	0.001	0.001
Tracking efficiency Down [F]	0.001	-0.000	-0.000	-0.001	0.001	0.002	0.000
Trigger efficiency Up [D]	0.003	-0.000	-0.001	-0.002	-0.001	0.000	0.001
Trigger efficiency Down [D]	0.004	0.001	-0.001	-0.000	-0.002	-0.000	0.002

Table 40: Variation of A_9 when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
fileNominal	-0.045 +/- 0.002	-0.745 +/- 0.003	0.511 +/- 0.001	0.697 +/- 0.001	0.960 +/- 0.001	0.626 +/- 0.001	-0.556 +/- 0.002
AC CTL Up [C]	0.005	-0.002	0.002	-0.000	0.001	0.002	-0.001
AC CTL Down [C]	0.005	0.002	0.001	-0.001	0.001	0.000	-0.002
AC CTK Up [C]	0.008	0.004	0.005	0.004	0.003	0.004	-0.006
AC CTK Down [C]	0.007	0.003	-0.005	-0.004	-0.001	-0.002	0.003
AC CTL Up CTK Down [C]	0.008	-0.003	0.006	0.003	0.003	0.005	-0.006
AC CTL Down CTK Up [C]	-0.000	-0.001	0.006	0.003	0.001	0.007	-0.006
AC CTL Up CTK Down [C]	0.004	-0.002	-0.005	-0.005	-0.001	-0.004	0.009
AC CTL Down CTK Down [C]	0.003	0.002	-0.004	-0.004	-0.001	-0.001	0.004
AC Non-factorisable Up [C]	0.001	0.007	0.002	0.001	0.001	0.003	-0.005
AC Non-factorisable Down [C]	0.000	-0.001	0.001	-0.002	0.000	-0.001	-0.001
AC q^2 binning +1 [B]	0.006	-0.002	-0.001	0.003	0.001	0.005	-0.003
AC q^2 binning -1 [B]	0.003	-0.001	0.003	-0.005	0.000	0.002	-0.001
1st Order Background Model [K]	0.010	0.002	0.001	0.001	0.002	-0.001	-0.001
3rd Order Background Model [K]	0.003	0.002	0.003	-0.001	-0.001	-0.000	-0.003
K^{*0} mis-ID Down [L]	0.003	0.001	-0.001	-0.002	-0.001	-0.003	-0.000
K^{*0} mis-ID Up [L]	0.003	-0.003	0.003	0.002	0.003	0.004	-0.000
Signal mass width Down [J]	0.001	-0.003	-0.000	0.001	0.001	0.000	-0.001
Signal mass width Up [J]	-0.001	-0.001	-0.000	0.000	0.001	0.002	-0.001
S-wave component [P]	0.005	0.039	-0.032	-0.032	-0.021	-0.022	0.042
IP Smearing [H]	0.004	0.004	0.001	0.002	0.001	-0.003	0.001
B^0 p_T re-weighting [I]	0.006	-0.003	0.000	-0.002	0.000	0.002	-0.001
B^0 p re-weighting [I]	0.007	-0.001	0.002	0.000	0.001	-0.001	-0.002
IsMuon efficiency Up [E]	0.000	0.001	-0.001	0.000	0.002	0.002	-0.001
IsMuon efficiency Down [E]	0.007	-0.005	-0.001	-0.000	0.001	0.001	-0.001
Removal of soft tracks [O]	0.003	-0.000	-0.005	-0.001	0.000	-0.001	-0.001
PID performance +5% [G]	0.001	0.005	0.004	0.003	0.001	0.002	-0.003
PID performance -5% [G]	0.003	-0.005	0.001	0.000	0.002	0.003	-0.005
PID performance +10% [G]	-0.000	0.000	0.004	0.002	0.001	0.004	-0.001
PID performance -10% [G]	0.004	-0.001	0.002	0.002	0.002	0.001	-0.003
PID performance +30% [G]	0.004	0.001	-0.000	0.000	0.000	0.003	-0.006
PID performance -30% [G]	0.001	-0.001	0.001	0.001	0.001	0.002	-0.005
Tracking efficiency Up [F]	0.003	-0.001	-0.002	-0.001	0.002	0.001	-0.000
Tracking efficiency Down [F]	0.001	-0.003	0.001	-0.002	0.003	0.001	0.000
Trigger efficiency Up [D]	0.002	-0.005	-0.000	-0.003	0.001	0.000	0.002
Trigger efficiency Down [D]	0.002	-0.001	-0.001	-0.002	0.001	0.002	-0.002

Table 41: Variation of A_{T}^{Re} when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
fileNominal	0.361 +/- 0.001	0.703 +/- 0.001	0.552 +/- 0.001	0.466 +/- 0.001	0.312 +/- 0.001	0.374 +/- 0.001	0.640 +/- 0.001
AC CTL Up [C]	-0.008	-0.005	-0.008	-0.007	-0.007	-0.005	-0.009
AC CTL Down [C]	0.012	0.008	0.010	0.010	0.007	0.012	0.010
AC CTK Up [C]	0.018	0.015	0.017	0.017	0.015	0.017	0.016
AC CTK Down [C]	-0.019	-0.016	-0.020	-0.020	-0.018	-0.016	-0.020
AC CTL Up CTK Down [C]	0.009	0.009	0.011	0.011	0.008	0.013	0.007
AC CTL Down CTK Up [C]	0.028	0.023	0.028	0.026	0.023	0.028	0.026
AC CTL Up CTK Down [C]	-0.027	-0.024	-0.028	-0.029	-0.025	-0.027	-0.027
AC CTL Down CTK Down [C]	-0.008	-0.007	-0.010	-0.010	-0.007	-0.009	-0.009
AC Non-factorisable Up [C]	0.002	0.001	-0.001	-0.000	-0.001	0.000	-0.000
AC Non-factorisable Down [C]	0.001	0.001	-0.001	-0.001	0.000	0.001	-0.000
AC q^2 binning +1 [B]	0.006	-0.002	-0.007	-0.001	0.004	0.016	-0.002
AC q^2 binning -1 [B]	-0.005	0.007	0.008	-0.011	-0.012	-0.010	0.005
1st Order Background Model [K]	-0.001	0.004	-0.001	-0.001	0.002	0.001	-0.000
3rd Order Background Model [K]	0.002	-0.001	-0.000	-0.001	-0.001	-0.001	-0.000
K^{*0} mis-ID Down [L]	-0.000	0.002	-0.000	-0.000	0.001	0.001	-0.001
K^{*0} mis-ID Up [L]	0.001	0.001	0.000	-0.001	-0.001	0.001	0.000
Signal mass width Down [J]	-0.001	0.001	0.000	-0.001	0.000	0.001	-0.000
Signal mass width Up [J]	-0.000	0.000	-0.001	-0.001	-0.000	0.000	-0.001
S-wave component [P]	0.004	-0.013	-0.009	-0.001	0.011	0.005	-0.014
IP Smearing [H]	0.019	-0.001	0.003	0.002	-0.006	-0.004	-0.009
B^0 p_T re-weighting [I]	-0.002	-0.002	-0.005	-0.005	-0.002	-0.001	-0.002
B^0 p re-weighting [I]	0.002	-0.004	-0.009	-0.012	-0.007	-0.004	-0.006
IsMuon efficiency Up [E]	0.008	0.008	0.002	-0.000	0.000	0.001	0.006
IsMuon efficiency Down [E]	-0.012	-0.010	-0.004	0.000	0.001	0.004	-0.008
Removal of soft tracks [O]	0.005	-0.003	-0.012	-0.015	-0.009	-0.011	-0.005
PID performance +5% [G]	0.007	0.003	0.003	0.001	0.002	0.010	0.004
PID performance -5% [G]	0.005	0.003	0.003	0.006	0.002	0.009	0.004
PID performance +10% [G]	0.011	0.002	0.005	0.003	0.001	0.005	0.007
PID performance -10% [G]	-0.005	-0.002	0.002	0.002	0.007	0.007	0.004
PID performance +30% [G]	0.013	0.003	0.003	0.002	0.002	0.006	0.008
PID performance -30% [G]	-0.002	-0.001	0.002	0.001	0.005	0.008	0.004
Tracking efficiency Up [F]	0.001	-0.001	-0.002	-0.000	-0.000	0.001	-0.002
Tracking efficiency Down [F]	0.001	-0.000	0.001	0.000	0.001	0.002	0.002
Trigger efficiency Up [D]	-0.001	-0.002	-0.005	-0.007	-0.003	-0.001	-0.004
Trigger efficiency Down [D]	0.007	-0.001	-0.005	-0.006	-0.004	-0.002	-0.001

Table 42: Variation of F_{L} when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
fileNominal	-0.158 +/- 0.003	-0.151 +/- 0.005	0.318 +/- 0.003	-0.527 +/- 0.003	0.082 +/- 0.003	-0.649 +/- 0.003	0.135 +/- 0.004
AC CTL Up [C]	-0.001	0.006	0.007	0.010	0.003	0.009	-0.002
AC CTL Down [C]	-0.008	-0.004	0.010	-0.009	0.000	-0.008	0.004
AC CTK Up [C]	-0.004	0.000	0.012	-0.006	-0.001	-0.005	-0.001
AC CTK Down [C]	-0.007	-0.001	0.006	0.001	-0.008	0.007	-0.004
AC CTL Up CTK Down [C]	-0.007	-0.007	0.013	-0.004	-0.000	-0.000	-0.007
AC CTL Down CTK Up [C]	-0.002	0.001	0.015	-0.005	0.003	-0.005	0.007
AC CTL Up CTK Down [C]	0.004	0.007	0.002	0.010	-0.002	0.011	-0.004
AC CTL Down CTK Down [C]	-0.004	-0.010	0.010	-0.000	-0.003	-0.000	0.010
AC Non-factorisable Up [C]	-0.003	0.001	0.009	-0.003	0.002	0.000	-0.002
AC Non-factorisable Down [C]	-0.004	-0.000	0.008	0.004	0.001	-0.000	-0.000
AC q^2 binning +1 [B]	-0.012	0.002	-0.004	0.007	0.015	-0.005	-0.002
AC q^2 binning -1 [B]	0.011	-0.022	-0.007	-0.014	0.011	0.015	0.013
1st Order Background Model [K]	-0.007	-0.006	0.013	-0.006	0.006	-0.009	0.005
3rd Order Background Model [K]	-0.002	0.008	0.008	0.003	-0.001	0.002	-0.001
K^{*0} mis-ID Down [L]	-0.008	0.006	0.007	0.003	0.001	0.004	-0.014
K^{*0} mis-ID Up [L]	-0.004	0.008	0.009	0.003	-0.003	0.006	-0.007
Signal mass width Down [J]	0.003	-0.002	0.005	0.010	0.003	0.003	0.000
Signal mass width Up [J]	-0.005	-0.001	0.010	-0.002	0.002	-0.002	0.008
S-wave component [P]	0.009	0.011	-0.032	0.049	-0.006	0.051	-0.023
IP Smearing [H]	0.002	-0.005	0.004	0.008	0.013	-0.031	0.024
B^0 p_T re-weighting [I]	-0.009	-0.006	0.004	0.005	0.000	0.002	-0.002
B^0 p re-weighting [I]	-0.009	-0.002	0.010	0.009	0.004	0.013	-0.004
IsMuon efficiency Up [E]	-0.004	0.002	0.019	0.001	0.003	-0.004	0.002
IsMuon efficiency Down [E]	-0.006	0.007	0.000	-0.006	-0.010	-0.003	-0.006
Removal of soft tracks [O]	-0.013	0.007	-0.001	0.012	0.005	-0.000	-0.015
PID performance +5% [G]	0.003	0.007	-0.001	0.004	-0.006	-0.005	-0.007
PID performance -5% [G]	-0.006	0.005	0.001	0.002	0.002	0.000	-0.009
PID performance +10% [G]	0.010	0.003	-0.005	-0.001	-0.003	0.007	-0.009
PID performance -10% [G]	0.003	-0.010	0.003	0.002	-0.017	-0.013	-0.008
PID performance +30% [G]	0.009	0.001	-0.006	0.001	-0.006	0.009	-0.011
PID performance -30% [G]	0.003	0.006	-0.002	0.001	-0.018	-0.007	-0.004
Tracking efficiency Up [F]	-0.004	-0.022	0.005	-0.001	0.003	0.001	-0.006
Tracking efficiency Down [F]	-0.008	0.003	0.007	-0.005	0.001	-0.008	-0.002
Trigger efficiency Up [D]	-0.003	0.012	0.005	0.003	-0.001	0.006	0.003
Trigger efficiency Down [D]	-0.009	-0.003	0.011	-0.000	0.000	0.003	0.009

Table 43: Variation of A_{T}^2 when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
fileNominal	0.176 +/- 0.003	-0.094 +/- 0.006	0.044 +/- 0.003	-0.076 +/- 0.003	-0.001 +/- 0.003	0.184 +/- 0.003	0.349 +/- 0.004
AC CTL Up [C]	0.001	0.001	-0.000	0.009	0.008	-0.001	-0.001
AC CTL Down [C]	0.009	-0.004	-0.002	0.005	0.004	-0.001	0.013
AC CTK Up [C]	0.004	0.007	0.003	0.001	0.008	-0.001	0.004
AC CTK Down [C]	-0.000	-0.002	-0.000	0.008	0.006	-0.000	0.004
AC CTL Up CTK Down [C]	0.002	0.004	0.002	0.010	0.008	-0.002	0.000
AC CTL Down CTK Up [C]	0.006	-0.004	0.004	0.003	0.005	0.004	0.015
AC CTL Up CTK Down [C]	-0.005	0.003	-0.000	0.001	0.006	0.002	-0.009
AC CTL Down CTK Down [C]	-0.002	0.006	0.005	0.004	0.004	0.003	0.009
AC Non-factorisable Up [C]	0.000	-0.006	0.001	0.005	0.004	0.003	0.007
AC Non-factorisable Down [C]	0.002	0.003	0.002	0.004	0.005	-0.001	0.000
AC q^2 binning +1 [B]	0.002	0.002	-0.001	0.011	0.008	0.002	0.007
AC q^2 binning -1 [B]	-0.003	-0.011	-0.003	0.004	0.003	-0.006	0.011
1st Order Background Model [K]	-0.002	-0.001	0.007	0.001	0.006	-0.005	0.006
3rd Order Background Model [K]	0.005	0.012	-0.002	0.002	0.004	-0.002	0.000
K^{*0} mis-ID Down [L]	0.003	-0.005	-0.000	0.007	0.002	0.001	0.003
K^{*0} mis-ID Up [L]	0.005	-0.006	0.005	0.008	0.005	-0.001	0.004
Signal mass width Down [J]	0.004	0.015	-0.004	0.009	0.008	0.002	0.003
Signal mass width Up [J]	0.003	-0.001	0.001	0.000	0.007	-0.003	0.009
S-wave component [P]	-0.016	0.000	-0.001	0.012	0.007	-0.014	-0.038
IP Smearing [H]	0.007	0.001	-0.002	0.004	0.004	-0.001	0.002
B^0 p_T re-weighting [I]	0.001	0.004	0.002	0.009	0.007	-0.006	0.013
B^0 p re-weighting [I]	0.006	0.005	0.003	0.001	0.008	0.001	0.001
IsMuon efficiency Up [E]	0.005	-0.001	0.002	0.005	0.004	-0.003	0.003
IsMuon efficiency Down [E]	-0.005	0.002	-0.001	0.006	0.003	-0.008	0.006
Removal of soft tracks [O]	0.008	0.001	-0.007	0.006	0.005	-0.001	0.003
PID performance +5% [G]	-0.001	-0.007	0.006	-0.001	0.005	0.005	0.005
PID performance -5% [G]	0.006	0.006	0.004	0.007	0.013	0.002	0.007
PID performance +10% [G]	0.004	0.003	-0.006	0.006	0.007	0.002	0.005
PID performance -10% [G]	0.004	0.002	0.000	-0.001	0.006	-0.002	0.007
PID performance +30% [G]	0.001	-0.008	-0.004	0.006	0.012	-0.002	0.003
PID performance -30% [G]	0.001	0.003	-0.003	0.002	0.005	0.001	-0.003
Tracking efficiency Up [F]	-0.000	-0.012	-0.005	0.005	0.002	-0.002	0.001
Tracking efficiency Down [F]	0.007	-0.001	-0.002	-0.003	0.006	-0.003	0.013
Trigger efficiency Up [D]	0.003	0.004	0.001	0.008	0.006	-0.000	0.010
Trigger efficiency Down [D]	0.006	-0.002	0.005	0.003	0.008	-0.004	0.007

Table 44: Variation of A_{T}^{Im} when systematically varying fit parameters or the weights applied to the input data set.

20 Result plots and tables

Figures. 43-45 show the results of the fits for F_{L} and the two sets of observables A_{FB}, S_3, A_9 and A_{T}^{Re}, A_{T}^{2}, A_{T}^{Im} in the six q^2-bins. The statistical uncertainty on the points was obtained using the Feldman-Cousins technique. The results are also presented in Table. 45 below.

The SM prediction for the angular observables, and the prediction rate-averaged over the q^2 bin, are also indicated on the figures. No SM prediction is included for the region between the $c\bar{c}$ resonances where the assumptions made in the prediction break down. No theory band is included for A_9 and A_{T}^{Im}, which are expected to be small, $\mathcal{O}(10^{-3})$ [25], in the SM. The theory band is also omitted for another reason, unlike the other observables, it could be sensitive to the SM contributions from helicity suppressed (by m_s/m_b) right-handed currents, that are usually neglected in the calculation. The observable S_9 is suppressed by the small size of the strong phase difference and is expected to be vanishingly small.

20.1 Normal variables

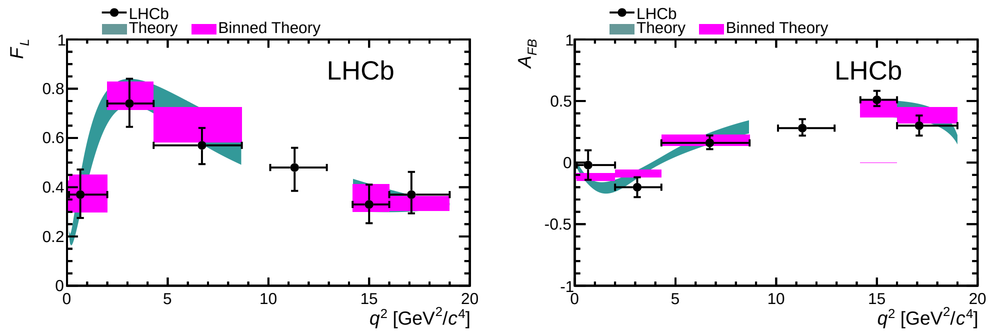


Figure 43: Fraction of longitudinal polarisation of the K^{*0}, F_{L} and dimuon forward-backward asymmetry, A_{FB}, as a function of q^2.

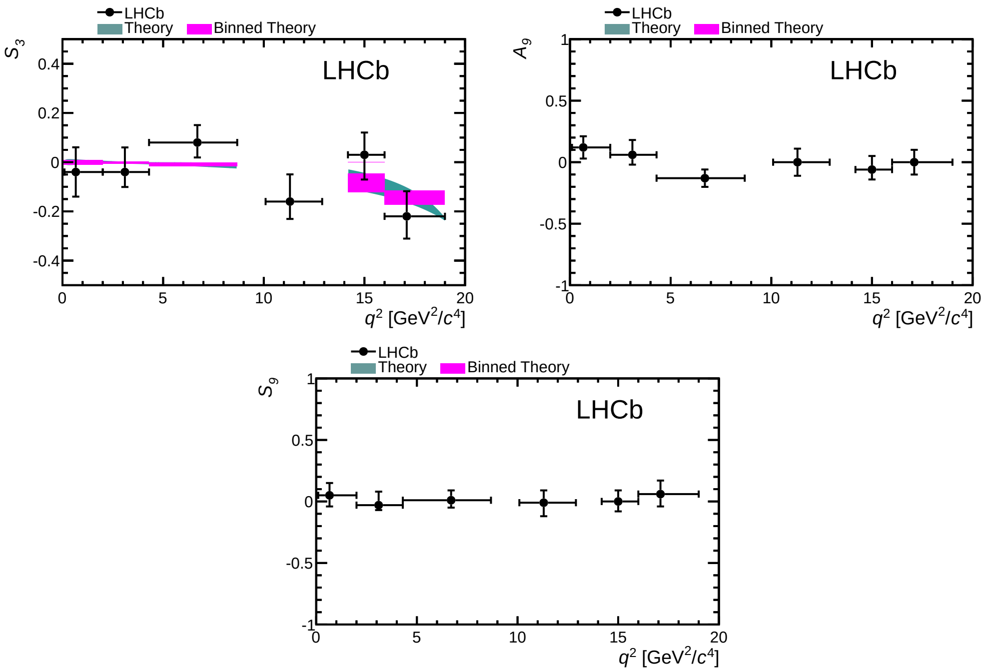


Figure 44: The observables S_3, S_9 and A_9 as a function of q^2.

20.2 Reparam variables

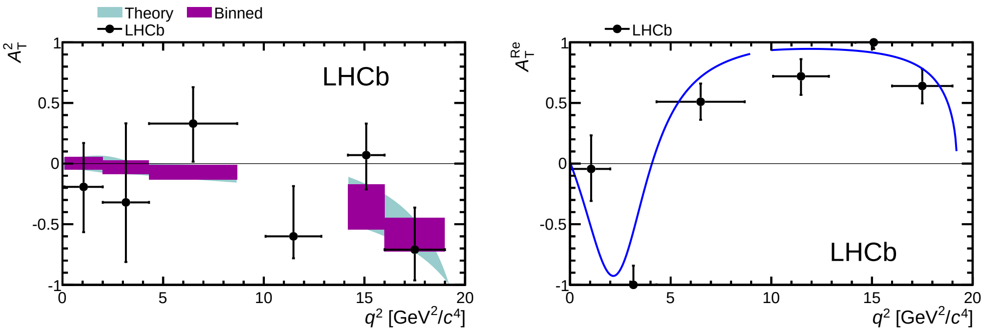


Figure 45: Transverse asymmetries, A_{T}^{Re} and A_T^2 as a function of q^2. No theory band is included for the A_{T}^{Re} prediction, the central value of the theory prediction is however indicated by the continuous (blue) curve.

Table 45: Central values for, and statistical and systematic uncertainties on A_{FB}, F_{L}, S_9, S_3, A_{T}^{Re}, A_{T}^2 and A_{T}^{Im} in bins of q^2. The first uncertainty is statistical and the second systematic.

	A_{FB}	F_{L}	S_3	S_9
$0.10 < q^2 < 2.00$	$-0.02^{+0.12+0.00}_{-0.12-0.00}$	$+0.37^{+0.10+0.02}_{-0.09-0.03}$	$-0.04^{+0.10+0.01}_{-0.10-0.00}$	$+0.05^{+0.10+0.00}_{-0.09-0.01}$
$2.00 < q^2 < 4.30$	$-0.20^{+0.08+0.01}_{-0.08-0.01}$	$+0.74^{+0.10+0.01}_{-0.09-0.03}$	$-0.04^{+0.10+0.00}_{-0.06-0.01}$	$-0.03^{+0.11+0.00}_{-0.04-0.00}$
$4.30 < q^2 < 8.68$	$+0.16^{+0.06+0.00}_{-0.05-0.01}$	$+0.57^{+0.07+0.01}_{-0.07-0.03}$	$+0.08^{+0.07+0.01}_{-0.06-0.01}$	$+0.01^{+0.08+0.00}_{-0.06-0.00}$
$10.09 < q^2 < 12.86$	$+0.28^{+0.07+0.02}_{-0.06-0.01}$	$+0.48^{+0.08+0.00}_{-0.09-0.03}$	$-0.16^{+0.11+0.01}_{-0.07-0.01}$	$-0.01^{+0.10+0.00}_{-0.11-0.00}$
$14.18 < q^2 < 16.00$	$+0.51^{+0.07+0.02}_{-0.05-0.01}$	$+0.33^{+0.08+0.01}_{-0.07-0.03}$	$+0.03^{+0.09+0.01}_{-0.10-0.01}$	$0.00^{+0.09+0.00}_{-0.08-0.01}$
$16.00 < q^2 < 19.00$	$+0.30^{+0.08+0.02}_{-0.08-0.01}$	$+0.37^{+0.09+0.02}_{-0.07-0.03}$	$-0.22^{+0.10+0.02}_{-0.09-0.01}$	$+0.06^{+0.11+0.00}_{-0.10-0.01}$
$1.00 < q^2 < 6.00$	$-0.17^{+0.06+0.02}_{-0.06-0.00}$	$+0.65^{+0.08+0.01}_{-0.07-0.03}$	$+0.03^{+0.07+0.00}_{-0.07-0.01}$	$+0.07^{+0.09+0.00}_{-0.08-0.00}$

	A_{T}^{Re}	A_{T}^2	A_{T}^{Im}	A_9
$0.10 < q^2 < 2.00$	$-0.05^{+0.26+0.02}_{-0.24-0.00}$	$-0.14^{+0.34+0.02}_{-0.30-0.02}$	$+0.16^{+0.31+0.02}_{-0.28-0.02}$	$+0.12^{+0.09+0.01}_{-0.09-0.01}$
$2.00 < q^2 < 4.30$	$-1.00^{+0.13+0.04}_{-0.00-0.01}$	$-0.29^{+0.65+0.02}_{-0.46-0.01}$	$-0.23^{+0.77+0.02}_{-0.27-0.01}$	$+0.06^{+0.12+0.01}_{-0.08-0.00}$
$4.30 < q^2 < 8.68$	$+0.50^{+0.16+0.01}_{-0.14-0.03}$	$+0.36^{+0.30+0.03}_{-0.31-0.03}$	$+0.05^{+0.31+0.01}_{-0.31-0.01}$	$-0.13^{+0.07+0.01}_{-0.07-0.01}$
$10.09 < q^2 < 12.86$	$+0.71^{+0.15+0.01}_{-0.15-0.03}$	$-0.60^{+0.42+0.05}_{-0.27-0.01}$	$-0.06^{+0.43+0.03}_{-0.41-0.00}$	$-0.00^{+0.11+0.00}_{-0.11-0.01}$
$14.18 < q^2 < 16.00$	$+1.00^{+0.00+0.01}_{-0.05-0.02}$	$+0.07^{+0.26+0.03}_{-0.28-0.02}$	$+0.02^{+0.27+0.02}_{-0.27-0.00}$	$-0.06^{+0.11+0.01}_{-0.08-0.01}$
$16.00 < q^2 < 19.00$	$+0.64^{+0.15+0.01}_{-0.15-0.02}$	$-0.71^{+0.35+0.06}_{-0.26-0.03}$	$+0.18^{+0.35+0.01}_{-0.32-0.02}$	$-0.00^{+0.10+0.01}_{-0.10-0.01}$
$1.00 < q^2 < 6.00$	$-0.66^{+0.24+0.04}_{-0.22-0.00}$	$+0.17^{+0.39+0.03}_{-0.41-0.02}$	$+0.41^{+0.42+0.02}_{-0.45-0.03}$	$+0.03^{+0.08+0.00}_{-0.08-0.01}$

21 Zero crossing point extraction

As discussed in Sec. 1 the zero-crossing point of $A_{\rm FB}$ (q_0^2) is well defined in the SM and it is sensitive to New Physics through differences in the Wilson coefficients C_7,C_9 and C_{10} which determine the zero-crossing point. A measurement of q_0^2 is therefore an important input to determine whether there are New Physics contributions to the $B^0 \to K^{*0}\mu^+\mu^-$ decay. This measurement is however not straightforward with limited statistics. The simplest imaginable method to determine q_0^2 would be to fit a straight line around the region where $A_{\rm FB}$ changes sign. This procedure is unbiased if $A_{\rm FB}$ can be assumed linear within a known interval around q_0^2. Unfortunately such assumption does not always hold. This method is therefore not applicable unless assumptions on the model are made, e.g. that $A_{\rm FB}$ follows a SM-like curve. In practice, the estimate of q_0^2 becomes dependent on how the data is binned in q^2 and over which range q^2 is assumed to be linear. Moreover, in order to decide a suitable fit range it would be necessary to examine $A_{\rm FB}$ itself. To ensure an unbiased result, this decision should be made without reference to the shape of $A_{\rm FB}$ (q^2).

Instead of performing an angular analysis (and fitting $\cos\theta_l$) to extract $A_{\rm FB}$ in a bin of q^2, an alternative strategy is adopted. Two independent, unbinned, maximum likelihood fits are performed to the distribution of candidates in q^2 for forward- and backward-going events. This procedure is referred to below as an *unbinned counting method*. The PDFs for forward- and backward-going events are expected to have a smooth behaviour as a function as of q^2 in the range $1 - 7.8 GeV^2$, i.e. far from the photon pole and from the J/ψ resonance. The range $1 < q^2 < 7.8\,{\rm GeV}^2/c^4$ is a natural choice. Above $7.8\,{\rm GeV}^2/c^4$ there can be a non-negligible contribution from the radiative tails of the J/ψ (see Sec. 3.4). Below $1\,{\rm GeV}^2/c^4$ the shape of the q^2 spectrum can vary rapidly and can be difficult to parametrise as a smoothly varying polynomial.

In the $1 < q^2 < 7.8\,{\rm GeV}^2/c^4$ range the distribution of forward- and backward-going events can be fitted with polynomial distributions in q^2 and consequently $A_{\rm FB}$ can be computed according to:

$$AFB(q^2) = \frac{N_F PDF_F(q^2) - N_B PDF_B(q^2)}{N_F PDF_F(q^2) + N_B PDF_B(q^2)}. \tag{18}$$

where $N_{F,B}$ is the number of forward- and backward-going events and $PDF_{F,B}$ is the fitted PDFs as a function of q^2 for forward- and backward-going signal events. To separate signal and background, the fit is performed in two dimensions: in the invariant mass of the B^0 candidate and q^2. The q^2 distribution of the signal has been parametrised with a third order Chebychev polynomial. The mass model described in Sec. 4 is used for the signal mass shape. The impact of the detector acceptance is accounted for by weighting candidates in the fit as described in Sec. 11.

In summary the analysis strategy for measuring the zero-crossing point consists of fitting separately the q^2-dependence of forward and backward events. The goodness of fit for forward- and backward going events will be estimated before computing A_{FB} using the point-to-point dissimilarity technique described in Ref. [1]. Finally the A_{FB} is estimated by combining the q^2 dependence of the forward- and backward-going events. The estimation of the uncertainty on the zero-crossing point is described in Sec. 21.1 below.

21.1 Estimating the 68% confidence level on q_0^2

MC studies have shown that the error distribution of the coefficients of the polynomials is often not Gaussian and therefore an estimate for the uncertainty on the crossing point can not be calculated directly from the covariance matrix of the fit. The use of event weights, can also lead the $\Delta LL = 1/2$ estimate to under-estimate the 68% confidence interval.

Two methods have been explored to estimate the uncertainty on q_0^2:

- the use of bootstrapping to obtain a confidence interval.
- Toy MC generated from the fitted forward and backward pdf.

These methods are described in more detail below.

21.1.1 Bootstrapped confidence interval

A 'bootstrap' method is used to calculate the 68% confidence interval on the zero-crossing point. Bootstrapping uses a re-sampling technique to generate many individual data samples.

Schematically, what is done is to take the dataset of N events,

$$d = \{\vec{\Omega}_0, \vec{\Omega}_1, \ldots, \vec{\Omega}_{N-2}, \vec{\Omega}_{N-1}\}$$

and to create a new, re-sampled dataset from it of the same size (the number of events is varied according with a Poisson distribution), d_1. The re-sampling allows events to be duplicated, e.g.:

$$d_1 = \{\vec{\Omega}_0, \vec{\Omega}_0, \ldots, \vec{\Omega}_{N-2}, \vec{\Omega}_{N-1}\}$$

would be allowed where event '0' appears twice and event '1' is omitted from d_1. The likelihood fit for the zero-crossing point is the performed on each of the re-sampled datasets, leading to a distribution of zero-crossing points. This distribution is then used to estimate the 68% confidence interval on q_0^2.

21.1.2 Confidence interval with toy study

To crosscheck the estimation of the uncertainty obtained with bootstrapping, a slightly different approach was performed as well. The pdfs for the forward and backward distributions were used as an input to a toy simulation. In this simulation, many datasets were created, where the events where distributed following the input pdfs and the number of events in the datasets were fluctuated following a poissonian distribution around the value measured in collision data. For all these samples the zero-crossing point was determined and the 68% confidence interval evaluated in the same was as for the bootstrapping. The resulting interval is a bit more narrow than the one obtained with the bootstrapping but still in good agreement. The difference may be a consequence of randomising the weights in the bootstrapping, which is not the case for this technique.

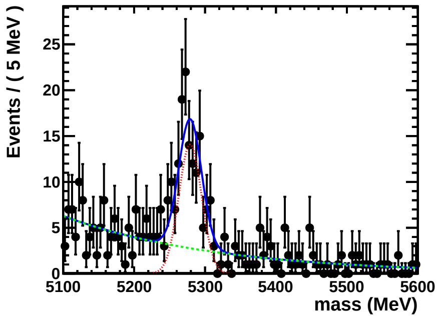
21.2 MC study for the zero-crossing extraction

Toy Monte Carlo studies have been performed before the unblinding to validate the method described above and study its sensitivity to a SM-like A_{FB}. The toys were generated with a SM-like q^2 dependence of forward- and backward going events and the expected signal-to-background ratio and signal yield in $1 < q^2 < 7.8\,\mathrm{GeV}^2/c^4$. The distribution of forward- and backward-going background events was taken from the upper mass sideband of the data. Fig. 46 shows the $K^+\ \pi^-\ \mu^+\mu^-$ invariant mass and q^2 distribution for a single toy experiment. A fit to the B^0 mass and q^2 is overlaid.

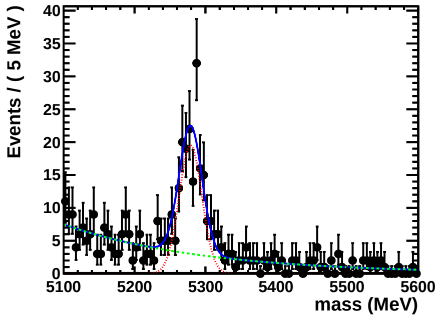
The result of performing 200 toys with a SM-like zero-crossing point is shown in Fig. 47. The mean value of A_{FB} in the 200 toys is found to be consistent, as expected, with the SM input distribution.

Unfortunately, due to statistical fluctuations, with $1\,\mathrm{fb}^{-1}$ it is not guaranteed that there will be a single, well-defined zero-crossing point. According to MC simulations, in the SM, there is about a 20% probability to measure either no zero-crossing point, or more than one zero-crossing, in a data sample corresponding to $1\,\mathrm{fb}^{-1}$. An illustration of this effect is shown in Fig. 48. It was decided before unblinding the data to quote a zero-crossing point only if the fit to data shows a single well defined value (alternatively the 90% CL will be given).

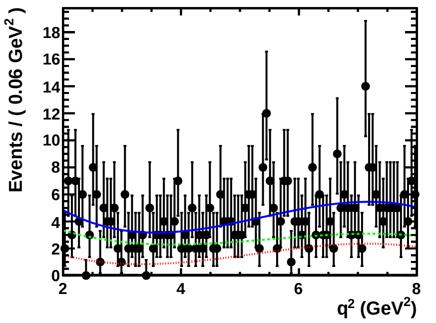
It is also apparent from toy-studies that the errors on the fit parameters are not Gaussian. The covariance matrix from the fit is therefore not a good



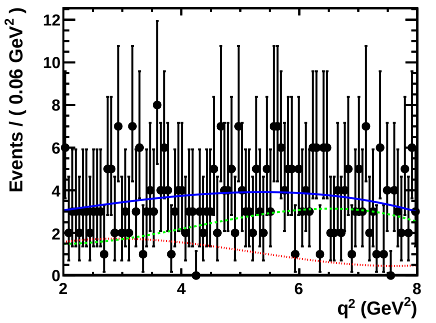
(a) Forward-going events



(b) Backward-going events



(c) Forward-going events



(d) Backward-going events

Figure 46: Fit to the invariant mass of the B-meson candidate, for forward (a) and backward (b) events and fit to the q^2 distribution for forward (c) and backward (d). The signal component (red) and background component (green) are indicated.

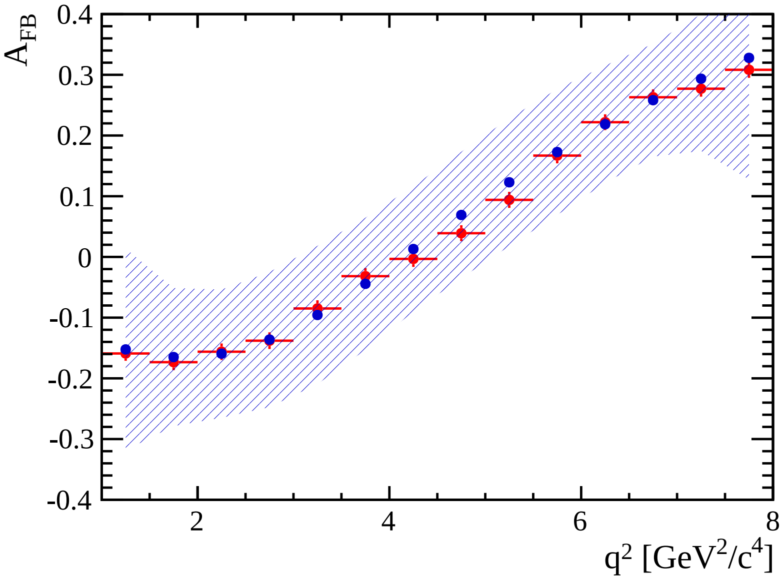
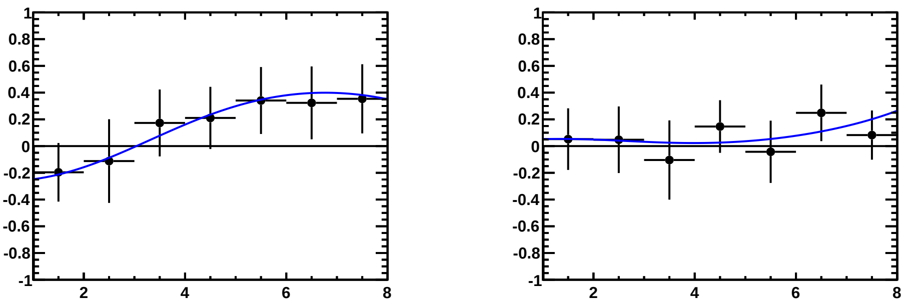


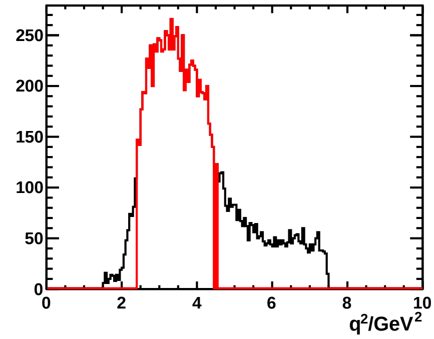
Figure 47: The hashed region represents the 68% confidence region from 200 toys at each q^2 value for a SM-like q^2 dependence of forward- and backward-going events. The blue marker is the mean value of $A_{\rm FB}(q^2)$ for the 200 toys, and the red marker is the true value of $A_{\rm FB}(q^2)$ used as input to the toy-MC.



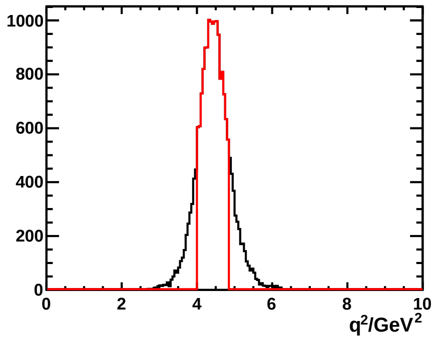
(a) Example of a sub-sample with one single zero-crossing point

(b) Example of a sub-sample with no zero-crossing point

Figure 48: Two examples of $A_{\rm FB}$ obtained from toy-studies with the unbinned counting method. The toy experiment were carried out with statistics equivalent to $1\,{\rm fb}^{-1}$ and a SM-like $A_{\rm FB}(q^2)$. The data-points in the figure are a binned estimate of $A_{\rm FB}$ in $1\,{\rm GeV}^2/c^4$ q^2 bins. The left-hand figure is indicative of an 'unlucky' result where, due to statistical fluctuations, no zero-crossing point is visible.



(a)

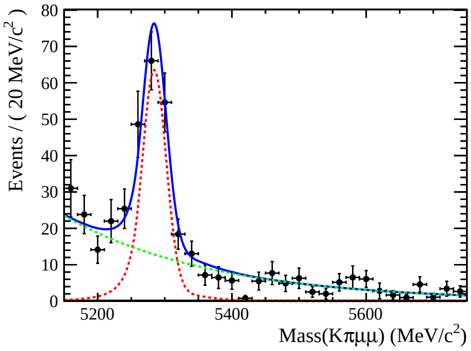


(b)

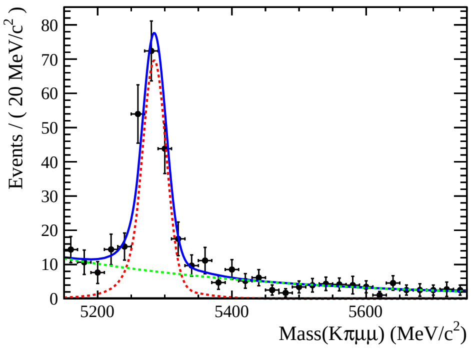
Figure 49: Examples of 'posterior' distributions obtained for the zero-crossing point of the A_{FB} for two different toy-MC experiments.

estimate of parameter errors, and cannot be used to estimate the uncertainty on the zero-crossing point. Two examples are show in Fig. 49.

The impact of the order of the polynomials has also been studied by using the MC simulation and found to be negligible for polynomials of order higher than three.



(a) Forward-going events



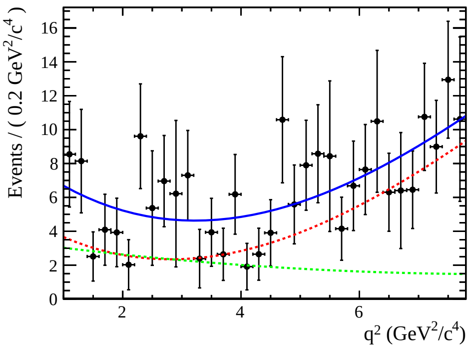
(b) Backward-going events

Figure 50: Fit to the invariant mass of the B-meson candidate, for forward and backward going events in data.

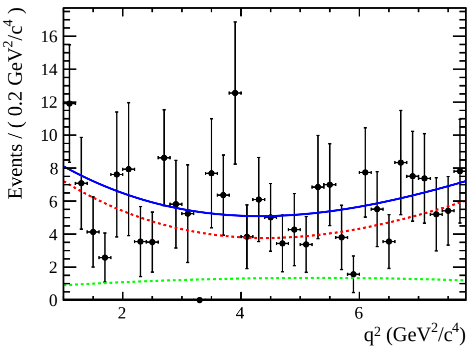
22 Zero crossing point result

The procedure described in the previous sections for the extraction of the zero-crossing point is here applied to data. The invariant mass of the B^0 candidates is shown in Fig 50 for forward- and backward-going events, the result of the fit is also shown. The q^2 distribution for forward- and backward-going events in the signal region is shown in Fig. 51. After fitting separately forward and backward events the quality of the fit was investigated with the point-to-point dissimilarity technique, the p-value obtained was 0.6 for the fit to the forward events and 0.9 for the fit to the backward events.

The forward-backward asymmetry is shown in Fig 52, the curve is the result of the *unbinned counting method* applied to data, the points are the result of a simple counting experiment used as a cross check. The distribution of the zero-crossing points for several toy distribution assuming the PDF



(a) Forward-going events



(b) Backward-going events

Figure 51: Fit to q^2 distribution for forward and backward going events in data.

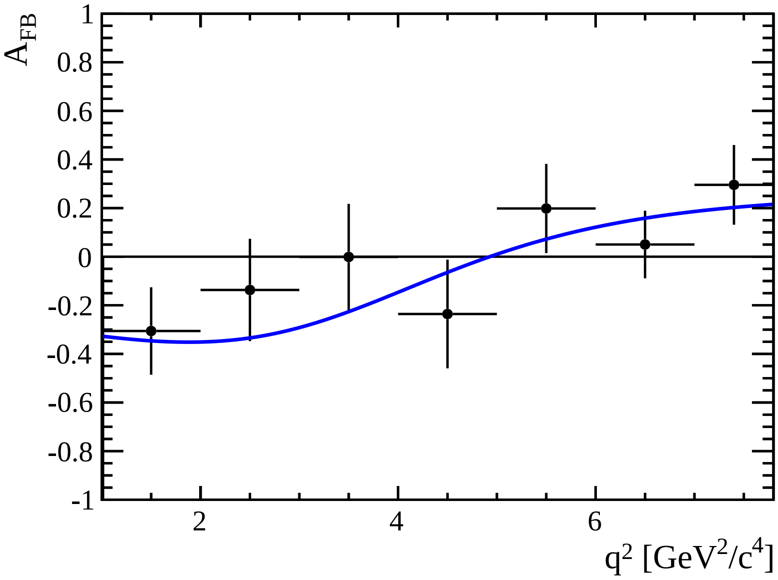
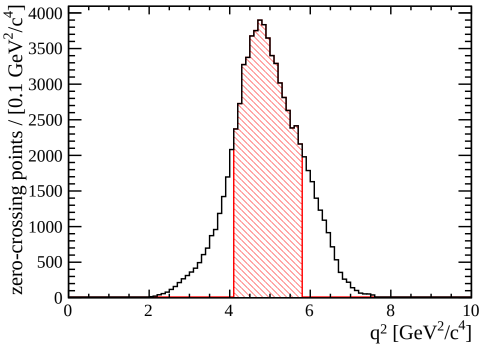


Figure 52: The A_{FB} as a function of q^2, that comes from the unbinned counting experiment (blue dashed line). The data-points are the result of counting forward- and backward-going events in $1\,\mathrm{GeV}^2/c^4$ bins of q^2.



(a) Toy MonteCarlo

Figure 53: The distribution of the zero-crossing points for toy experiments generated by assuming the forward and backward Pdfs measured in data.

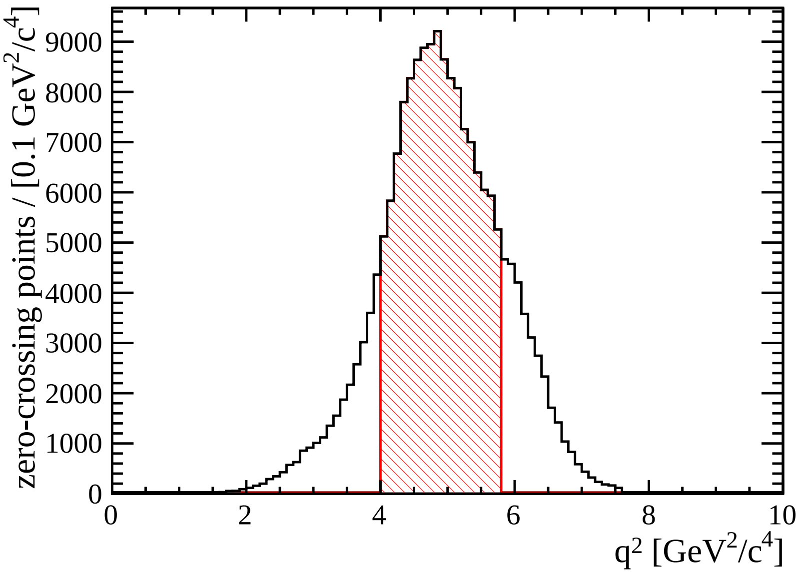


Figure 54: The distribution of the zero-crossing points in the bootstrapping method. The red region shows the 68% CL.

measured in data is shown Fig. 53.

The distribution of zero crossing points for the bootstrapping (re-sampling) technique is shown in Fig. 54. The result, which only includes the statistical error is:

$$q_0^2 = (4.9^{+0.9}_{-0.9})\,\mathrm{GeV}^2/c^4, \tag{19}$$

where the error has been determined by re-sampling (bootstrapping) the data 200'000 times, see Sec. 21.1.1 for a description of the method. The error is in very good agreement to what is expected when generating many toy-experiments, where the result is $q_0^2 = (4.9^{+0.9}_{-0.8})\,\mathrm{GeV}^2/c^4$ (compare Fig 54 with Fig. 53). The toy study was carried out by generating pseudo-experiments at the central value measured in data. The number of event observed in the data is Poisson-fluctuated in the toy-experiments. The method is described in more detail in Sect. 21.1.2.

22.1 Systematic uncertainties

The following sources of systematic errors were considered:

1 Uncertainty in the IP smearing: The fit is repeated using an acceptance model where the MC sample is not IP smeared.

2 Uncertainty in the binning of the PID variables: To account for this uncertainty, 50% of the events in the lowest 30% of a certain bin were migrated to the lower bin and 50% of the events in the highest 30% of the bin were migrated to the higher bin.

3 Uncertainty on the tracking efficiency: Possible systematic effects are taken into account by assigning the tracks with a momentum lower than $10\,\mathrm{GeV}/c$ an efficiency which is lower (higher) by one standard deviation and by assigning the tracks with a momentum higher than $10\,\mathrm{GeV}/c$ an efficiency which is higher (lower) by one standard deviation.

4 Uncertainty in the trigger efficiency: Systematic effects were accounted for by increasing or decreasing the trigger efficiency for muons with a momentum below $3\,\mathrm{GeV}/c$ by 3% for the acceptance correction.

5 Uncertainty of the `IsMuon` criterion: The systematic uncertainty is assessed by fluctuating downwards the efficiency for tracks with a momentum less than $10\,\mathrm{GeV}/c$ by the statistical uncertainty and by fluctuating upwards the efficiency for tracks with a momentum more than $10\,\mathrm{GeV}/c$ by the statistical uncertainty. The procedure is also repeated by changing the direction of fluctuation for the corresponding two categories.

6 Acceptance correction: The acceptance correction is varied as described in Sec. 18.3.

7 The widths (σ) of the Gaussian component of both crystal ball functions shows a slight dependence on q^2 which amounts to a slope corresponding to about 5%. These widths are therefore varied by $\pm 5\%$ in the fit and the result is recalculated.

Furthermore, some crosschecks were performed as well:

8 The fit was performed with and without reweighting the momentum of the B in the simulation to the values of the collision data.

9 The fit was performed with and without reweighting the transverse momentum of the B in the simulation to the values of the collision data.

10 The fit was performed with and without cutting on the momentum of $3\,\mathrm{GeV}/c$ on the hadrons.

The zero crossing points, evaluated under the changes to the data sample corresponding to the systematic checks, are listed in Table 46. Even when summing the systematic uncertainties and the deviations from the cross-checks in quadrature, which clearly overestimates the uncertainty, the overall systematic uncertainty is small compared to the statistical uncertainty and was not included in the overall uncertainty.

Table 46: Values for the zero-crossing point and deviation from the nominal value for all evaluations of the systematic uncertainty and the performed crosschecks. The type corresponds to the type given in the list of systematic uncertainties and crosschecks. The overall systematic uncertainty is calculated by adding all contributions (also the ones from the crosschecks) in quadrature.

Type	q_0^2 [GeV^2/c^4]	**Deviation** [GeV^2/c^4]
1	4.92	0.01
2	4.93	0.00
	4.94	0.01
3	4.92	0.01
	4.93	0.00
4	4.93	0.00
	4.93	0.00
5	4.95	0.02
	4.92	0.01
6	4.93	0.00
	4.92	0.01
	4.92	0.01
	4.93	0.00
	4.92	0.01
	4.93	0.00
	4.95	0.02
	4.90	0.03
7	4.93	0.00
	4.92	0.01
8	4.94	0.01
9	4.93	0.00
10	4.94	0.01
total	—	0.05

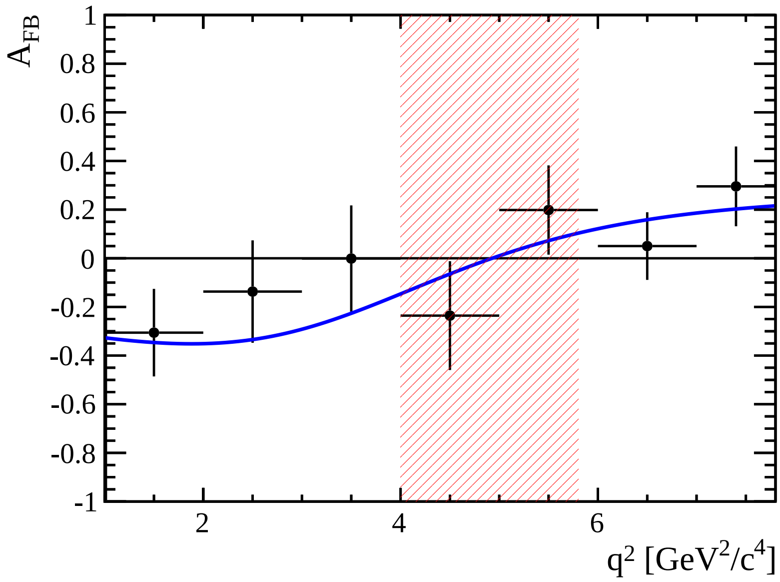


Figure 55: The A_{FB} as a function of q^2, obtained with unbinned counting (blue dashed line). The black data-points are the result of counting forward- and backward-going events in $1\,\mathrm{GeV}^2/c^4$ bins of q^2. The red hashed region corresponds to the 68% confidence interval.

22.1.1 Result plot

A plot of A_{FB} obtained with the unbinned counting method, the counting experiment in $1\,\mathrm{GeV}^2/c^4$ bins and the 68% confidence interval on q_0^2 can be seen in Fig. 55.

22.2 Changes with respect to the preliminary result

The preliminary result quoted in Ref. [8], based on the same dataset, has

$$q_0^2 = 4.9^{+1.3}_{-1.1}\,\mathrm{GeV}^2/c^4 \quad .$$

The difference between the result presented here and this preliminary result is due (predominatly) to a bug that was discovered in the preliminary result. The bug related to the use of weighted datasets in RooFit. It was discovered that when cloning a weighted dataset, information about the weights was lost (even though the dataset still had a flag set to say that it was weighted). Without the weights applied the forward backward asymmetry is reduced, reducing the gradient of A_{FB} in the region around the zero-crossing point and increasing the error on q_0^2. As expected, the value of q_0^2 itself is almost unchanged by turning on/off the weights to correct for the acceptance correction. The effect is largest for low q^2 where the acceptance effects in $\cos\theta_\ell$ can be large.

23 Conclusions

Measurements of the differential branching fraction and angular observables S_3 (A_T^2), F_{L}, S_9, A_{FB} (A_T^{Re}) and the CP asymmetry A_9 of the $B^0 \to K^{*0}\mu^+\mu^-$ decay have been presented, using $1\,\mathrm{fb}^{-1}$ of integrated luminosity collected by LHCb in 2011. These are the most precise measurements of these quantities to date and are consistent with the SM predictions. A first measurement of the zero-crossing point of the forward-backward asymmetry has also been presented. The zero-crossing point is determined to be $q_0^2 = (4.9^{+0.9}_{-0.9})\,\mathrm{GeV}^2/c^4$.

The angular analysis and zero-crossing point measurement are currently statistically limited. For the differential branching fraction the statistical uncertainties are comparable to the size of the systematic uncertainties. The measurement would, however, no longer be systematically limitted if it were binned finer in q^2 and this should be considered for future iterations of the analysis.

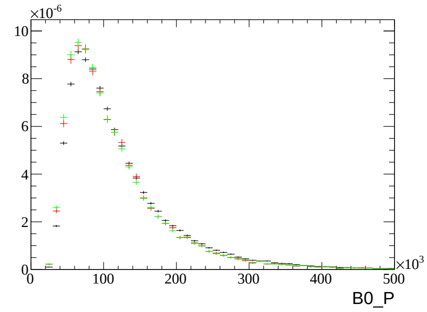
The systematic uncertainty coming from the acceptance correction can be viewed as being fairly conservative and could improve with increased MC statistics and a better understanding of the $B^0 \to K^{*0}J/\psi$ control channel (where at the extremes of $\cos\theta_K$ the data disagrees with the fit-model at the level of $\sim 5\%$).

Appendix

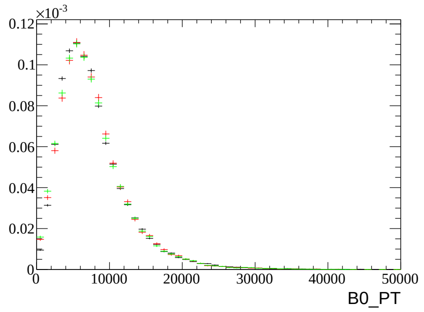
This appendix includes supplementary information for the analysis.

A Data/MC comparison

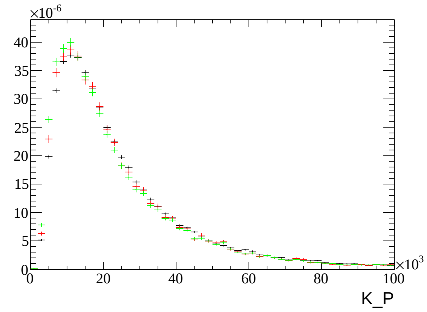
The momentum and p_{T} distribution of $B^0 \to K^{*0} J/\psi$ candidates in the MC (`MC11a`) have been cross checked with the data after the application of the full offline selection (and IP smearing of the MC) and are found to be in good agreement. The distributions of the B^0 and daughter momentum are shown in Fig. 56. The DLL distribution of the daughters is shown in Fig. 58. The IP smearing of the daughter track states tends to over smear the end vertex quality of the fitted B vertex (see Fig. 59). This quantiy is not very correlated to q^2 or to the angualr distribution of the K^{*0} or dimuon system and differences between data and MC can be safely ignored.



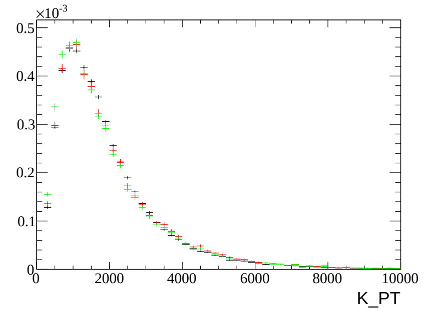
(a) B^0 P distribution



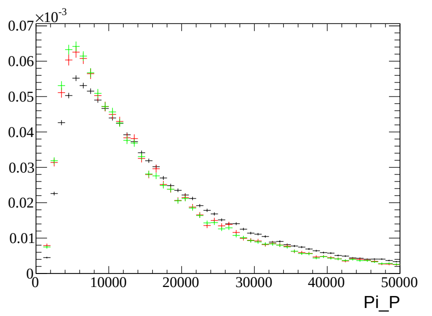
(b) B^0 p_{T} distribution



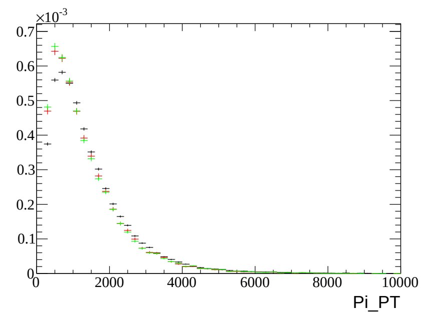
(c) K P distribution



(d) K p_{T} distribution



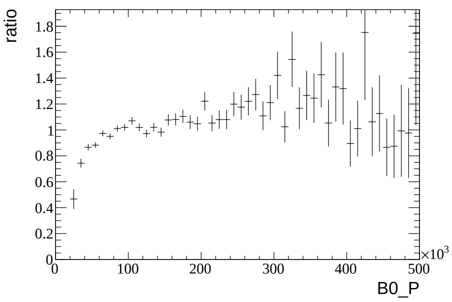
(e) π P distribution



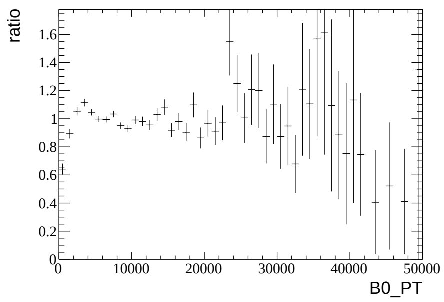
(f) π p_{T} distribution

Figure 56: Comparison of the B^0 and daughter momentum and p_{T} distributions for $B^0 \to J/\psi\, K^{*0}$ candidates in the data and the MC. The three distributions are Data (Black), data-corrected simulated events (Red) and uncorrected simulated events (Green)

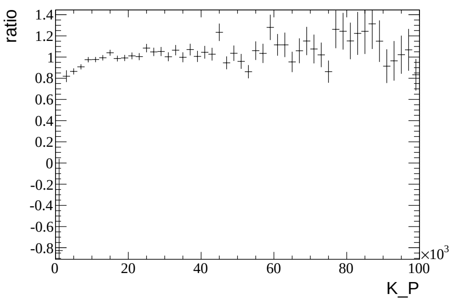
The comparison between the data and the simulation has been investigated after re-weighting to correct for the small disagreement in the underlying B-momentum spectrum. This is shown is Fig. 57. Even after re-weighting for difference in the underlying B^0 momentum spectrum between data and MC, a perfect agreement is still not expected between the daughter momentum and transverse-momentum spectrums. Difference are expected due to a $\sim$ 7% S-wave contribution in the data, that is not present in the MC. The intereference between the S-wave and P-wave resutls in a forward backward asymmetry in $\cos\theta_K$, which in turn produces a harder pion momentum spectrum in data than in the MC.



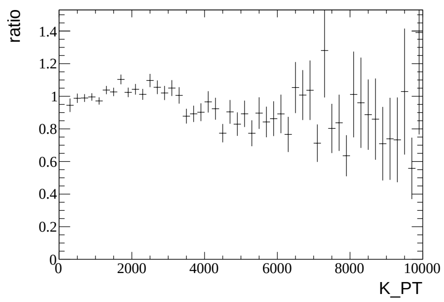
(a) B^0 P distribution



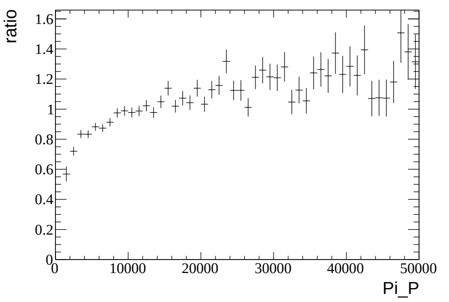
(b) B^0 p_{T} distribution



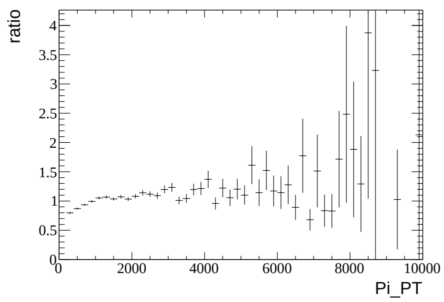
(c) K P distribution



(d) K p_{T} distribution

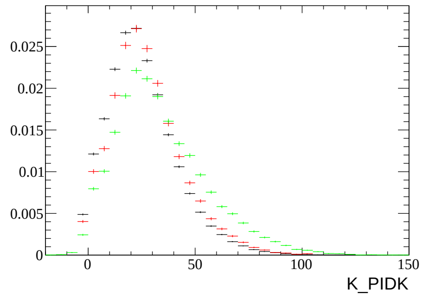


(e) π P distribution

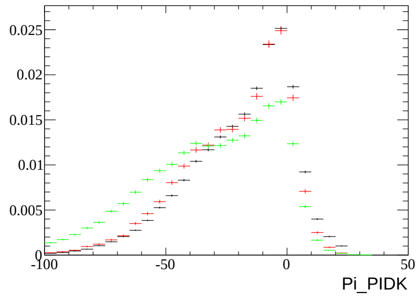


(f) π p_{T} distribution

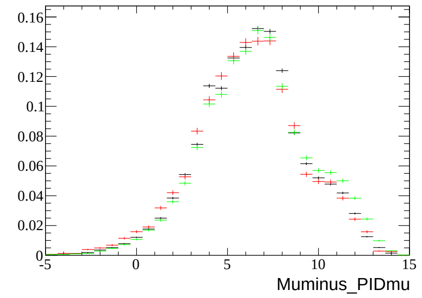
Figure 57: Ratio of the B^0 and daughter momentum and p_{T} distributions for $B^0 \to J/\psi K^{*0}$ candidates in the data and the MC. The three distributions are Data/corrected simulation (Black), data / uncorrected simulated events (Red)



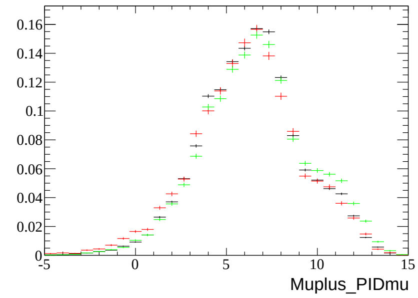
(a)



(b)



(c)



(d)

Figure 58: Comparison of the daughter DLL distributions for $B^0 \to J/\psi K^{*0}$ candidates in the data and the MC. The three distributions are Data (Black), data-corrected simulated events (Red) and uncorrected simulated events (Green)

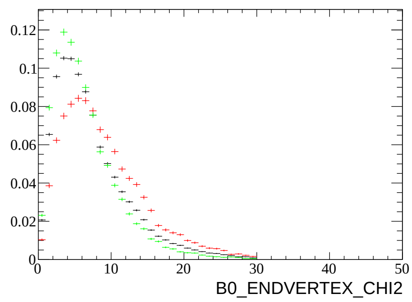


Figure 59: Comparison of the B end vertex χ^2 distributions for $B^0 \to J/\psi K^{*0}$ candidates in the data and the MC. The three distributions are Data (Black), data-corrected simulated events (Red) and uncorrected simulated events (Green)

In general there is good agreement between data and MC for all of the input variables that are used in the BDT. The first order correlations between the different variables are also in general very well re-produced. The only a couple of places where the correaltions are not faithfully reproduced: the correlation between the B end vertex and the impact parameter of the daughters and the correlation between the various daughter DLL dsitributions. The latter is dilluted in the MC by the re-sampling that is applied.

A.1 Comparison of data and MC efficiency

As a further check of the data-MC agreement, Fig. 60 shows the ratio of offline selected to stripped candidates as a function of $\cos\theta_\ell$, $\cos\theta_K$ and the ϕ angle in data and MC for a BDT cut at 0.1. Within the present statistics, the MC accuratley reproduces the distribution seen in the data.

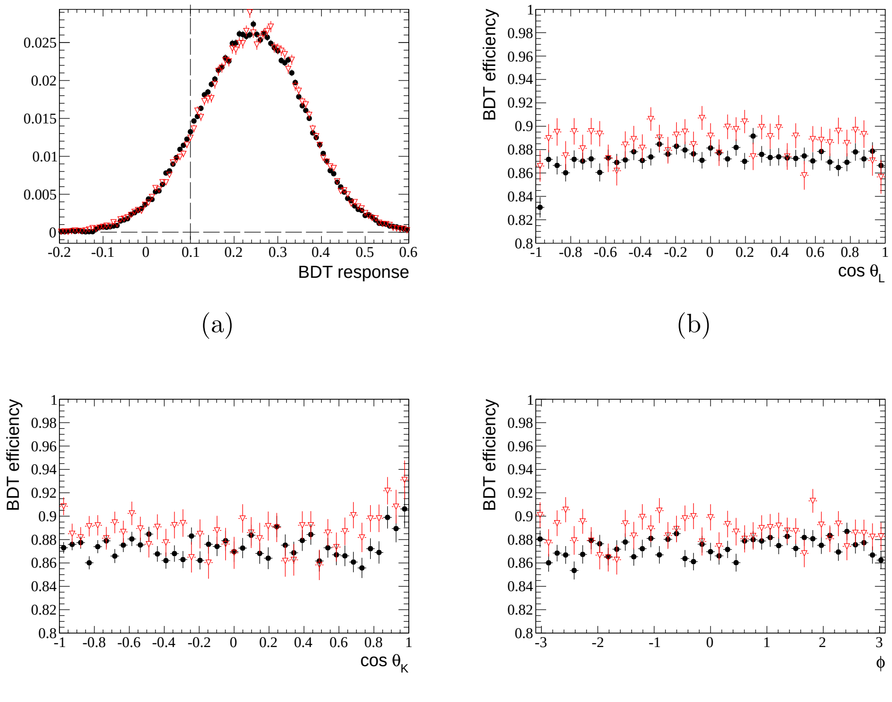


Figure 60: Comparison of the BDT cut “efficiency” as a function of $\cos\theta_\ell$, $\cos\theta_K$ and ϕ between data and MC for background subtracted $B^0 \to J/\psi\, K^{*0}$ candidates. The solid (black) markers are fromthe data. The open (red) markers from MC. Fig (a) shows the BDT distribution for data/MC.Events are selected offline if the BDT response is larger than 0.1.

B Factorisation of the acceptance correction

If the efficiency in a narrow bin of q^2 can be factorised into separate functions of $\cos\theta_l$, $\cos\theta_K$ and ϕ:

$$\varepsilon(\cos\theta_l, \cos\theta_K, \phi) = \varepsilon(\cos\theta_l)\varepsilon(\cos\theta_K)\varepsilon(\phi)$$

and the underlying 'physics' distribution of the events can also be factorised, then the efficiency as a function of ϕ can be written as:

$$\varepsilon(\phi) = \frac{\int\int \frac{d^3\Gamma}{d\cos\theta_l\, d\cos\theta_K\, d\phi}\varepsilon(\cos\theta_l, \cos\theta_K, \phi)\, d\cos\theta_l\, d\cos\theta_K}{\int\int \frac{d^3\Gamma}{d\cos\theta_l\, d\cos\theta_K\, d\phi} d\cos\theta_l\, d\cos\theta_K}$$

It is a simple ratio of the distribution of the number of events after selection as a function of ϕ to the distribution at generator level (before production cuts). If the underlying physics does not factorise into three separate angular dsitributions, then even if the acceptance factorises it is not possible to estimate the efficiency in ϕ from the distribution of events in the ϕ angle alone. This is the case for $B^0 \to K^{*0}\mu^+\mu^-$ when $F_L \neq 0$. If the phyics is non-factorisable then the factorised efficiencies can still be taken from physics-MC but would require a fit to the distribution of events in $(\cos\theta_l, \cos\theta_K, \phi)$, not just a single angular projection.

For phase-space MC the situation is particularly simple as:

$$\frac{d^3\Gamma}{d\cos\theta_l\, d\cos\theta_K\, d\phi} = \frac{1}{8\pi} \quad ,$$

which not only factorises, but is flat in all three angles. In phase-space MC $\varepsilon(\phi)$ can be trivially taken from the distribution of events after reconstruction, the trigger and offline selection. In a bin of q^2, "k", the efficiency is then given by:

$$\varepsilon(q^2, \cos\theta_l, \cos\theta_K, \phi)_k = 8\pi\frac{N_{\mathrm{Sel.};k}}{N_{\mathrm{Gen.};k}} f(\phi)_k f(\cos\theta_l)_k f(\cos\theta_K)_k$$

where e.g.

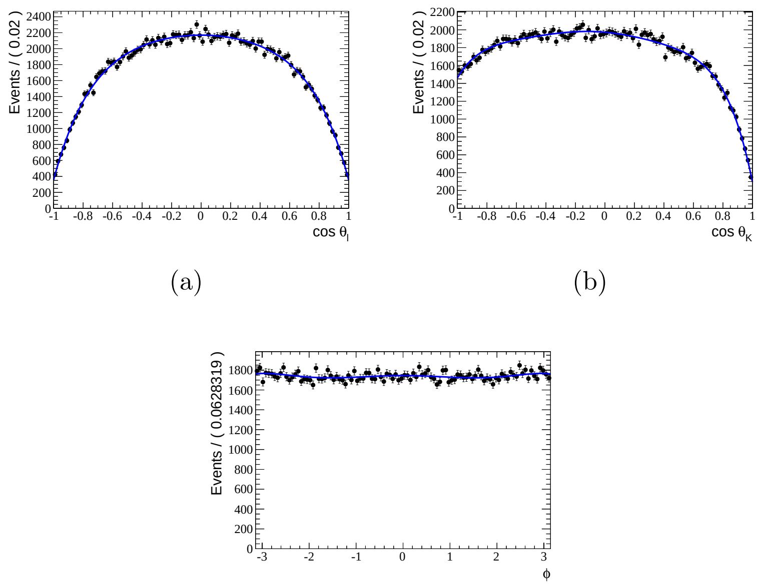
$$f(\phi) = \int\int \frac{d^3\Gamma}{d\cos\theta_l\, d\cos\theta_K\, d\phi}\varepsilon(\cos\theta_l, \cos\theta_K, \phi)\, d\cos\theta_l\, d\cos\theta_K$$

is a probability density function that describes the distribution of events in ϕ after reconstruction, selection etc. The ratio, $N_{\mathrm{Sel.}}/N_{\mathrm{Gen.}}$, of events in a bin of q^2 after selection to the number at generator level is used to normalise the relative efficiency between q^2 bins. The functions $f(\phi)$, $f(\cos\theta_l)$ and $f(\cos\theta_K)$ are normalised such that the integrals:

$$\int_{-\pi}^{\pi} f(\phi)_k d\phi = 1 \ , \quad \int_{-1}^{1} f(\cos\theta_l)_k d\cos\theta_l = 1 \ \text{ and } \ \int_{-1}^{1} f(\cos\theta_K)_k d\cos\theta_K = 1 \ .$$

B.1 Example dsitributions at low- and high-q^2

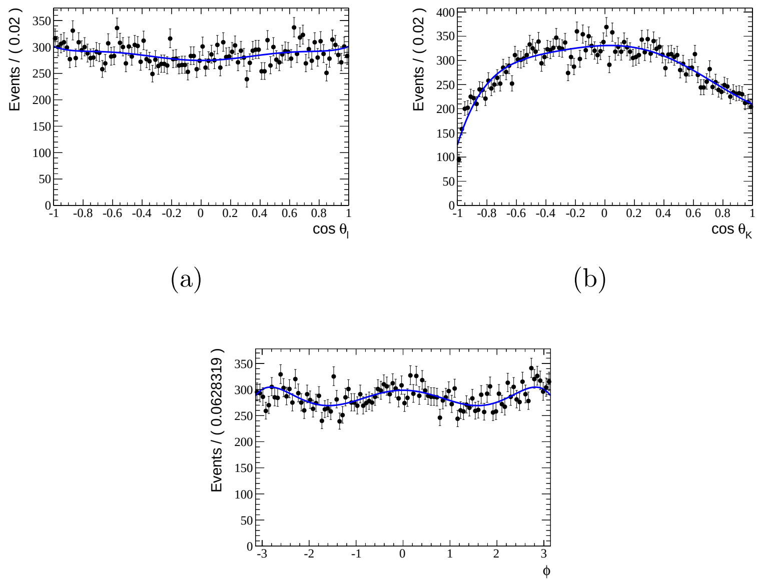
The distribution of events after reconstruction, the trigger and selection in $\cos\theta_l$, $\cos\theta_K$ and ϕ with $1 < q^2 < 1.5\,\text{GeV}^2/c^4$ and $17 < q^2 < 17.5\,\text{GeV}^2/c^4$ are shown in Figs. 61 and 62 respectively. They are fitted with a 6^{th} order Chebychev polynomial, which for $\cos\theta_l$ and ϕ only contains even order terms.



(a) (b)

(c)

Figure 61: One dimensional projections of the distribution of events in $\cos\theta_l$, $\cos\theta_K$ and ϕ in phase-space MC after applying the full selection in the $1 < q^2 < 1.5\,\text{GeV}^2/c^4$ region.

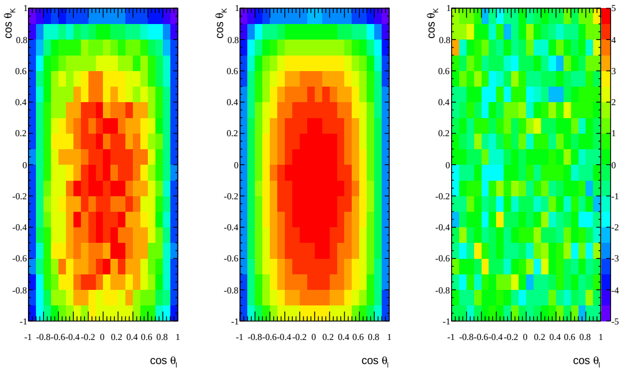


(a) (b)

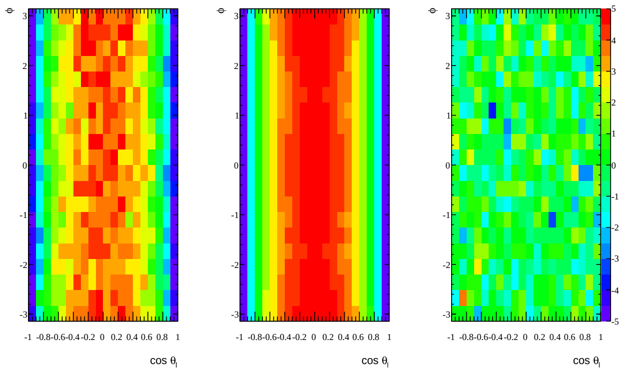
(c)

Figure 62: One dimensional projections of the distribution of events in $\cos\theta_l$, $\cos\theta_K$ and ϕ in phase-space MC after applying the full selection in the $17 < q^2 < 17.5\,\mathrm{GeV}^2/c^4$ region.

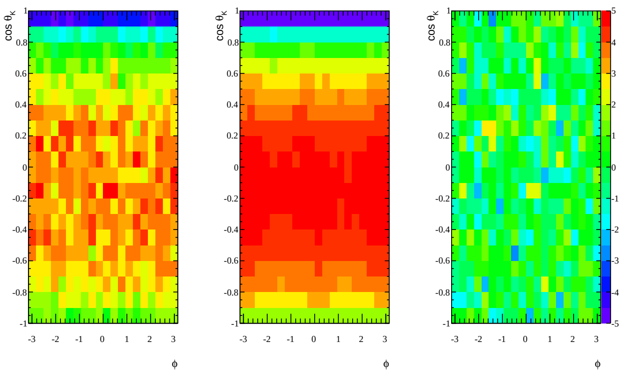
The degree to which the efficiencies factorise is explored for $1 < q^2 < 1.5\,\mathrm{GeV}^2/c^4$ and $17 < q^2 < 17.5\,\mathrm{GeV}^2/c^4$ in Figs. 63 and 64 below. The two dimensional dsitribution of phase-space MC events after reconstruction, the trigger and offline selection is compared to the distribution that would be obtained using toy-MC if it is assumed that the efficiency factorises into three one dimensional distributions in Figs. 61 and 62. Qualitatively, the toy-MC reproduces many of the features seen in the phase-space MC. To try and quantify any potential differences a plot of the difference between phase-space MC and the toy-MC (divided by the error on the phase-space MC) is included. There are no regions where the factorisation is seen to break down. This agrees with the result of the unbinned goodness of fit test pefromed in three dimensions that was reported in Sec. 11.



(a) $\cos\theta_l$ versus $\cos\theta_K$ for $1 < q^2 < 1.5\,\mathrm{GeV}^2/c^4$

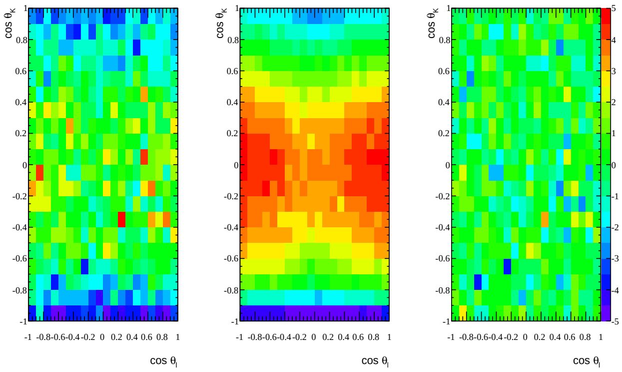


(b) $\cos\theta_l$ versus ϕ for $1 < q^2 < 1.5\,\mathrm{GeV}^2/c^4$

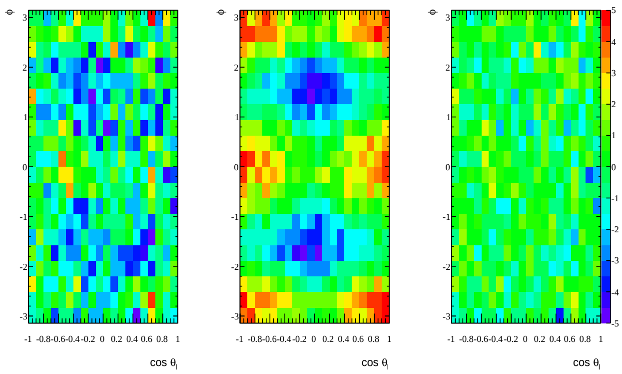


(c) ϕ versus $\cos\theta_K$ for $1 < q^2 < 1.5\,\mathrm{GeV}^2/c^4$

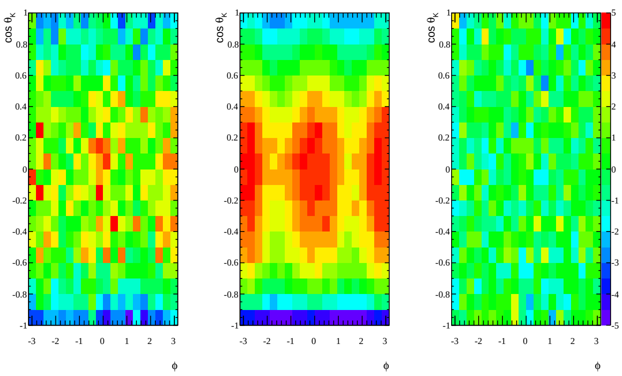
Figure 63: The distribution of events in phase-space MC in the $1 < q^2 < 1.5\,\mathrm{GeV}^2/c^4$ mass region after reconstruction, the trigger and offline selection (left). The corresponding dsitribution in toy-MC if it is assumed that the efficiency can be factorised (centre) and the difference between the toy-MC and phase-space MC, divided by the error on the phase-space MC (right).



(a) $\cos\theta_l$ versus $\cos\theta_K$ for $17 < q^2 < 17.5\,\mathrm{GeV}^2/c^4$



(b) $\cos\theta_l$ versus ϕ for $17 < q^2 < 17.5\,\mathrm{GeV}^2/c^4$

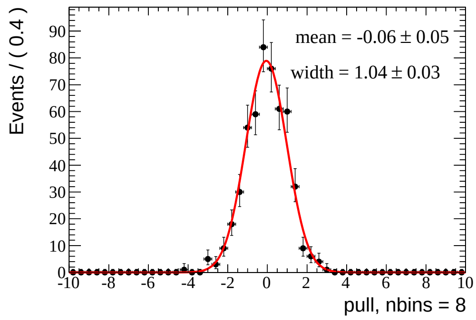


(c) ϕ versus $\cos\theta_K$ for $17 < q^2 < 17.5\,\mathrm{GeV}^2/c^4$

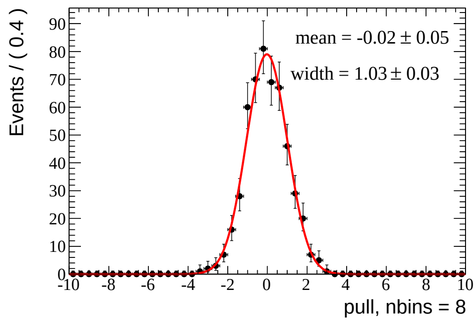
Figure 64: The distribution of events in phase-space MC in the $17 < q^2 < 17.5\,\mathrm{GeV}^2/c^4$ mass region after reconstruction, the trigger and offline selection (left). The corresponding dsitribution in toy-MC if it is assumed that the efficiency can be factorised (centre) and the difference between the toy-MC and phase-space MC, divided by the error on the phase-space MC (right).

B.2 Pull distributions from the factorisation

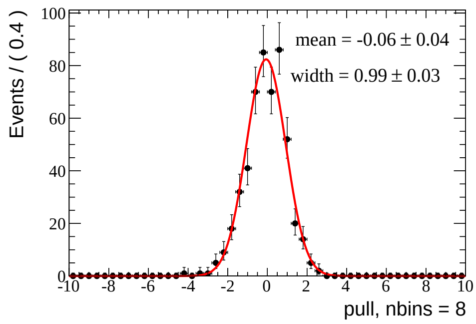
The agreement between the phase-space MC after the application of the reconstruction, stripping, trigger and offline selection and a factorised model is explored further by calculating between the MC and the factorised model in bins of $\cos\theta_l$, $\cos\theta_K$ and ϕ. The "pull" distributions for the J/ψ region, $1 < q^2 < 1.5\,\mathrm{GeV}^2/c^4$ and $17 < q^2 < 17.5\,\mathrm{GeV}^2/c^4$ are shown in Fig. 65. Eight bins have been used in each of the angles, i.e. 512 bins in total appear in the figure. There are no visible outliers and each of the "pull" distributions has a mean of zero and is consistent with having width one.



(a) J/ψ region



(b) $1 < q^2 < 1.5\,\mathrm{GeV}^2/c^4$



(c) $17 < q^2 < 17.5\,\mathrm{GeV}^2/c^4$

Figure 65: The "pull" distribution of the difference between the number of phase-space MC events in a bin of $\cos\theta_l$, $\cos\theta_K$ and ϕ and the number predicted by a factorised model divided by the error on the difference.

B.3 Sensitivity to non-factorisable effects

The level to which we are sensitive to non-factorisable effects in the efficiency distribution has been investigated using toy simulations. First a set of toys was generated according to the factorised efficiency distribution that is seen in the phase-space MC. This dsitribution was then fitted with the same factorised model and the pull distribution was plotted for bins of the dataset with respect to the factorised model. As expected this data set has a well behaved pull distribution with respect to the model, with width of one and a mean of zero.

To simulate a non-factorisable efficiency distribution, a new set of toys was generated. The PDF used to generate the first set of toys was multiplied by a non-factorisable contribution:

$$1 + a\sin(\pi\cos\theta_l)\sin(\pi\cos\theta_K) \tag{20}$$

where a is a scaling factor indicating the size of the non-facotrisable effect. This set of toys was then fitted with the factorised model. For small values of a, the pull dsitribution looks reasonable, but as a increases a large number of bins in the toy dataset are seen to be poorly described by the factorised model. This test was performed for 40 scaling factors between 0 and 1. The number of extreme pulls is significant for $a \geq 0.1$. The value for this test when performed on the phase space simulation data-set used to obtain the efficiency PDF is 5.

C Comparison of B^0 and $\overline{B}^0$ distributions for $B^0 \rightarrow K^{*0} J/\psi$

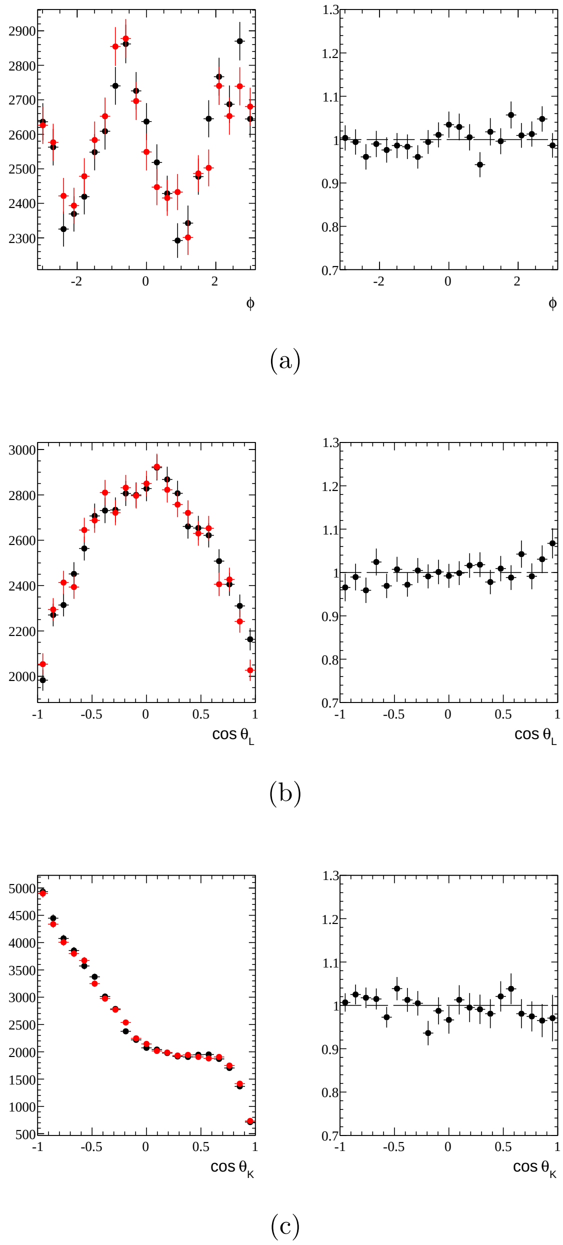
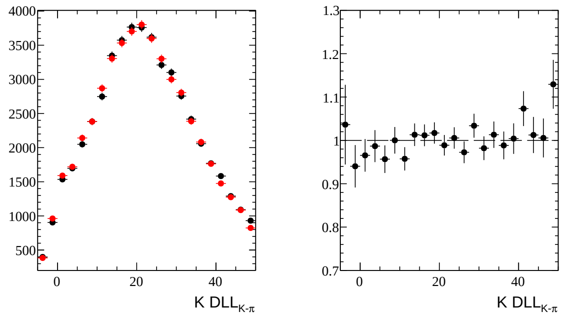
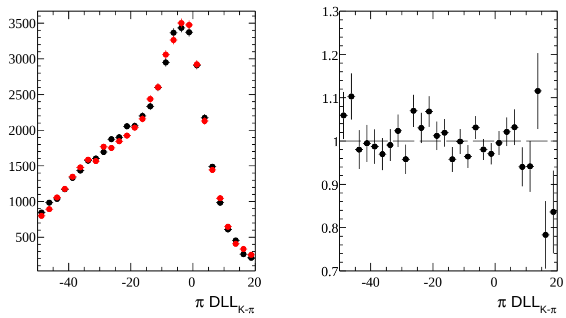


Figure 66: A comparison of the angular distribution of B^0 and $\overline{B}^0$ decays for the channel $B^0 \rightarrow K^{*0} J/\psi$.



(a)



(b)

Figure 67: A comparison of the kaon and pion DLL distributions for B^0 and $\overline{B}^0$ decays for the channel $B^0 \to K^{*0} J/\psi$.

D Lepton mass terms

If the lepton mass is not neglected then extra terms are introduced into the angular distribution and the I_i terms can be writen as:

$$
\begin{aligned}
\frac{1}{\Gamma}I_1^S &= \left(\frac{3}{4}(1 - F_{\rm L}) \times (1 - \frac{4m_\mu^2}{3q^2}) + \frac{1}{\Gamma}\frac{4m_\mu^2}{q^2}\Re\left(A_{\perp\ L}A^*_{\perp\ R} + A_{\parallel\ L}A^*_{\parallel\ R}\right)\right)\sin^2\theta_K \\
\frac{1}{\Gamma}I_1^C &= \left(F_{\rm L} + \frac{1}{\Gamma}\frac{4m_\mu^2}{q^2} \times \left(|A_t|^2 + 2\Re(A_{0\ L}A^*_{0\ R})\right)\right)\cos^2\theta_K \\
\frac{1}{\Gamma}I_2^S &= \frac{1}{4}(1 - F_{\rm L})(1 - \frac{4m_\mu^2}{q^2})\sin^2\theta_K \\
\frac{1}{\Gamma}I_2^C &= -F_{\rm L}(1 - \frac{4m_\mu^2}{q^2})\cos^2\theta_K \\
\frac{1}{\Gamma}I_3 &= \frac{1}{2}(1 - F_{\rm L})A_{\rm T}^2\left(1 - \frac{4m_\mu^2}{q^2}\right) \times \sin^2\theta_K \\
\frac{1}{\Gamma}I_6 &= 2A_T^{Re}(1 - F_{\rm L})\sqrt{(1 - \frac{4m_\mu^2}{q^2})} \times \sin^2\theta_K \\
\frac{1}{\Gamma}I_9 &= \frac{1}{2}(1 - F_{\rm L})A_T^{Im}\left(1 - \frac{4m_\mu^2}{q^2}\right) \times \sin^2\theta_K
\end{aligned}
$$

with the standard definitions for the parameters $F_{\rm L}$, $A_{\rm T}^2$ A_T^{Im} and A_T^{Re}. At low-q^2 where these additional terms can be significant, if the amplitudes coming from QCD factorisation, with soft form-factors are used[3], then I_1^S and I_1^C can be simplified - without requiring extra parameters in the fit.

Starting with the $\frac{1}{\Gamma}I_1^C$ term, one has:

$$\frac{|A_t|^2 + 2\Re(A_{0\ L}A^*_{0\ R})}{\Gamma} = F_{\rm L} \times \frac{|A_t|^2 + 2\Re(A_{0\ L}A^*_{0\ R})}{|A_0|^2} \tag{21}$$

and using the expressions for the amplitudes in terms of the soft form-factors in Ref. [26]:

$$\frac{|A_t|^2 + 2\Re(A_{0\ L}A^*_{0\ R})}{|A_0|^2} = 1 \quad . \tag{22}$$

Thus:

[3]These assumption are assumed to hold to $\mathcal{O}(\Lambda/m_b) \sim 10\%$ for small values of q^2

$$\frac{1}{\Gamma} I_1^C \;= F_{\mathrm{L}} \times \left(1 + \frac{4m_\mu^2}{q^2}\right) \cos^2\theta_K \quad . \tag{23}$$

$\frac{1}{\Gamma} I_1^S$ term is slightly more complicated:

$$\begin{aligned} \frac{\Re(A_{\perp\ L} A^*_{\perp\ R} + A_{\parallel\ L} A^*_{\parallel\ R})}{\Gamma} =& (1 - F_{\mathrm{L}}) \times \frac{\Re(A_{\perp\ L} A^*_{\perp\ R} + A_{\parallel\ L} A^*_{\parallel\ R})}{|A_\parallel|^2 + |A_\perp|^2} \\ =& \frac{1}{2}(1 - F_{\mathrm{L}}) \times \left[1 - f(\mathcal{C}_7^{eff(\prime)}, \mathcal{C}_9^{eff(\prime)}, \mathcal{C}_{10}^{eff(\prime)})\right] \end{aligned} \tag{24}$$

where:

$$\begin{aligned} f(\mathcal{C}_7^{eff(\prime)}, \mathcal{C}_9^{eff(\prime)}, \mathcal{C}_{10}^{(\prime)}) = 2(|\mathcal{C}_{10}|^2 + |\mathcal{C}_{10}'|^2) / \Big[\quad & |\mathcal{C}_9^{eff}|^2 + |\mathcal{C}_9^{eff\prime}|^2 + |\mathcal{C}_{10}|^2 + |\mathcal{C}_{10}'|^2 \;\; + \\ & 2\frac{m_b m_B}{q^2}\left(\mathcal{C}_9^{eff}\mathcal{C}_7^{eff*} + \mathcal{C}_9^{eff*}\mathcal{C}_7^{eff}\right) \;\; + \\ & 2\frac{m_b m_B}{q^2}\left(\mathcal{C}_9^{eff\prime}\mathcal{C}_7^{eff\prime*} + \mathcal{C}_9^{eff\prime*}\mathcal{C}_7^{eff\prime}\right) \;\; + \\ & 4\frac{m_b^2 m_B^2}{q^4}\left(|\mathcal{C}_7^{eff}|^2 + |\mathcal{C}_7^{eff\prime}|^2\right) \quad \Big] \end{aligned}$$

and then:

$$\frac{1}{\Gamma} I_1^S = \frac{3}{4}(1 - F_{\mathrm{L}}) \times \left[1 + \frac{4m_\mu^2}{3q^2} - \frac{8m_\mu^2}{3q^2} f(\mathcal{C}_7^{eff(\prime)}, \mathcal{C}_9^{eff(\prime)}, \mathcal{C}_{10}^{(\prime)})\right]$$

If $f(\mathcal{C}_7^{eff(\prime)}, \mathcal{C}_9^{eff(\prime)}, \mathcal{C}_{10}^{(\prime)})$ is small, this simplifies to:

$$\frac{1}{\Gamma} I_1^S \simeq \frac{3}{4}(1 - F_{\mathrm{L}}) \times \left[1 + \frac{4m_\mu^2}{3q^2}\right]$$

For this to be true:

$$\begin{aligned} & 2\frac{m_b m_B}{q^2}\left(\mathcal{C}_9^{eff}\mathcal{C}_7^{eff*} + \mathcal{C}_9^{eff*}\mathcal{C}_7^{eff} + \mathcal{C}_9^{eff\prime}\mathcal{C}_7^{eff\prime*} + \mathcal{C}_9^{eff\prime*}\mathcal{C}_7^{eff\prime}\right) \;\; + \\ & \qquad 4\frac{m_b^2 m_B^2}{q^4}\left(|\mathcal{C}_7^{eff}|^2 + |\mathcal{C}_7^{eff\prime}|^2\right) + |\mathcal{C}_9^{eff}|^2 + |\mathcal{C}_9^{eff\prime}|^2 \gg |\mathcal{C}_{10}|^2 + |\mathcal{C}_{10}'|^2 \end{aligned}$$

which will tend to be true for $q^2 \leq 1$ where the contribution from $\mathcal{C}_7^{(\prime)}$ dominates, i.e. $4m_b^2 m_B^2(|\mathcal{C}_7|^2 + |\mathcal{C}_7'|^2)/q^4$ is large compared to $|\mathcal{C}_{10}|^2$. $|\mathcal{C}_7|^2 + |\mathcal{C}_7'|^2$ is known to $\sim 10\%$ from $b \to s\gamma$.

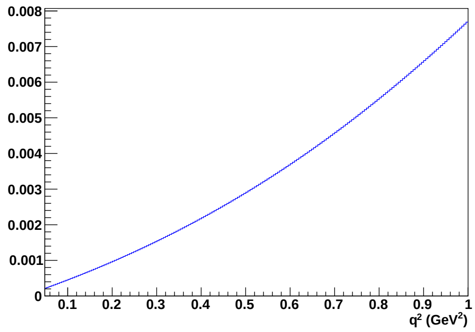


Figure 68: Variation of the function $8m_\mu^2 f(\mathcal{C}_7^{eff(\prime)}, \mathcal{C}_9^{eff(\prime)}, \mathcal{C}_{10}^{(\prime)})/3q^2$ with q^2.

1921 Using the SM values for the Wilson coefficients and neglecting $\mathcal{C}_7^{eff'}$, $\mathcal{C}_9^{eff'}$
1922 and $\mathcal{C}_{10}^{eff'}$ with respect to $\mathcal{C}_7^{eff}$, one can draw the variation of:

$$\frac{8m_\mu^2}{3q^2} f(\mathcal{C}_7^{eff(\prime)}, \mathcal{C}_9^{eff(\prime)}, \mathcal{C}_{10}^{(\prime)})$$

1923 as a function of q^2. It is shown in Fig 68.
1924 In summary, no additional parameters are introduced but kinematical
1925 factors, that depend on m_μ^2/q^2, appear in front of the usual terms.

E Threshold Terms

E.1 Testing the correction procedure.

In order to test its validity, the correction procedure has been applied to a large statistics MC sample. The events have been generated according to the SM predictions for the physics parameters of interest. The first q^2 bin, between 0.1 and $2\,\mathrm{GeV}^2/c^4$ is divided into 19 sub-bins of width $0.1\,\mathrm{GeV}^2/c^4$. In each of these bins, two fits are performed:

- the first fit neglecting the threshold terms completely;
- the second fit includes threshold terms.

In both cases the q^2 variation over the bin is neglected. In the second case this amounts to treating x as a constant over the sub-bin. The impact of neglecting the threshold terms can be clearly seen in Fig 69, which shows the angular distribution of simulated events with $0.1 < q^2 < 0.2\,\mathrm{GeV}^2/c^4$. The $\cos\theta_l$ distribution is only correctly described if the threshold terms are taken into account.

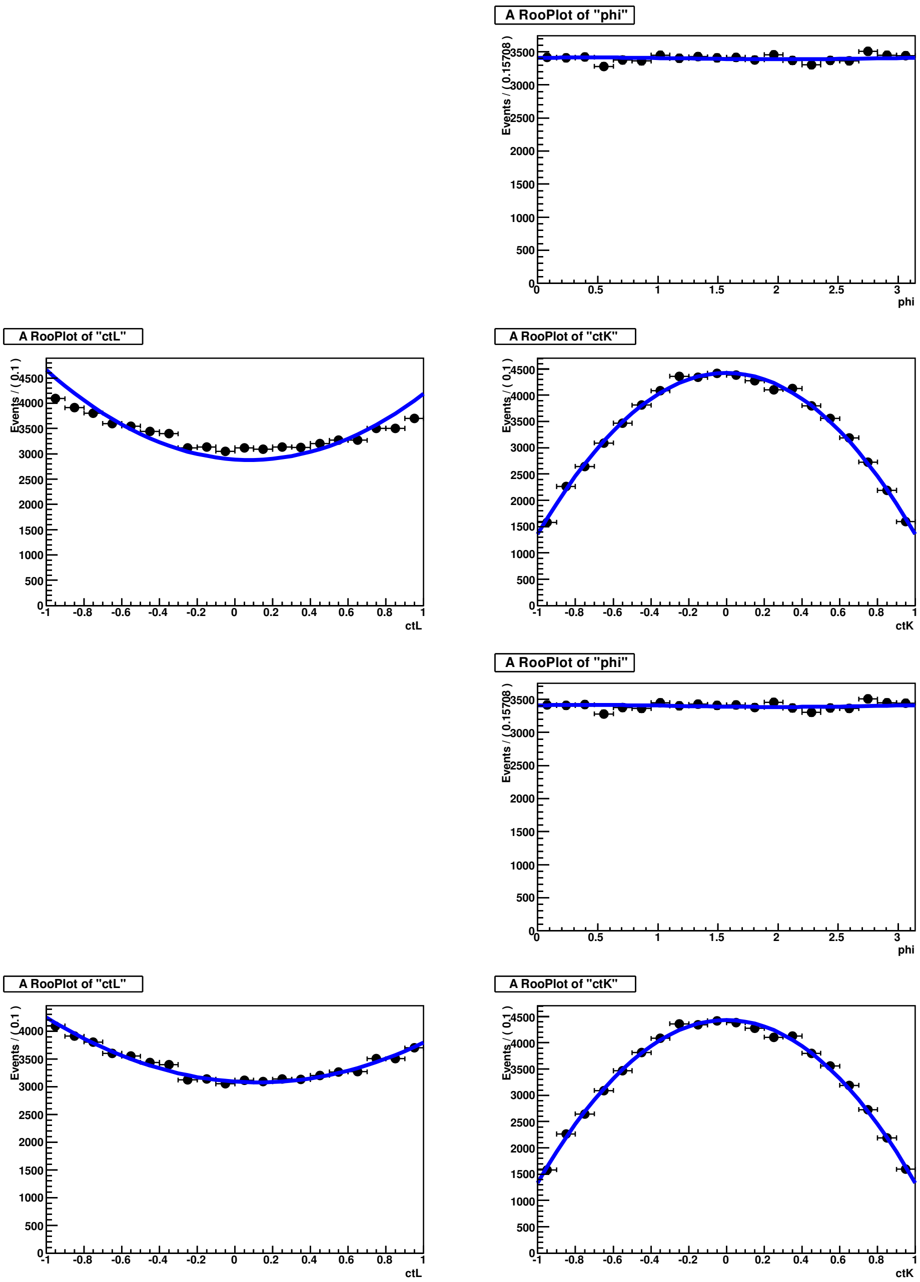


Figure 69: Fit of the angular distributions in simulation with $0.1 < q^2 < 0.2\,\mathrm{GeV}^2/c^4$ with a pdf without threshold terms (three top plots) and with threshold terms (three bottom plots). The $\cos\theta_l$ distribution is clearly not well fitted in the first case.

The results of the fits for each observable in the 19 small bins, i.e. as a function of q^2, are shown on Figs. 70 and 71 for the fits without and with threshold terms respectively. As expected, the ratio of the two fit results approaches one as q^2 becomes large, Fig. 72.

In the MC, where the statistics is large, the true value of the physics parameters over the $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$ bin can be obtained by averaging the results of the fits to the 19 sub-bins, taking into account the threshold terms in the fits (the assumption here is that the q^2 variation over the sub-bins is negligible). The averages are calculated as follows:

$$< F_L >= \frac{\sum_{i=1}^{nbins} F_{L,i} N_i}{\sum_{i=1}^{nbins} N_i} \tag{25}$$

$$< A_T^2 >= \frac{\sum_{i=1}^{nbins} A_{T,i}^2 N_i (1 - F_{L,i})}{\sum_{i=1}^{nbins} N_i (1 - F_{L,i})} \tag{26}$$

$$< A_T^{Im} >= \frac{\sum_{i=1}^{nbins} A_{T,i}^{Im} N_i (1 - F_{L,i})}{\sum_{i=1}^{nbins} N_i (1 - F_{L,i})} \tag{27}$$

$$< A_T^{Re} >= \frac{\sum_{i=1}^{nbins} A_{T,i}^{Re} N_i (1 - F_{L,i})}{\sum_{i=1}^{nbins} N_i (1 - F_{L,i})} \tag{28}$$

and are listed in Table 48, third row.

The results of the fit to the whole $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$ bin without taking into account the threshold terms in the PDF are also shown: on the first row without applying the correction procedure and on the second row applying the correction procedure. The values in the second row of table 48 are in general in good agreement with the reference values in the third row.

The values of the corrections, evaluated with formulas 8 and 9 using the 400 k SM MC candidates with $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$, are shown in table 47. Three different values of the parameter a of $F_L(q^2)$, defined in eq. 7, have been considered. The results for $a = 0.66$ and $a = 1.5$ are shown on tables 49 and 50 respectively.

We can notice that assuming a linear behavior for A_T^{Re} allows to get a correction which gives a more reliable result. The differences in the values for A_T^2 are due to statistical fluctuations, which have a large impact here since the generation value for A_T^2 is about zero. The same analysis for a non SM MC, having a generation value for A_T^2 different from zero, gives a good agreement also for the value of A_T^2, as can be seen in Tables 52 and 53.

	$a = 0.66$	$a = 1$	$a = 1.5$
Correction on A_T^2	1.24	1.26	1.28
Correction on $err(A_T^2)$	1.22	1.24	1.26
Correction on A_T^{Re}	1.16	1.17	1.18
Correction on A_T^{Re} (linear approx)	1.08	1.08	1.09
Correction on $err(A_T^{Re})$	1.15	1.16	1.17

Table 47: Values of the corrections evaluated with formulas 8 and 9 using 400 k SM MC candidates in the range $(0.1 - 2)\,\mathrm{GeV}^2/c^4$. Three different values of the parameter a of $F_L(q^2)$, defined in Eq. 7, have been considered.

SM MC a=1					
	F_L	A_T^{Re}	A_T^{Re} (linear approx.)	A_T^2	A_T^{Im}
Bin not corrected	0.4469±0.0011	-0.2783±0.0029	-	0.0004±0.0060	0.0090±0.0060
Bin corrected	0.4468±0.0011	-0.3244±0.0033	-0.3019±0.0033	0.0006±0.0075	0.0114±0.0075
Average of results in small bins	0.4477±0.0011	-0.2956±0.0033	-	0.0030±0.0077	0.0115±0.0078

Table 48: Results of the validation of the correction procedure on high statistics SM MC, assuming a=1

SM MC a=0.66					
	F_L	A_T^{Re}	A_T^{Re} (linear approx.)	A_T^2	A_T^{Im}
Bin corrected	0.4469±0.0011	-0.3217±0.0033	-0.3000±0.0033	0.0006±0.0074	0.0112±0.0074
Average of results in small bins	0.4477±0.0011	-0.2956±0.0033	-	0.0030±0.0077	0.0115±0.0078

Table 49: Results of the validation of the correction procedure on high statistics SM MC, assuming a=0.66

SM MC a=1.5					
	F_L	A_T^{Re}	A_T^{Re} (linear approx.)	A_T^2	A_T^{Im}
Bin corrected	0.4468±0.0011	-0.3273±0.0034	-0.3039±0.0039	0.0006±0.0076	0.0115±0.0076
Average of results in small bins	0.4477±0.0011	-0.2956±0.0033		0.0030±0.0077	0.0115±0.0078

Table 50: Results of the validation of the correction procedure on high statistics SM MC, assuming a=1.5

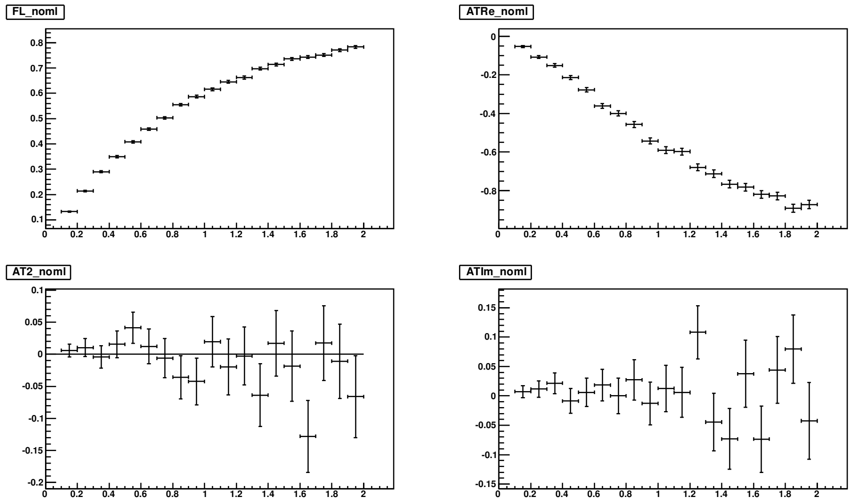


Figure 70: Results of the fits in small bins of $0.1\,\mathrm{GeV}^2/c^4 2$ width for the high statistics SM MC, not taking into account the threshold terms.

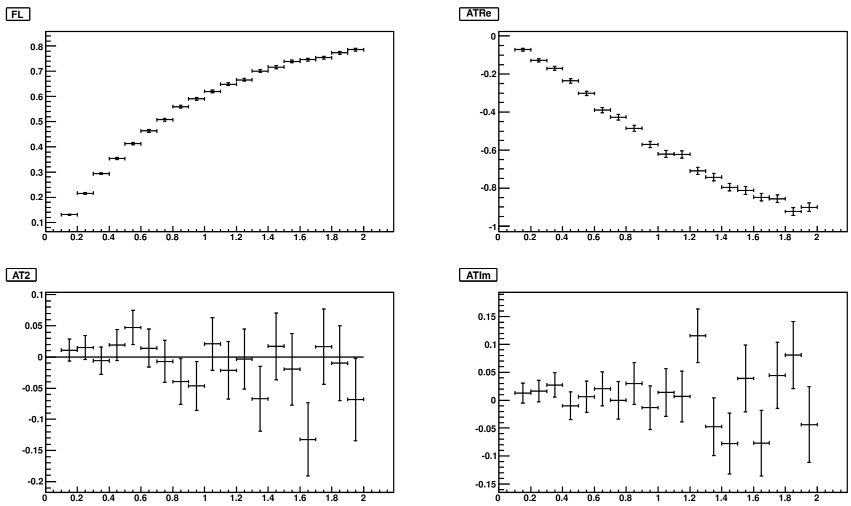


Figure 71: Results of the fits in small bins of $0.1\,\mathrm{GeV}^2/c^4 2$ width for the high statistics SM MC, taking into account the threshold terms.

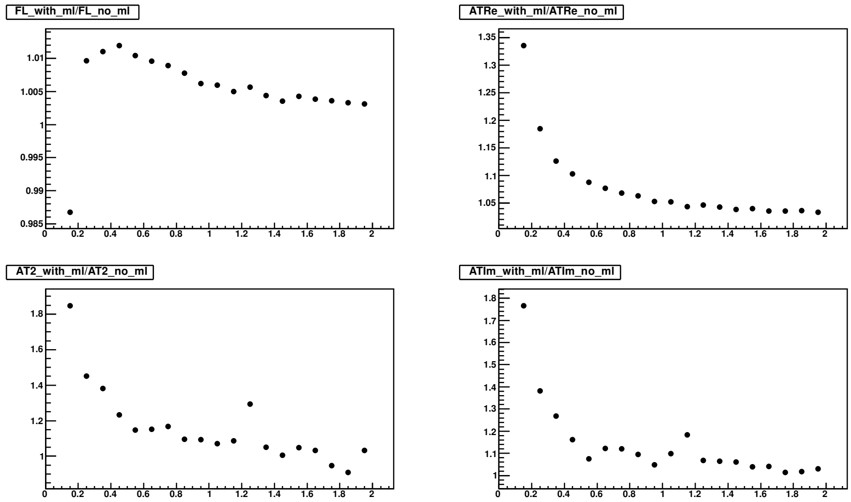


Figure 72: Ratio of the results of the fits in small bins of $0.1\,\mathrm{GeV}^2/c^4$ width taking into account the threshold terms over the results not taking into accountthem for the high statistics SM MC.

In order to test the precision to which the correction factors can be determined, the high statistics MC sample has been divided in 1832 samples, each containing 143 signal events as expected in $1\,\mathrm{fb}^{-1}$ in the range $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$. The corrections have been evaluated for each of these toy samples and the results are shown in Fig. 73. The distributions of the corrections are fit with a Gaussian function, and the results are reported on Table 51 for the mean and the sigma. We can see that the corrections are determined with an uncertainty lower than 1%.

Fit with no threshold terms (a=1)		
Parameter	m	σ
A_T^2	1.259	0.016
$err(A_T^2)$	1.240	0.014
A_T^{Re}	1.1662	0.0099
A_T^{Re} (linear approx)	1.0851	0.0043
$err(A_T^{Re})$	1.1583	0.0092

Table 51: Results of the Gaussian fit to the distributions of the corrections obtained from 1832 MC toys based on SM MC. Each toy has a statistic corresponding 143 signal events as expected in $1\,\text{fb}^{-1}$ in the range $0.1 < q^2 < 2\,\text{GeV}^2/c^4$.

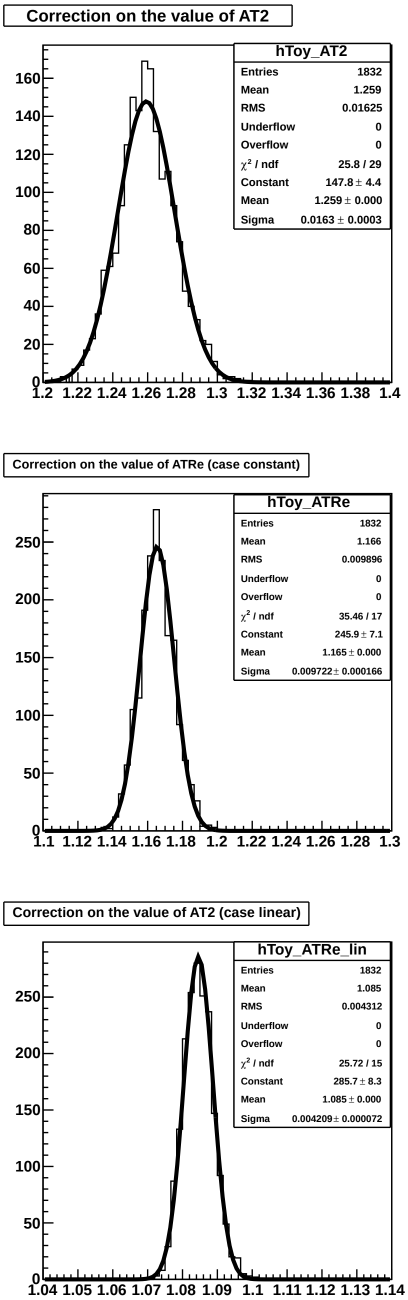


Figure 73: Distributions of the corrections obtained from 1832 MC toys based on SM MC. Each toy has a statistic corresponding 143 signal candidates as expected in $1\,\mathrm{fb}^{-1}$ of data. The distribution is fitted with a Gaussian function.

	$a = 0.66$	$a = 1$	$a = 1.5$
Correction on A_T^2	1.23	1.24	1.26
Correction on $err(A_T^2)$	1.21	1.22	1.24
Correction on A_T^{Re}	1.15	1.16	1.17
Correction on A_T^{Re} (linear approx)	1.07	1.08	1.09
Correction on $err(A_T^{Re})$	1.14	1.15	1.16

Table 52: Values of the corrections evaluated with formulas 8 and 9 using 70 k events of non SM MC in the range $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$. Three different values of the parameter a of $F_L(q^2)$, defined in eq. 7, have been considered.

Non-SM MC, a=1					
	F_L	A_T^{Re}	A_T^{Re} (linear approx.)	A_T^2	A_T^{Im}
Bin not corrected	0.5305±0.0038	-0.3177±0.0108	-	0.1250±0.0227	-0.0168±0.0228
Bin corrected	0.5305±0.0038	-0.3673±0.0124	-0.3428±0.0124	0.1555±0.0279	-0.0209±0.0279
Average of results in small bins	0.5348±0.0036	-0.3433±0.0122	-	0.1613±0.0291	-0.0252±0.0290

Table 53: Results of the validation of the correction procedure on high statistics non-SM MC, assuming a=1

E.2 Cross-checking the assumption on the dependence of F_L from q^2.

As a cross-check we also computed the correction assuming a linear behavior for F_L , i.e. using the following expression instead of that in equation 7:

$$F_L(q_i^2) = bq_i^2 \tag{29}$$

Table 54 shows the size of the corresponding correction factors for the three value of b. The measured value of b on data, shown on figure 74, is $b = 0.29 \pm 0.08$.

	$b = 0.21$	$b = 0.29$	$b = 0.37$
Correction on A_T^2, S_3, A_T^{Im}, A_{Im}	1.18	1.21	1.24
Correction on $err(A_T^2)$, $err(S_3)$, $err(A_T^{Im})$, $err(A_{Im})$	1.17	1.19	1.22
Correction on A_T^{Re}, A_{FB}	1.12	1.13	1.15
Correction on A_T^{Re}, A_{FB} (linear approx)	1.06	1.07	0.09
Correction on $err(A_T^{Re})$, $err(A_{FB})$	1.11	1.13	1.15

Table 54: Values of the corrections evaluated with Eq. 15 and 16 using 254 events of data in the range $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$, assuming linear behavior for F_L as in Eq. 29. Three different values of the parameter b of $F_L(q^2)$, defined in Eq. 29, have been considered.

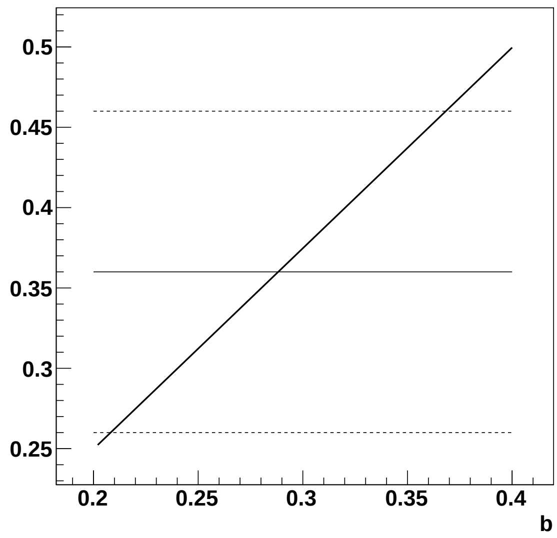


Figure 74: The curve represent the values of $< F_L >$ as function of b as calculated on data assuming linear behavior for F_L as in equation 29. The horizontal lines represent the measured value of F_L and its error. The intersection with the curve gives the measurement of $b = 0.29 \pm 0.08$.

F S-wave extraction

F.1 Validation of the S-wave extraction with $B^0 \to K^{*0} J/\psi$

To determine the S-wave parameters in data, we perform a simultaneous fit in the two mass regions: above and below the K^{*0} mass.

The signal is described by the angular Pdf including the extra terms due to the S-wave, as discussed in Sec. 16, while the B^0-mass Pdf is identical to the one used in the main fit. In the simultaneous fit all parameters of the two Pdfs, apart for the value of A_s^+ and A_s^- and the signal fraction, are shared. While in the main fit the S-wave parameters are fixed to zero, in this more complex fit an iterative procedure is used. The fit is performed as follow: F_{S} is first fixed to 0, while A_S^+ and A_S^- are free to float. After the first fit, F_{S} is computed using Eq. 4 and fixed to this new value. A second fit is performed to determine again A_S^+ and A_S^-, so a new value of F_{S} is obtained. We found that F_{S} varies slightly between the two fits, so there is no need to iterate again. This procedure assumes implicitly that the acceptance corrections calculated for the full sample can be used for both the $K\pi$ mass regions, i.e. that the acceptance has a small dependence on the $K\pi$-mass, which is reasonable to expect.

The iterative fit to extract the S-wave has been validated on $B^0 \to K^{*0} J/\psi$ events. The results are shown in Table 55. After the second iteration, the F_{S} value is found to be 0.0835 ± 0.0024, consistent with expectations. The value obtained using A_S^+ and A_S^- from the first iteration was $F_{\mathrm{S}} = 0.0838$, which shows how quickly this procedure converges for the $B^0 \to K^{*0} J/\psi$.

The projection of the four fitted quantities for the two $K\pi$ mass regions are shown on Figures 75 and 76.

Observable	Fit result
A_T^{Re}	0.010 ± 0.007
F_L	0.567 ± 0.002
A_T^2	0.050 ± 0.017
A_T^{Im}	-0.390 ± 0.017
A_S^+	-0.054 ± 0.004
A_S^-	-0.288 ± 0.004

Table 55: Fit results on $B^0 \to J/\psi\ K^{*0}$ including the S-wave and exploiting the phase information.

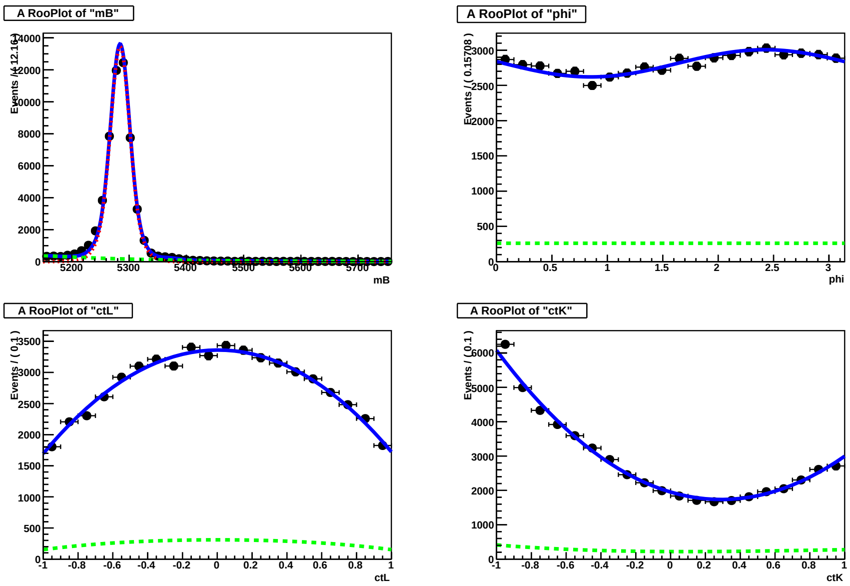


Figure 75: 1D projections of the four fitted quantities for the $B^0 \to J/\psi\ K^{*0}$ dataset with $M(K\pi) < M(K^{*0})$. The fitted pdf (blue), the background only pdf (green) are overlaid.

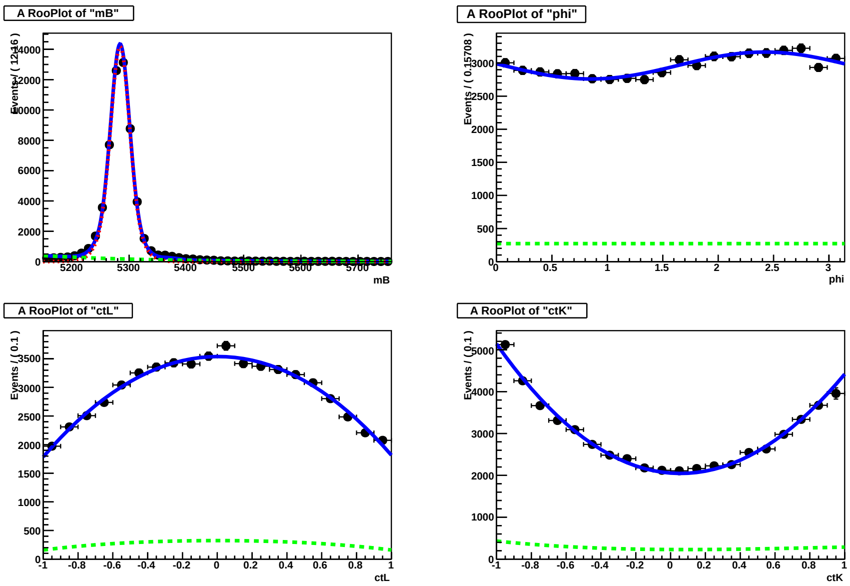


Figure 76: 1D projections of the four fitted quantities for the $B^0 \to J/\psi\ K^{*0}$ dataset with $M(K\pi) > M(K^{*0})$. The fitted pdf (blue), the background only pdf (green) are overlaid.

For comparison, a simple fit with $A_{\rm S}$ and $F_{\rm S}$ as free parameters is performed on $B^0 \to K^{*0} J/\psi$ events. The results are shown in Table 56. The $A_{\rm S}$ value can be compared with the mean of A_S^+ and A_S^- from Table 55. The fit results are compatible with the ones of Table 55 but the method exploiting the phase change gives an error on $F_{\rm S}$ smaller by a factor ~ 3.

Observable	Fit result
A_T^{Re}	0.010 $\pm$ 0.007
F_L	0.566 $\pm$ 0.003
A_T^2	0.052 $\pm$ 0.017
A_T^{Im}	-0.382 $\pm$ 0.017
F_S	0.0771 $\pm$ 0.0062
A_S	-0.169 $\pm$ 0.003

Table 56: Fit results on $B^0 \to J/\psi\ K^{*0}$ including the S-wave, fitting directly $F_{\rm S}$ and $A_{\rm S}$.

We have also tested the method to extract the S-wave splitting the $B^0 \to J/\psi\ K^{*0}$ dataset in 152 files of 1000 events. The value obtained for $F_{\rm S}$ and its error after the second fit are shown on Figure 77 and 78, it demonstrates that this method gives reliable results on small samples.

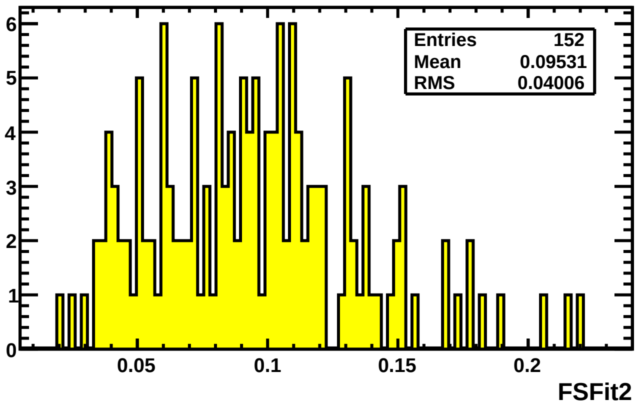


Figure 77: F_{S} values obtained from fits on $B^0 \to J/\psi\ K^{*0}$ data samples of 1000 events.

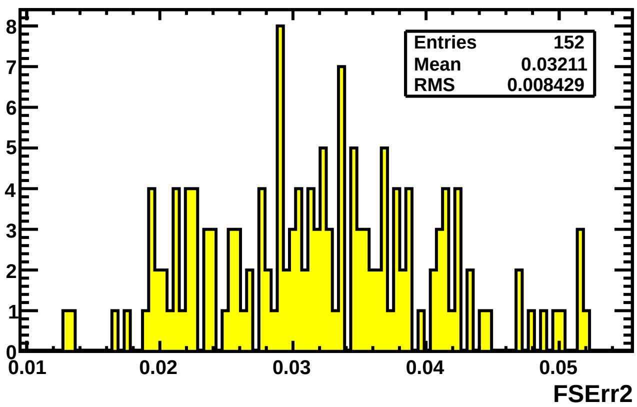


Figure 78: F_{S} errors obtained from fits on $B^0 \to J/\psi\ K^{*0}$ data samples of 1000 events.

Using the $B^0 \to K^{*0} J/\psi$ events, it was also checked how the calculated values of F_{S} depends on the assumptions: the S-wave was parametrised as varying by $\pm 20\%$ over $\pm 100\,\mathrm{MeV}/c^2$ instead of being taken as constant. The Breit Wigner was parametrised as a P-wave relativistic Breit Wigner instead of the simple BW and central value and sigma of the BW were varied within their errors, resulting among others from different background subtraction . All these variations resulted in $< F_{\mathrm{S}} >$ variations by less than 10%. This 10%

is much smaller than the statistical error on F_s obtained with $B^0 \to K^{*0}\mu^+\mu^-$ events.

F.2 Fit distribution for the extraction of a $K^+ \; \pi^-$ system S-wave in $B^0 \to K^{*0}\mu\mu$

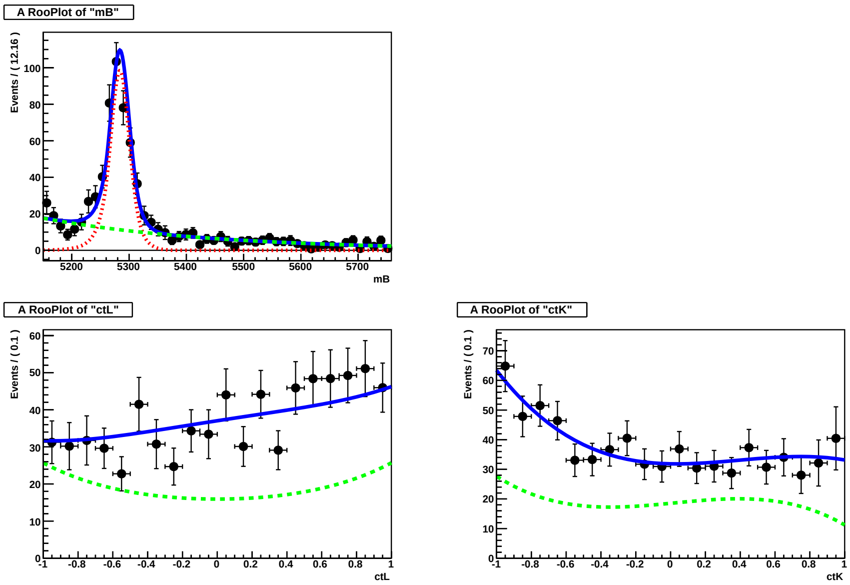


Figure 79: 1D projections of the four fitted quantities for the $B^0 \to K^{*0}\mu\mu$ dataset with $M(K\pi) < M(K^{*0})$ in the q^2 region from 1 to 19 GeV^2/c^4. The fitted pdf (blue), the background only pdf (green) are overlaid.

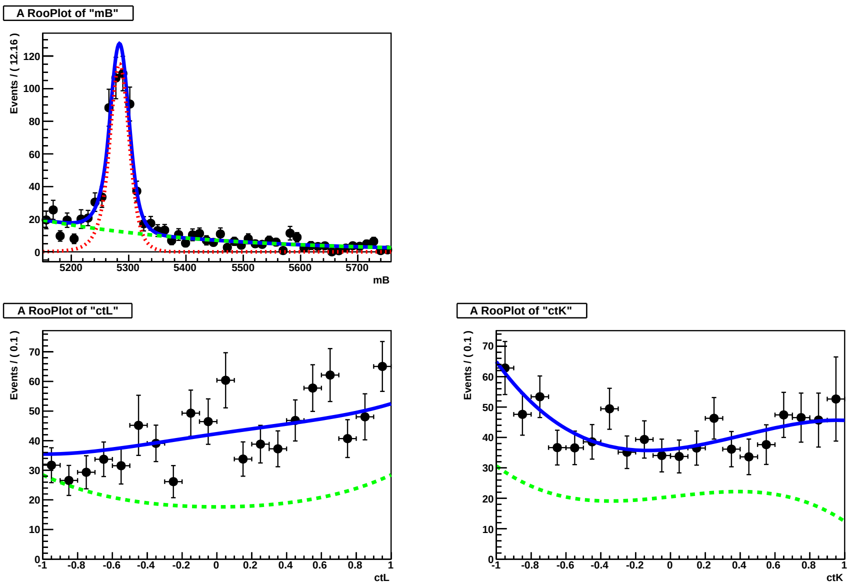


Figure 80: 1D projections of the four fitted quantities for the $B^0 \to K^{*0}\mu\mu$ dataset with $M(K\pi) > M(K^{*0})$ in the q^2 region from 1 to 19 GeV^2/c^4. The fitted pdf (blue), the background only pdf (green) are overlaid.

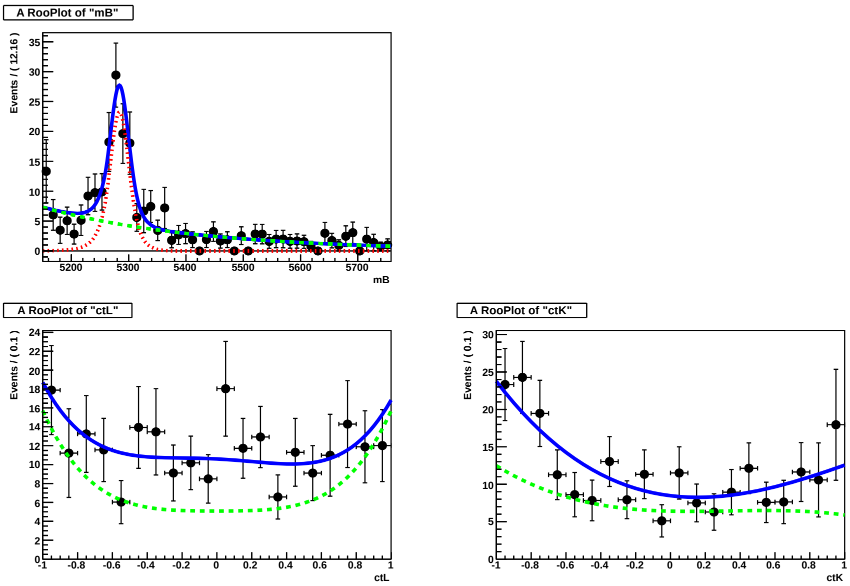


Figure 81: 1D projections of the four fitted quantities for the $B^0 \to K^{*0}\mu\mu$ dataset with $M(K\pi) < M(K^{*0})$ in the q^2 region from 1 to 6 GeV^2/c^4. The fitted pdf (blue), the background only pdf (green) are overlaid.

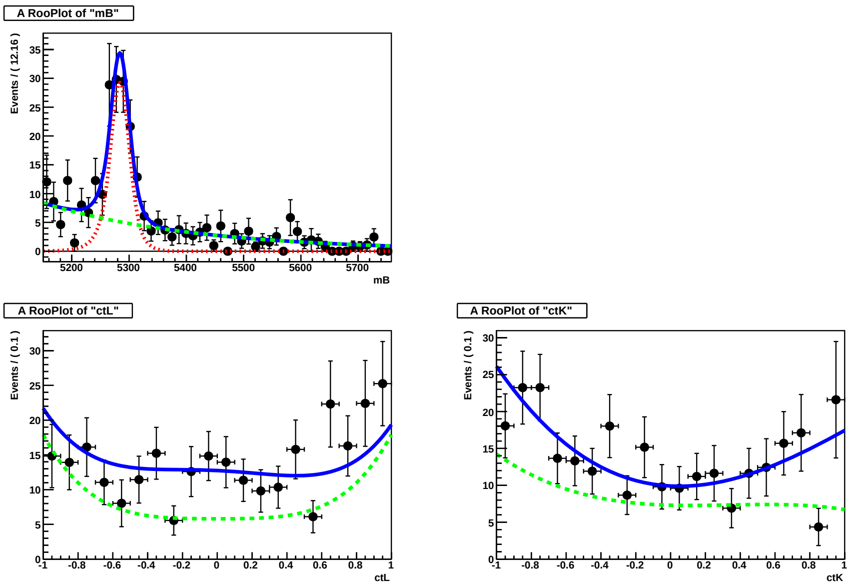


Figure 82: 1D projections of the four fitted quantities for the $B^0 \to K^{*0}\mu\mu$ dataset with $M(K\pi) > M(K^{*0})$ in the q^2 region from 1 to 6 GeV^2/c^4. The fitted pdf (blue), the background only pdf (green) are overlaid.

G Profile Likelihood

G.1 Profile-likelihoods

The 1D likelihood scans can be found at this location (`http://www.hep.ph.ic.ac.uk/~cp309/FCandMINOS\_Results/L1/`)

The 2D likelihood scans are shown in Figs. 83-89.

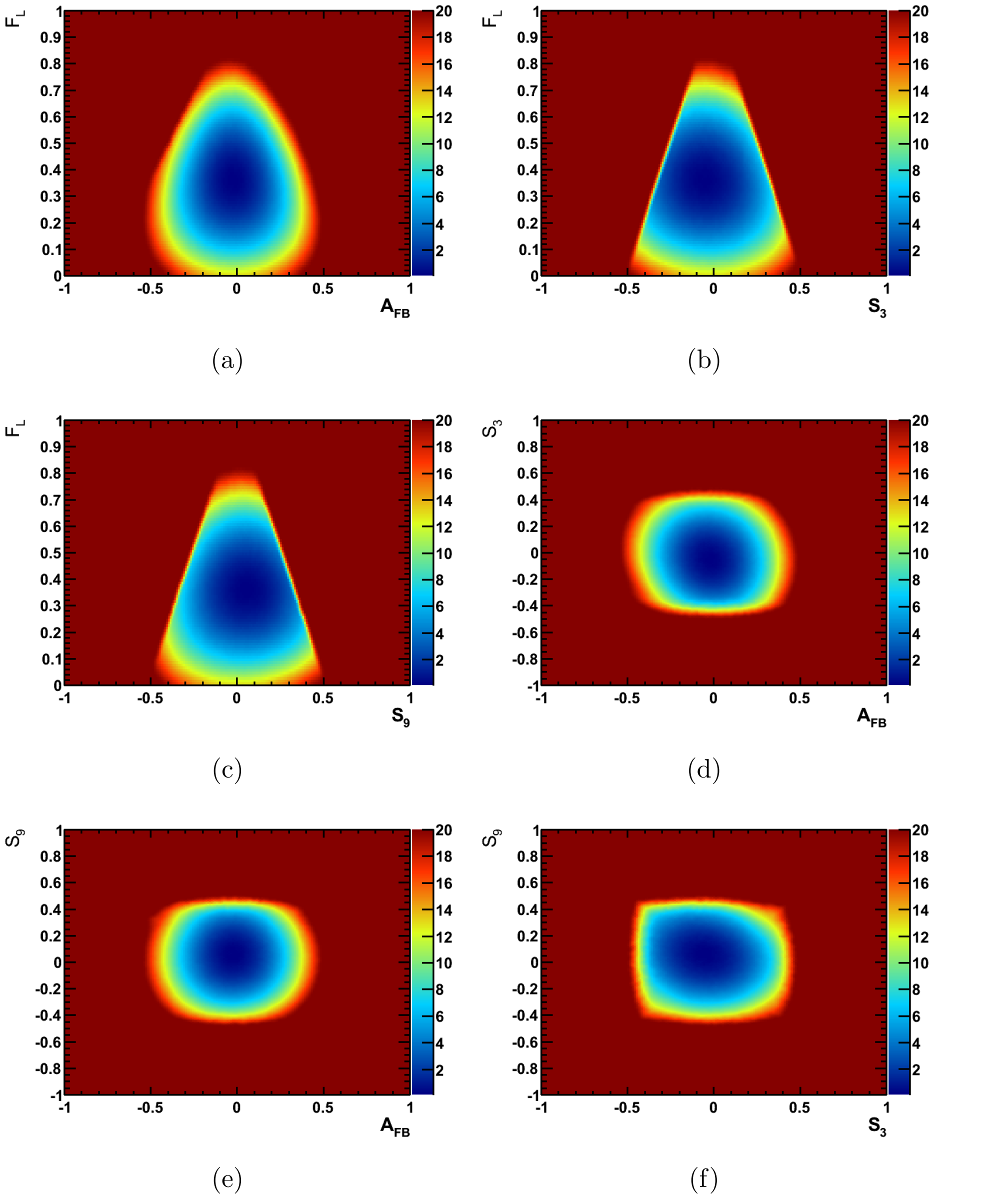


Figure 83: Two dimensional log-likelihood scans for F_{L}, A_{FB}, §3 and §9 in the $0.1 < q^2 < 2\,\mathrm{GeV}^2/c^4$ q^2-bin.

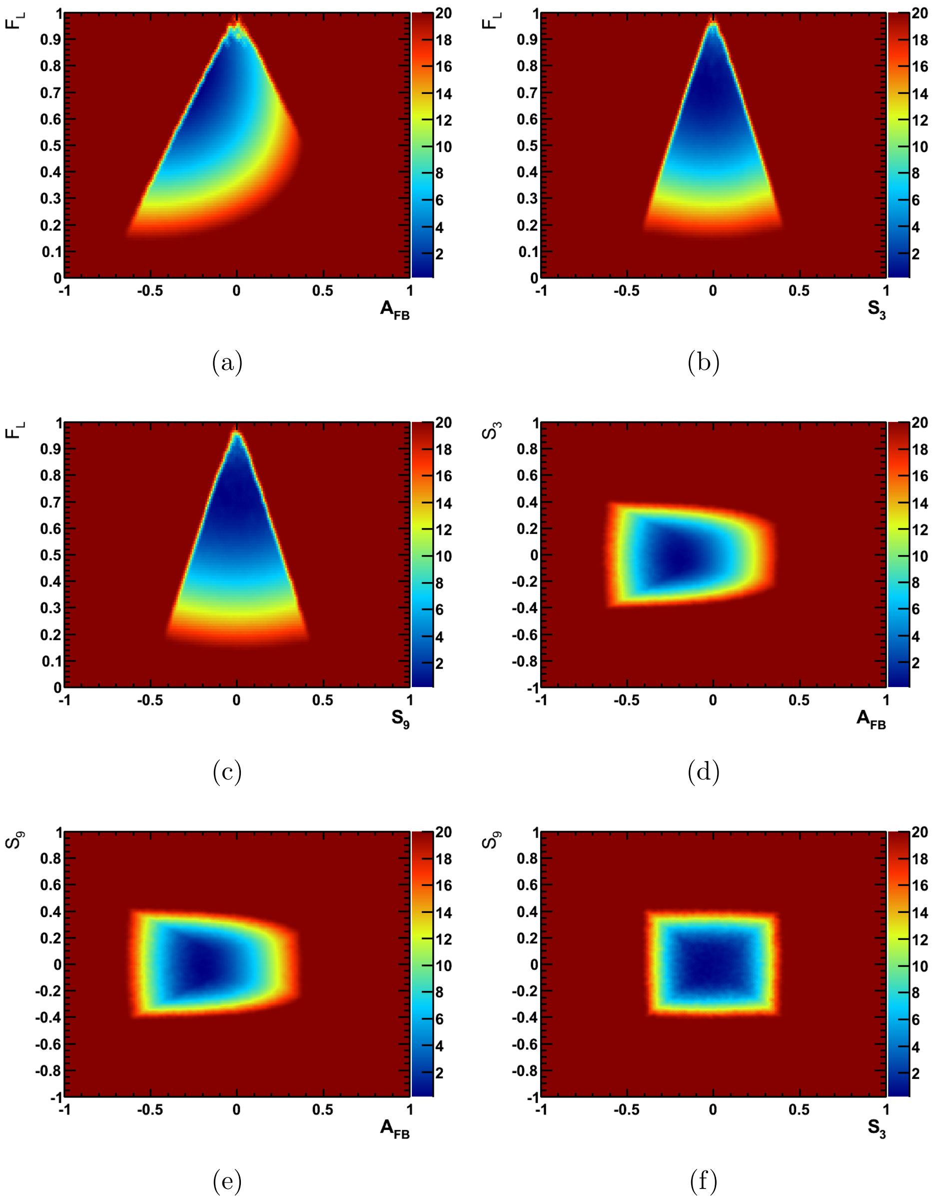
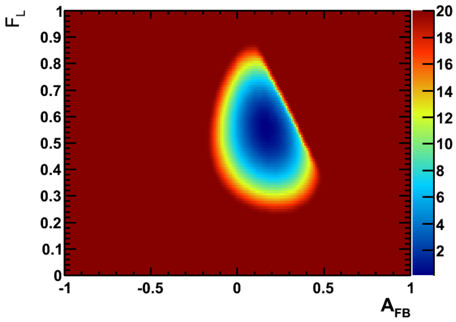
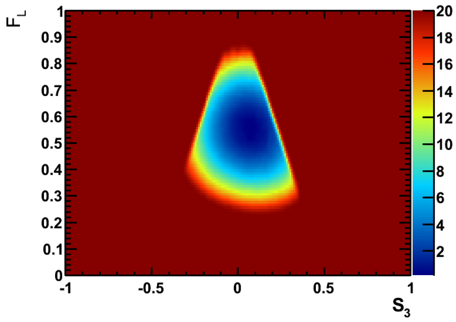


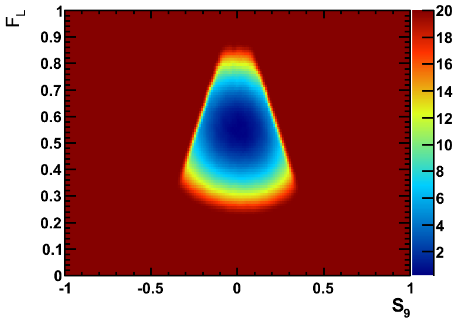
Figure 84: Two dimensional log-likelihood scans for F_{L}, A_{FB}, §3 and §9 in the $2 < q^2 < 4.3\,\mathrm{GeV}^2/c^4$ q^2-bin.



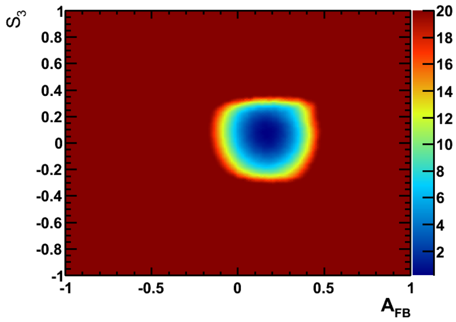
(a)



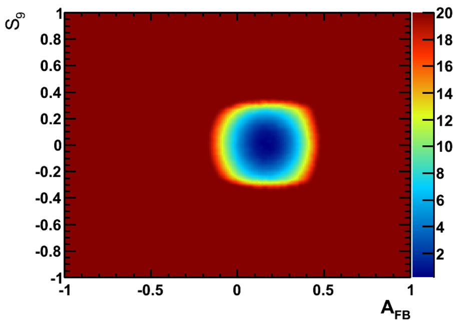
(b)



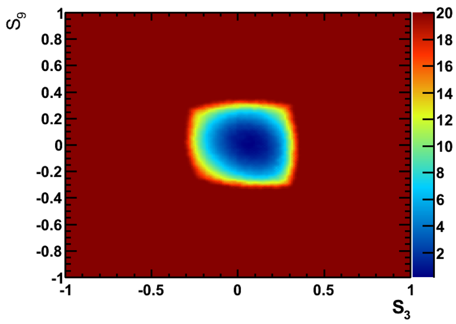
(c)



(d)



(e)



(f)

Figure 85: Two dimensional log-likelihood scans for F_{L}, A_{FB}, §3 and §9 in the $4.3 < q^2 < 8.68\,\mathrm{GeV}^2/c^4$ q^2-bin.

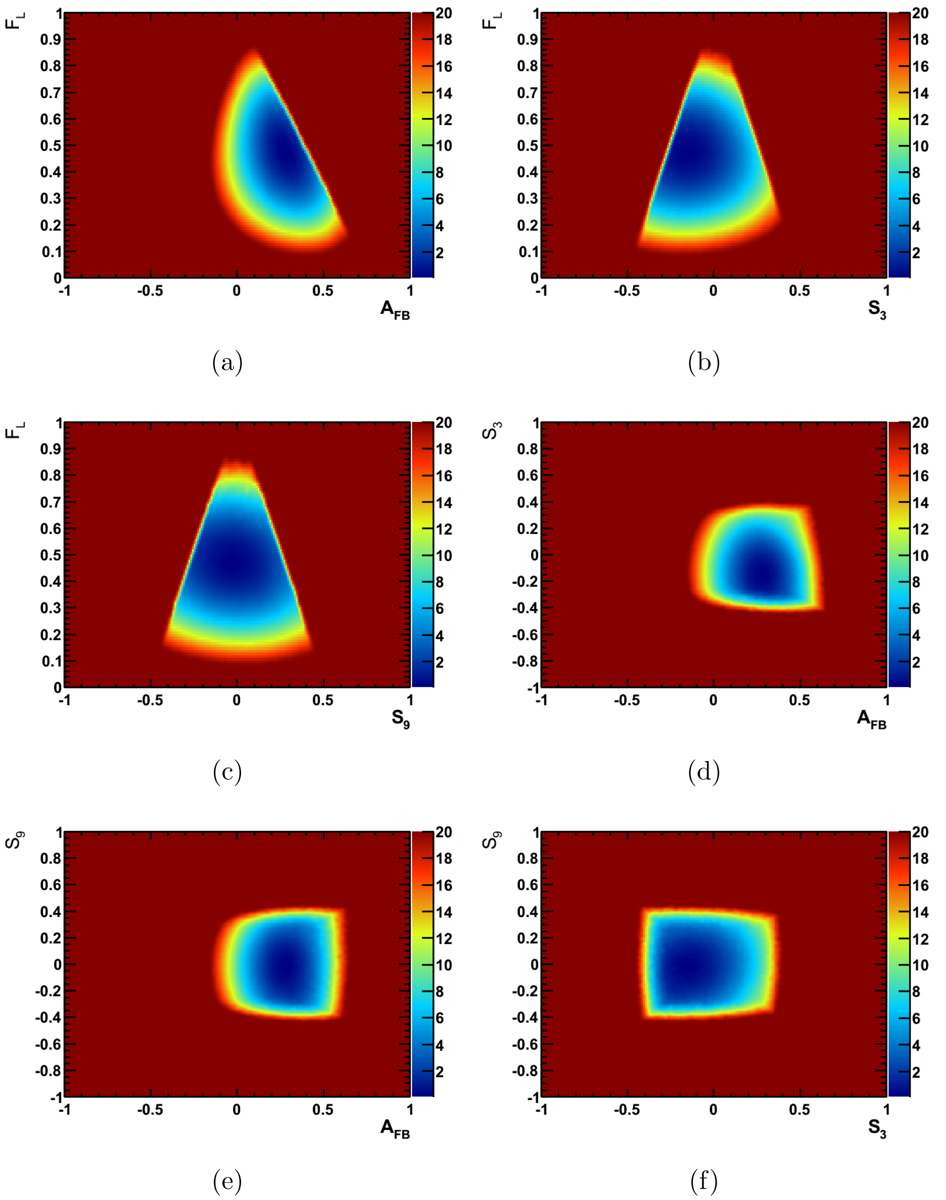


Figure 86: Two dimensional log-likelihood scans for F_{L}, A_{FB}, §3 and §9 in the $10.09 < q^2 < 12.86\,\mathrm{GeV}^2/c^4$ q^2-bin.

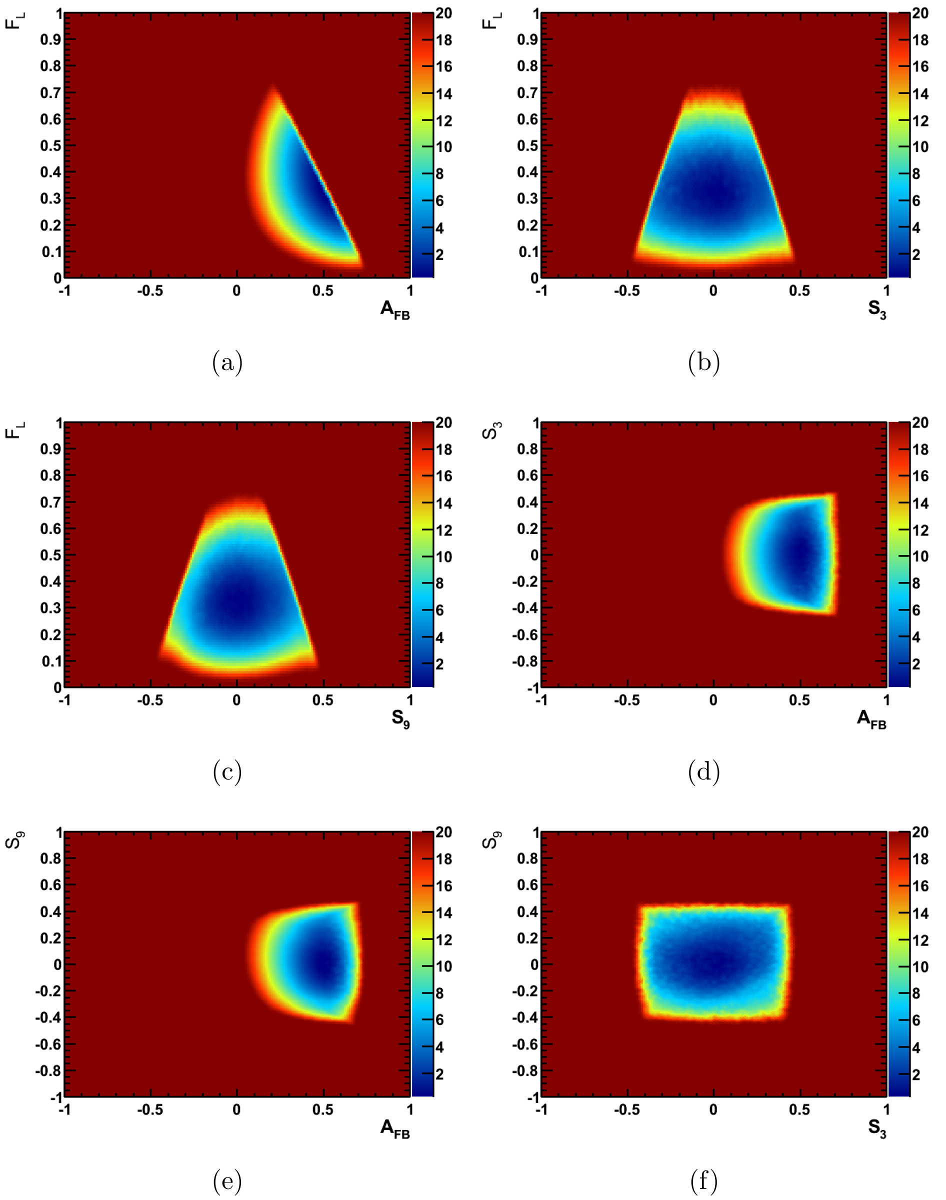


Figure 87: Two dimensional log-likelihood scans for $F_{\rm L}$, $A_{\rm FB}$, §3 and §9 in the $14.18 < q^2 < 16\,{\rm GeV}^2/c^4$ q^2-bin.

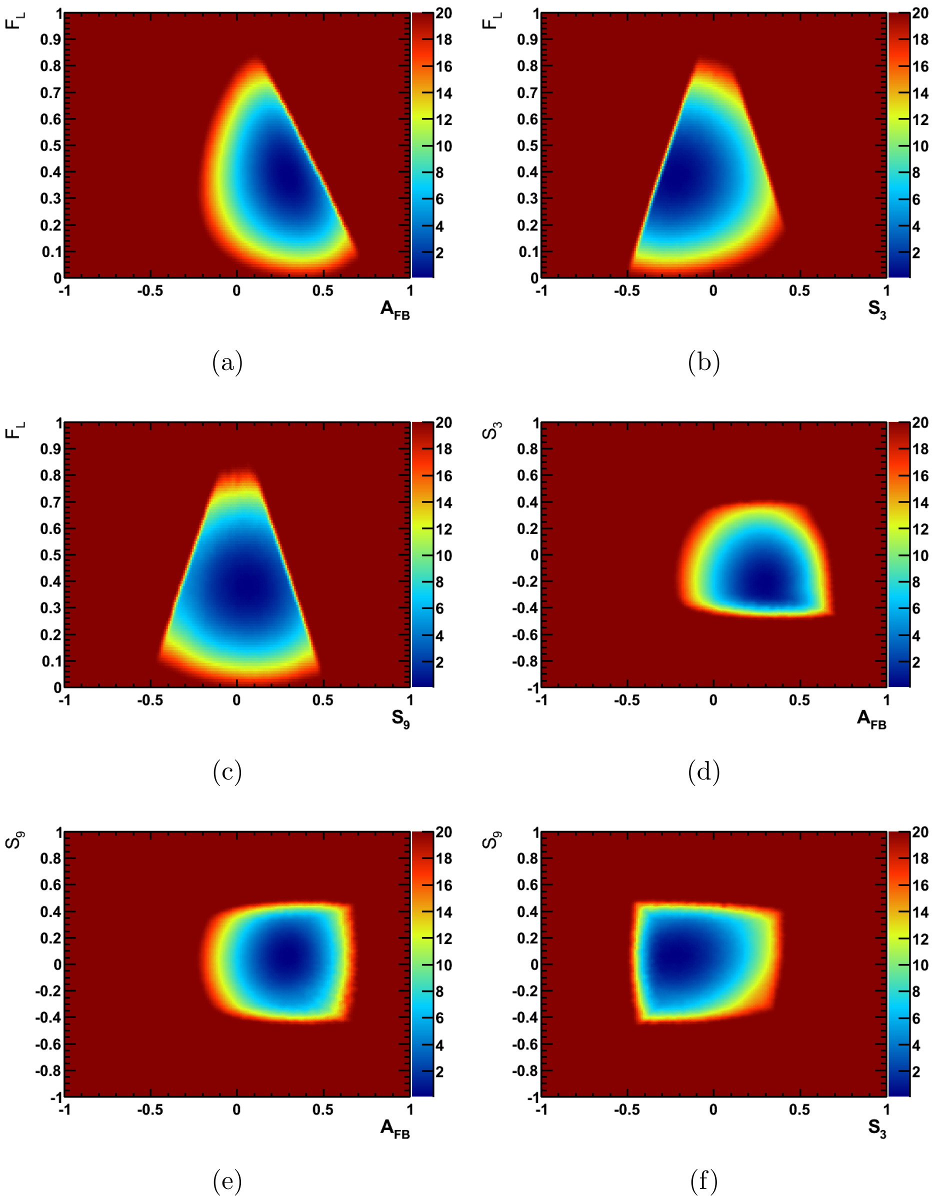
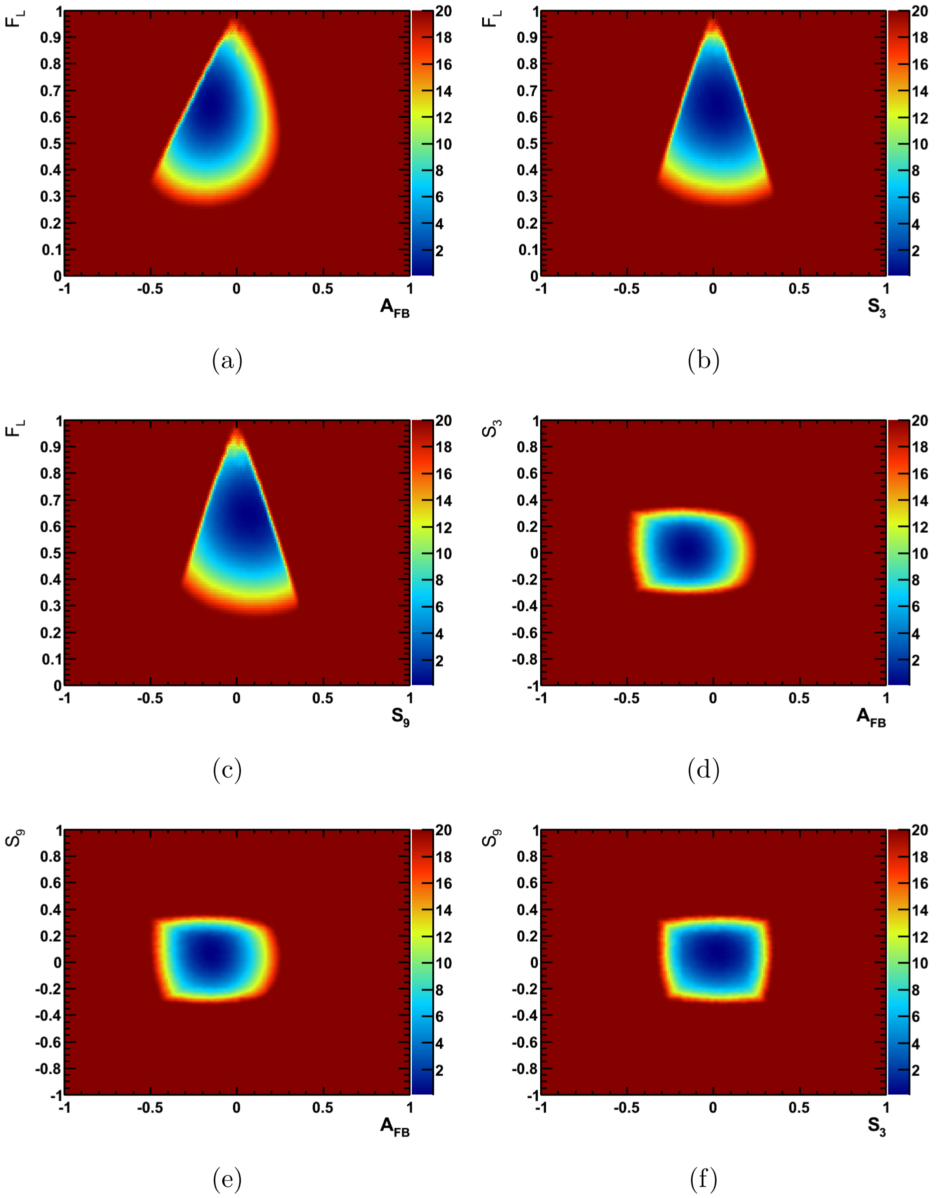


Figure 88: Two dimensional log-likelihood scans for F_{L}, A_{FB}, §3 and §9 in the $16 < q^2 < 19\,\mathrm{GeV}^2/c^4$ q^2-bin.



(a) (b) (c) (d) (e) (f)

Figure 89: Two dimensional log-likelihood scans for F_{L}, A_{FB}, §3 and §9 in the $1 < q^2 < 6\,\mathrm{GeV}^2/c^4$ q^2-bin.

H Systematic variations when re-fitting

In addition to the toy-based method detailed in section 18 of this note, an alternative procedure for estimating the systematic uncertainties is performed.

The following systematic uncertainties are extracted as follows. The standard angular fit is performed on candidates from the data with the nominal acceptance correction applied. The fit is then repeated with a systematically varied acceptance correction applied. The difference in the result of the two fits is taken as an estimate of the systematic uncertainty.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal value	-0.022	-0.198	0.163	0.280	0.511	0.304	-0.170
B^0 p re-weighting [I]	0.001	-0.001	0.002	0.004	0.006	0.002	0.002
B^0 p_T re-weighting [I]	0.000	-0.001	0.001	0.002	0.002	0.001	-0.000
AC CTK Down [C]	-0.001	-0.011	0.005	0.009	0.015	0.007	-0.006
AC CTK Up [C]	0.001	0.010	-0.005	-0.008	-0.013	-0.006	0.005
AC CTL Down CTK Down [C]	-0.001	-0.004	0.001	0.005	0.010	0.005	-0.000
AC CTL Up CTK Down [C]	0.001	0.004	-0.001	-0.005	-0.010	-0.004	0.000
AC CTL Down [C]	0.001	0.007	-0.004	-0.003	-0.005	-0.002	0.006
AC CTL Up [C]	-0.001	-0.006	0.004	0.003	0.005	0.002	-0.005
PID performance -5% [G]	-0.001	0.002	-0.000	-0.003	-0.002	-0.001	0.001
PID performance -10% [G]	-0.001	-0.001	-0.000	-0.003	-0.004	0.001	-0.002
PID performance -30% [G]	-0.002	-0.001	-0.000	-0.004	-0.004	0.001	-0.002
PID performance +5% [G]	-0.001	0.001	-0.000	-0.004	-0.004	-0.002	0.001
PID performance +10% [G]	-0.001	0.004	-0.001	-0.005	-0.001	0.001	0.004
PID performance +30% [G]	-0.001	0.004	-0.001	-0.004	-0.001	0.001	0.004
Removal of soft tracks [O]	0.002	-0.002	0.003	0.007	0.007	0.006	0.000
IP Smearing [H]	-0.001	0.003	-0.001	0.007	0.004	0.002	0.000
IsMuon efficiency Down [E]	-0.003	-0.006	0.001	0.000	-0.001	-0.000	-0.005
IsMuon efficiency Up [E]	0.003	0.004	-0.000	-0.000	-0.001	-0.001	0.004
0th Order Background Model [K]	-0.038	0.002	-0.015	-0.001	-0.017	-0.028	-0.064
1st Order Background Model [K]	0.003	0.020	0.004	0.004	-0.013	-0.005	-0.001
3rd Order Background Model [K]	-0.003	0.001	0.000	-0.001	-0.002	0.001	0.005
K^{*0} mis-ID Down [L]	-0.000	-0.000	0.000	0.000	-0.000	0.000	-0.000
K^{*0} mis-ID Up [L]	0.000	-0.000	-0.000	-0.000	-0.001	-0.000	0.000
No multiple candidates [N]	-0.006	-0.002	0.002	-0.009	0.001	-0.015	0.003
Signal mass width Down [J]	-0.001	-0.001	0.001	0.000	0.002	0.000	0.001
Signal mass width Up [J]	0.001	0.002	-0.001	-0.000	-0.001	-0.000	-0.001
S-wave component [P]	0.002	-0.011	-0.005	-0.009	0.003	-0.022	0.009
AC Non-factorisable Down [C]	0.003	0.000	0.001	0.003	0.001	0.000	-0.001
AC Non-factorisable Up [C]	-0.003	-0.000	-0.001	-0.003	-0.002	-0.000	0.001
Tracking efficiency Down [F]	-0.000	0.000	-0.000	-0.001	-0.001	-0.001	0.000
Tracking efficiency Up [F]	0.000	-0.000	0.000	0.001	0.000	-0.000	0.000
Trigger efficiency Down [D]	0.000	0.001	0.001	0.002	0.003	0.001	0.002
Trigger efficiency Up [D]	0.001	-0.002	0.001	0.002	0.003	0.001	0.000

Table 57: Variation of A_{FB} when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal value	0.369	0.740	0.566	0.477	0.331	0.372	0.654
B^0 p re-weighting [I]	-0.002	-0.001	-0.008	-0.009	-0.008	-0.005	-0.003
B^0 p_T re-weighting [I]	-0.002	-0.002	-0.003	-0.003	-0.002	-0.003	-0.002
AC CTK Down [C]	-0.020	-0.015	-0.020	-0.020	-0.020	-0.020	-0.020
AC CTK Up [C]	0.018	0.014	0.019	0.018	0.017	0.018	0.018
AC CTL Down CTK Down [C]	-0.013	-0.006	-0.007	-0.009	-0.013	-0.011	-0.010
AC CTL Up CTK Down [C]	0.012	0.005	0.007	0.008	0.013	0.010	0.010
AC CTL Down [C]	0.007	0.009	0.013	0.011	0.007	0.008	0.010
AC CTL Up [C]	-0.006	-0.008	-0.012	-0.010	-0.006	-0.008	-0.009
PID performance -5% [G]	-0.001	0.002	0.003	0.008	0.003	0.007	0.003
PID performance -10% [G]	-0.005	-0.001	0.001	0.007	0.005	0.007	-0.001
PID performance -30% [G]	-0.005	-0.001	0.001	0.008	0.005	0.007	-0.000
PID performance +5% [G]	0.001	0.001	0.003	0.008	0.005	0.009	0.003
PID performance +10% [G]	0.003	0.006	0.005	0.010	0.001	0.005	0.005
PID performance +30% [G]	0.003	0.005	0.005	0.008	0.002	0.005	0.005
Removal of soft tracks [O]	-0.006	-0.003	-0.013	-0.016	-0.010	-0.007	-0.007
IP Smearing [H]	-0.002	0.004	-0.005	-0.015	-0.006	-0.007	-0.002
IsMuon efficiency Down [E]	-0.003	-0.008	-0.003	0.000	0.001	-0.000	-0.006
IsMuon efficiency Up [E]	0.002	0.006	0.003	-0.000	0.002	0.001	0.004
0th Order Background Model [K]	0.006	0.002	0.020	-0.032	0.022	0.026	0.021
1st Order Background Model [K]	0.005	0.026	0.017	-0.033	0.017	0.024	0.026
3rd Order Background Model [K]	0.005	0.001	0.001	-0.002	0.002	0.006	0.001
K^{*0} mis-ID Down [L]	-0.000	-0.000	0.000	-0.000	0.000	0.000	0.000
K^{*0} mis-ID Up [L]	0.000	-0.000	-0.000	-0.000	0.001	-0.000	0.000
No multiple candidates [N]	-0.002	-0.003	0.001	0.004	-0.001	0.016	-0.004
Signal mass width Down [J]	0.004	-0.001	-0.000	0.001	-0.002	-0.001	-0.001
Signal mass width Up [J]	-0.004	0.002	0.000	-0.001	0.002	0.001	0.001
S-wave component [P]	0.012	-0.014	-0.013	0.003	-0.004	0.009	-0.017
AC Non-factorisable Down [C]	0.002	0.000	-0.000	0.002	-0.001	0.001	0.001
AC Non-factorisable Up [C]	-0.002	-0.000	0.000	-0.001	0.002	-0.000	-0.001
Tracking efficiency Down [F]	0.000	0.000	0.002	0.001	0.002	0.001	0.000
Tracking efficiency Up [F]	0.000	-0.000	-0.002	-0.002	-0.000	-0.001	-0.001
Trigger efficiency Down [D]	0.001	0.001	-0.004	-0.004	-0.004	-0.003	0.000
Trigger efficiency Up [D]	-0.002	-0.002	-0.005	-0.004	-0.004	-0.002	-0.003

Table 58: Variation of F_{L} when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal value	-0.043	-0.038	0.077	-0.157	0.028	-0.223	0.029
B^0 p re-weighting [I]	-0.003	-0.001	0.002	0.004	0.002	-0.000	0.000
B^0 p_T re-weighting [I]	-0.001	-0.001	0.000	0.001	0.001	-0.001	0.000
AC CTK Down [C]	-0.001	-0.001	0.002	0.001	0.002	-0.004	0.002
AC CTK Up [C]	0.001	0.000	-0.002	-0.001	-0.002	0.003	-0.002
AC CTL Down CTK Down [C]	-0.004	-0.001	0.001	0.001	0.001	-0.002	0.002
AC CTL Up CTK Down [C]	0.004	0.000	-0.001	-0.001	-0.001	0.002	-0.002
AC CTL Down [C]	-0.003	-0.000	-0.001	-0.000	-0.001	0.002	-0.000
AC CTL Up [C]	0.003	-0.000	0.001	-0.000	0.001	-0.002	0.000
PID performance -5% [G]	-0.002	0.001	-0.001	0.001	-0.000	0.002	-0.000
PID performance -10% [G]	0.003	-0.001	-0.001	-0.000	-0.001	-0.000	-0.000
PID performance -30% [G]	0.002	-0.001	-0.001	-0.000	-0.001	-0.000	-0.000
PID performance +5% [G]	0.001	-0.001	-0.000	0.002	-0.000	0.003	-0.000
PID performance +10% [G]	-0.002	-0.001	-0.001	0.002	0.000	0.003	-0.001
PID performance +30% [G]	-0.002	-0.001	-0.001	0.002	0.000	0.003	-0.001
Removal of soft tracks [O]	-0.000	0.002	0.003	0.005	0.003	-0.004	0.001
IP Smearing [H]	0.002	0.003	0.001	0.001	-0.001	-0.008	0.002
IsMuon efficiency Down [E]	0.002	-0.001	0.000	0.001	-0.001	-0.002	0.001
IsMuon efficiency Up [E]	-0.000	-0.000	-0.001	-0.000	0.000	-0.000	-0.001
0th Order Background Model [K]	-0.012	-0.007	-0.013	-0.061	-0.002	0.007	-0.022
1st Order Background Model [K]	-0.020	-0.009	-0.016	-0.062	0.009	0.014	-0.025
3rd Order Background Model [K]	-0.002	-0.002	-0.005	0.000	-0.020	-0.010	-0.002
K^{*0} mis-ID Down [L]	-0.000	-0.000	-0.000	0.000	-0.000	0.000	-0.000
K^{*0} mis-ID Up [L]	0.000	0.000	0.000	0.000	0.000	-0.000	-0.000
No multiple candidates [N]	-0.001	-0.001	0.010	-0.003	0.001	0.003	0.002
Signal mass width Down [J]	-0.001	0.002	-0.003	-0.000	0.000	0.000	0.001
Signal mass width Up [J]	0.001	-0.006	0.003	-0.000	-0.000	-0.000	-0.001
S-wave component [P]	-0.021	-0.014	-0.018	0.013	0.014	0.050	-0.014
AC Non-factorisable Down [C]	0.001	0.003	-0.001	0.006	-0.001	-0.003	-0.002
AC Non-factorisable Up [C]	-0.002	-0.004	0.001	-0.006	0.002	0.003	0.001
Tracking efficiency Down [F]	0.001	-0.000	-0.000	-0.000	-0.001	-0.001	0.000
Tracking efficiency Up [F]	-0.000	-0.000	0.000	0.000	0.000	-0.001	0.000
Trigger efficiency Down [D]	-0.002	-0.000	0.001	0.002	0.001	-0.000	-0.000
Trigger efficiency Up [D]	-0.001	-0.000	0.001	0.002	0.001	-0.000	0.000

Table 59: Variation of S_3 when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal value	0.052	-0.030	0.012	-0.015	0.004	0.056	0.071
B^0 p re-weighting [I]	0.002	-0.001	0.001	-0.003	-0.002	0.001	0.000
B^0 p_T re-weighting [I]	0.001	-0.001	0.000	-0.001	-0.001	0.002	0.000
AC CTK Down [C]	0.002	-0.003	0.002	-0.001	-0.002	0.002	0.002
AC CTK Up [C]	-0.001	0.003	-0.002	0.001	0.001	-0.002	-0.002
AC CTL Down CTK Down [C]	0.002	-0.004	0.000	0.000	0.002	0.002	0.002
AC CTL Up CTK Down [C]	-0.002	0.003	0.000	-0.000	-0.002	-0.002	-0.002
AC CTL Down [C]	0.001	-0.000	-0.002	0.001	0.003	0.000	-0.000
AC CTL Up [C]	-0.001	0.000	0.002	-0.001	-0.004	-0.000	0.000
PID performance -5% [G]	0.001	0.002	-0.001	-0.000	0.000	-0.002	-0.000
PID performance -10% [G]	0.001	0.002	0.000	-0.000	0.002	-0.002	0.001
PID performance -30% [G]	0.001	0.002	0.000	-0.000	0.002	-0.002	0.001
PID performance +5% [G]	0.001	0.001	-0.000	-0.002	0.001	-0.003	-0.000
PID performance +10% [G]	0.002	-0.001	-0.001	-0.001	0.001	-0.003	-0.001
PID performance +30% [G]	0.002	-0.001	-0.001	-0.001	0.001	-0.003	-0.001
Removal of soft tracks [O]	0.002	-0.001	0.003	-0.005	-0.002	0.004	0.001
IP Smearing [H]	0.005	-0.001	-0.000	-0.002	-0.001	-0.002	-0.001
IsMuon efficiency Down [E]	-0.001	-0.000	-0.000	-0.001	-0.001	0.001	0.000
IsMuon efficiency Up [E]	0.000	0.000	-0.000	0.001	0.000	-0.000	-0.000
0th Order Background Model [K]	0.012	0.009	-0.018	-0.002	0.021	-0.008	0.018
1st Order Background Model [K]	0.002	0.013	-0.005	0.004	0.005	0.010	0.004
3rd Order Background Model [K]	0.007	-0.027	0.013	-0.009	-0.012	-0.014	-0.008
K^{*0} mis-ID Down [L]	0.000	-0.000	0.000	-0.000	-0.000	0.000	0.000
K^{*0} mis-ID Up [L]	-0.000	0.000	-0.000	0.000	-0.000	-0.000	-0.000
No multiple candidates [N]	-0.003	-0.007	0.009	0.016	-0.002	0.030	0.004
Signal mass width Down [J]	0.002	0.003	0.002	0.002	-0.003	-0.005	0.002
Signal mass width Up [J]	-0.002	-0.006	-0.001	-0.001	0.002	0.004	-0.002
S-wave component [P]	0.003	-0.019	-0.009	0.003	-0.015	0.009	-0.013
AC Non-factorisable Down [C]	-0.004	-0.002	0.006	-0.005	0.004	-0.005	0.001
AC Non-factorisable Up [C]	0.003	0.002	-0.006	0.006	-0.004	0.005	-0.001
Tracking efficiency Down [F]	-0.000	0.001	-0.000	0.000	-0.000	-0.000	-0.000
Tracking efficiency Up [F]	0.000	-0.001	0.000	-0.000	-0.000	0.001	0.000
Trigger efficiency Down [D]	0.001	-0.001	0.001	-0.002	-0.001	0.001	0.000
Trigger efficiency Up [D]	0.001	-0.000	0.001	-0.002	-0.001	0.001	0.000

Table 60: Variation of S_9 when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal value	0.121	0.056	-0.133	-0.003	-0.061	-0.000	0.026
B^0 p re-weighting [I]	0.001	-0.005	-0.001	0.000	-0.003	-0.000	-0.000
B^0 p_T re-weighting [I]	0.000	0.025	-0.000	0.000	-0.002	-0.001	0.000
AC CTK Down [C]	0.003	-0.009	-0.003	-0.001	-0.001	0.001	0.002
AC CTK Up [C]	-0.002	0.004	0.003	0.000	-0.004	-0.001	-0.001
AC CTL Down CTK Down [C]	0.002	-0.003	-0.001	-0.002	-0.003	0.003	0.001
AC CTL Up CTK Down [C]	-0.002	0.010	0.001	0.001	0.000	-0.003	-0.001
AC CTL Down [C]	-0.000	0.002	0.001	-0.001	-0.002	0.002	-0.001
AC CTL Up [C]	0.000	-0.009	-0.001	0.001	0.003	-0.002	0.001
PID performance -5% [G]	-0.001	-0.012	0.001	-0.000	-0.003	-0.002	-0.000
PID performance -10% [G]	0.001	0.007	0.001	-0.001	0.001	-0.001	0.001
PID performance -30% [G]	0.001	0.004	0.001	-0.001	-0.002	-0.001	0.001
PID performance +5% [G]	-0.000	0.022	0.000	0.001	0.000	-0.003	0.000
PID performance +10% [G]	-0.001	0.021	0.000	0.001	-0.003	-0.001	-0.001
PID performance +30% [G]	-0.001	0.022	0.000	0.001	-0.001	-0.001	-0.000
Removal of soft tracks [O]	0.004	0.006	-0.002	0.000	-0.002	-0.001	0.001
IP Smearing [H]	-0.000	0.030	-0.002	-0.002	-0.001	-0.001	-0.002
IsMuon efficiency Down [E]	-0.001	0.002	0.000	0.000	0.001	0.000	0.001
IsMuon efficiency Up [E]	0.000	-0.028	0.000	-0.000	-0.004	0.000	-0.001
0th Order Background Model [K]	-0.014	0.023	-0.002	-0.004	-0.034	0.001	-0.013
1st Order Background Model [K]	0.071	0.099	-0.150	0.015	-0.060	-0.046	-0.042
3rd Order Background Model [K]	0.003	-0.026	0.009	0.003	-0.002	0.006	-0.009
K^{*0} mis-ID Down [L]	0.000	0.001	-0.000	-0.000	-0.003	-0.000	0.000
K^{*0} mis-ID Up [L]	-0.000	0.005	0.000	0.000	-0.003	0.000	-0.000
No multiple candidates [N]	-0.012	0.023	0.025	0.017	-0.001	0.017	0.015
Signal mass width Down [J]	-0.002	0.003	0.001	-0.002	0.002	0.003	-0.004
Signal mass width Up [J]	0.002	0.001	-0.001	0.001	-0.006	-0.002	0.003
S-wave component [P]	-0.037	-0.016	0.010	0.002	-0.015	-0.008	-0.011
AC Non-factorisable Down [C]	-0.004	-0.007	0.001	0.001	0.004	0.009	0.001
AC Non-factorisable Up [C]	0.004	0.007	-0.002	-0.001	-0.009	-0.009	-0.001
Tracking efficiency Down [F]	-0.000	-0.000	0.000	-0.000	-0.002	0.001	-0.000
Tracking efficiency Up [F]	-0.000	0.001	-0.000	0.000	-0.003	0.000	-0.000
Trigger efficiency Down [D]	0.000	0.007	-0.000	0.000	-0.002	-0.000	-0.000
Trigger efficiency Up [D]	0.001	-0.009	-0.001	0.000	-0.002	-0.000	0.000

Table 61: Variation of A_9 when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal value	-0.046	-1.000	0.500	0.713	1.000	0.644	-0.656
B^0 p re-weighting [I]	0.002	-0.000	-0.004	-0.002	-0.000	-0.000	0.013
B^0 p_T re-weighting [I]	0.001	-0.000	-0.001	-0.001	-0.000	-0.001	0.004
AC CTK Down [C]	-0.002	0.000	-0.008	-0.005	0.000	-0.005	0.016
AC CTK Up [C]	0.002	-0.000	0.007	0.004	-0.000	0.004	-0.014
AC CTL Down CTK Down [C]	-0.001	0.000	-0.005	0.001	0.000	-0.001	0.019
AC CTL Up CTK Down [C]	0.001	-0.000	0.005	-0.001	-0.000	0.001	-0.017
AC CTL Down [C]	0.001	-0.000	0.003	0.005	-0.000	0.003	0.003
AC CTL Up [C]	-0.001	-0.000	-0.002	-0.005	-0.000	-0.003	-0.002
PID performance -5% [G]	-0.002	-0.000	0.003	0.002	0.000	0.005	-0.000
PID performance -10% [G]	-0.003	-0.000	0.000	0.001	-0.000	0.009	-0.006
PID performance -30% [G]	-0.003	-0.000	0.001	0.001	0.000	0.009	-0.005
PID performance +5% [G]	-0.002	-0.000	0.003	0.001	0.000	0.006	-0.002
PID performance +10% [G]	-0.002	-0.000	0.003	0.001	-0.000	0.007	0.006
PID performance +30% [G]	-0.003	0.000	0.003	0.001	-0.000	0.007	0.006
Removal of soft tracks [O]	0.004	0.000	-0.006	-0.004	-0.000	0.005	0.014
IP Smearing [H]	-0.003	-0.000	-0.008	-0.004	-0.000	-0.003	0.003
IsMuon efficiency Down [E]	-0.007	-0.000	-0.001	0.001	-0.000	-0.001	-0.008
IsMuon efficiency Up [E]	0.005	0.000	0.001	-0.001	-0.000	-0.001	0.006
0th Order Background Model [K]	-0.080	0.000	-0.023	-0.048	-0.000	-0.031	-0.271
1st Order Background Model [K]	0.007	-0.000	0.031	-0.038	-0.000	0.013	-0.048
3rd Order Background Model [K]	-0.007	0.000	0.002	-0.006	-0.000	0.009	0.016
K^{*0} mis-ID Down [L]	-0.000	0.000	0.000	0.000	-0.000	0.000	-0.000
K^{*0} mis-ID Up [L]	0.000	0.000	-0.000	-0.000	-0.000	-0.000	0.000
No multiple candidates [N]	-0.012	-0.000	0.006	-0.017	-0.000	-0.015	0.020
Signal mass width Down [J]	-0.001	-0.000	0.002	0.002	-0.000	-0.000	0.005
Signal mass width Up [J]	0.001	-0.000	-0.002	-0.002	-0.000	0.000	-0.004
S-wave component [P]	0.003	0.000	-0.030	-0.018	-0.000	-0.037	0.071
AC Non-factorisable Down [C]	0.005	-0.000	0.003	0.010	-0.000	0.001	-0.006
AC Non-factorisable Up [C]	-0.006	-0.000	-0.003	-0.010	-0.000	-0.001	0.005
Tracking efficiency Down [F]	-0.001	-0.000	0.001	-0.000	-0.000	-0.001	-0.000
Tracking efficiency Up [F]	0.001	-0.000	-0.001	-0.001	0.000	-0.001	0.001
Trigger efficiency Down [D]	0.001	0.000	-0.002	-0.001	0.000	-0.000	0.007
Trigger efficiency Up [D]	0.001	0.000	-0.002	-0.001	-0.000	-0.000	0.006

Table 62: Variation of A_{T}^{Re} when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal value	0.369	0.740	0.566	0.477	0.329	0.372	0.654
B^0 p re-weighting [I]	-0.002	-0.001	-0.008	-0.009	-0.008	-0.005	-0.003
B^0 p_T re-weighting [I]	-0.002	-0.002	-0.003	-0.003	-0.002	-0.003	-0.002
AC CTK Down [C]	-0.020	-0.015	-0.020	-0.020	-0.020	-0.020	-0.020
AC CTK Up [C]	0.018	0.014	0.019	0.018	0.018	0.018	0.018
AC CTL Down CTK Down [C]	-0.013	-0.006	-0.007	-0.009	-0.013	-0.011	-0.010
AC CTL Up CTK Down [C]	0.012	0.005	0.007	0.008	0.012	0.010	0.010
AC CTL Down [C]	0.007	0.009	0.013	0.011	0.007	0.008	0.010
AC CTL Up [C]	-0.006	-0.008	-0.012	-0.010	-0.006	-0.008	-0.009
PID performance -5% [G]	-0.001	0.002	0.003	0.008	0.003	0.007	0.003
PID performance -10% [G]	-0.005	-0.001	0.001	0.007	0.004	0.007	-0.001
PID performance -30% [G]	-0.005	-0.001	0.001	0.008	0.005	0.007	-0.000
PID performance +5% [G]	0.001	0.001	0.003	0.008	0.004	0.009	0.003
PID performance +10% [G]	0.003	0.005	0.005	0.010	0.001	0.005	0.005
PID performance +30% [G]	0.003	0.005	0.005	0.008	0.001	0.005	0.005
Removal of soft tracks [O]	-0.006	-0.003	-0.013	-0.016	-0.010	-0.007	-0.007
IP Smearing [H]	-0.002	0.004	-0.005	-0.015	-0.006	-0.007	-0.002
IsMuon efficiency Down [E]	-0.003	-0.008	-0.003	0.000	0.000	-0.000	-0.006
IsMuon efficiency Up [E]	0.002	0.006	0.003	-0.000	0.001	0.001	0.004
0th Order Background Model [K]	0.006	0.004	0.020	-0.032	0.021	0.026	0.021
1st Order Background Model [K]	0.005	0.027	0.017	-0.033	0.017	0.024	0.026
3rd Order Background Model [K]	0.005	0.002	0.001	-0.002	0.001	0.006	0.001
K^{*0} mis-ID Down [L]	-0.000	-0.000	0.000	-0.000	-0.000	-0.000	0.000
K^{*0} mis-ID Up [L]	0.000	0.000	-0.000	0.000	0.000	0.000	0.000
No multiple candidates [N]	-0.002	-0.003	0.001	0.004	-0.001	0.016	-0.004
Signal mass width Down [J]	0.004	-0.002	-0.000	0.001	-0.002	-0.001	-0.001
Signal mass width Up [J]	-0.004	0.002	0.000	-0.001	0.002	0.001	0.001
S-wave component [P]	0.012	-0.015	-0.013	0.003	-0.004	0.009	-0.017
AC Non-factorisable Down [C]	0.002	0.000	-0.000	0.002	-0.001	0.001	0.001
AC Non-factorisable Up [C]	-0.002	-0.000	0.000	-0.001	0.002	-0.000	-0.001
Tracking efficiency Down [F]	0.000	0.000	0.002	0.001	0.002	0.001	0.000
Tracking efficiency Up [F]	0.000	-0.000	-0.002	-0.002	-0.000	-0.001	-0.001
Trigger efficiency Down [D]	0.001	0.001	-0.004	-0.004	-0.004	-0.003	0.000
Trigger efficiency Up [D]	-0.002	-0.002	-0.005	-0.004	-0.004	-0.002	-0.003

Table 63: Variation of F_{L} when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal value	-0.137	-0.290	0.355	-0.601	0.067	-0.709	0.167
B^0 p re-weighting [I]	-0.008	-0.001	0.000	0.026	0.006	0.004	-0.001
B^0 p_T re-weighting [I]	-0.002	-0.001	-0.002	0.008	0.001	-0.001	-0.000
AC CTK Down [C]	0.001	0.013	-0.006	0.029	0.004	0.011	0.004
AC CTK Up [C]	-0.001	-0.011	0.006	-0.025	-0.004	-0.010	-0.003
AC CTL Down CTK Down [C]	-0.010	0.003	-0.002	0.016	0.002	0.007	0.006
AC CTL Up CTK Down [C]	0.010	-0.003	0.002	-0.015	-0.001	-0.006	-0.005
AC CTL Down [C]	-0.011	-0.009	0.004	-0.012	-0.003	-0.004	0.002
AC CTL Up [C]	0.011	0.008	-0.004	0.011	0.002	0.004	-0.002
PID performance -5% [G]	-0.005	0.001	-0.000	-0.004	0.000	-0.003	-0.001
PID performance -10% [G]	0.009	-0.002	-0.004	-0.010	-0.003	-0.010	-0.001
PID performance -30% [G]	0.007	-0.002	-0.004	-0.010	-0.004	-0.009	-0.001
PID performance +5% [G]	0.002	-0.006	0.001	-0.002	-0.001	-0.001	0.000
PID performance +10% [G]	-0.007	-0.013	-0.001	-0.003	0.000	0.004	-0.004
PID performance +30% [G]	-0.006	-0.015	-0.002	-0.001	0.001	0.003	-0.006
Removal of soft tracks [O]	0.000	0.016	0.001	0.039	0.007	-0.005	0.005
IP Smearing [H]	0.006	0.020	0.001	0.021	-0.004	-0.019	0.013
IsMuon efficiency Down [E]	0.006	0.004	-0.000	0.002	-0.002	-0.005	0.005
IsMuon efficiency Up [E]	-0.002	-0.007	-0.000	-0.000	0.001	-0.002	-0.002
0th Order Background Model [K]	-0.037	-0.075	-0.043	-0.210	-0.003	-0.008	-0.109
1st Order Background Model [K]	-0.065	-0.085	-0.056	-0.213	0.029	0.017	-0.125
3rd Order Background Model [K]	-0.007	-0.018	-0.020	0.004	-0.052	-0.039	-0.009
K^{*0} mis-ID Down [L]	0.000	0.000	0.000	0.000	-0.000	0.000	0.000
K^{*0} mis-ID Up [L]	-0.000	-0.000	-0.000	-0.000	0.000	-0.000	0.000
No multiple candidates [N]	-0.002	-0.005	0.047	-0.015	0.002	-0.010	0.011
Signal mass width Down [J]	-0.005	0.046	-0.014	-0.001	0.002	0.002	0.008
Signal mass width Up [J]	0.005	-0.045	0.012	-0.001	-0.002	-0.001	-0.007
S-wave component [P]	-0.068	-0.095	-0.097	0.045	0.036	0.148	-0.092
AC Non-factorisable Down [C]	0.004	0.027	-0.006	0.020	-0.005	-0.010	-0.009
AC Non-factorisable Up [C]	-0.007	-0.027	0.007	-0.022	0.004	0.011	0.005
Tracking efficiency Down [F]	0.003	-0.002	0.001	-0.002	-0.001	-0.004	0.001
Tracking efficiency Up [F]	-0.001	-0.001	-0.001	0.004	0.001	-0.002	-0.000
Trigger efficiency Down [D]	-0.007	-0.002	0.000	0.012	0.003	0.003	-0.001
Trigger efficiency Up [D]	-0.003	0.001	0.000	0.012	0.002	0.001	-0.000

Table 64: Variation of A_{T}^2 when systematically varying fit parameters or the weights applied to the input data set.

Systematic	$0.1 < q^2 < 2.0$	$2.0 < q^2 < 4.3$	$4.3 < q^2 < 8.68$	$10.09 < q^2 < 12.86$	$14.18 < q^2 < 16.0$	$16.0 < q^2 < 19.0$	$0.1 < q^2 < 6.0$
Nominal value	0.164	-0.232	0.054	-0.056	0.019	0.179	0.411
B^0 p re-weighting [I]	0.006	-0.008	0.005	-0.013	-0.005	0.003	-0.001
B^0 p_T re-weighting [I]	0.002	-0.004	0.001	-0.005	-0.002	0.004	-0.002
AC CTK Down [C]	-0.000	-0.010	0.006	-0.001	-0.005	0.001	-0.010
AC CTK Up [C]	0.001	0.008	-0.005	0.001	0.004	-0.001	0.009
AC CTL Down CTK Down [C]	0.005	-0.019	-0.001	0.002	0.007	0.004	-0.000
AC CTL Up CTK Down [C]	-0.004	0.018	0.001	-0.001	-0.006	-0.004	0.001
AC CTL Down [C]	0.005	-0.010	-0.006	0.002	0.011	0.003	0.009
AC CTL Up [C]	-0.005	0.010	0.006	-0.002	-0.010	-0.003	-0.009
PID performance -5% [G]	0.004	0.010	-0.002	-0.002	0.001	-0.005	0.001
PID performance -10% [G]	0.000	0.019	0.002	-0.002	0.007	-0.003	0.003
PID performance -30% [G]	0.001	0.019	0.002	-0.003	0.007	-0.004	0.003
PID performance +5% [G]	0.005	0.006	-0.001	-0.008	0.004	-0.007	0.002
PID performance +10% [G]	0.007	-0.010	-0.003	-0.006	0.004	-0.009	0.002
PID performance +30% [G]	0.006	-0.011	-0.002	-0.005	0.003	-0.009	0.002
Removal of soft tracks [O]	0.004	-0.007	0.012	-0.019	-0.006	0.010	-0.003
IP Smearing [H]	0.014	-0.007	-0.002	-0.006	-0.001	-0.007	-0.006
IsMuon efficiency Down [E]	-0.003	0.007	-0.001	-0.003	-0.001	0.005	-0.005
IsMuon efficiency Up [E]	0.001	-0.005	-0.001	0.003	0.001	-0.000	0.004
0th Order Background Model [K]	0.039	0.093	-0.079	-0.004	0.062	-0.017	0.119
1st Order Background Model [K]	0.009	0.095	-0.018	0.019	0.014	0.039	0.049
3rd Order Background Model [K]	0.023	-0.204	0.060	-0.033	-0.037	-0.042	-0.043
K^{*0} mis-ID Down [L]	0.000	-0.000	0.000	-0.000	0.000	0.000	0.000
K^{*0} mis-ID Up [L]	-0.000	0.000	-0.000	0.000	-0.000	-0.000	-0.000
No multiple candidates [N]	-0.009	-0.054	0.042	0.059	-0.004	0.097	0.021
Signal mass width Down [J]	0.008	0.044	0.008	0.006	-0.009	-0.016	0.013
Signal mass width Up [J]	-0.007	-0.046	-0.006	-0.005	0.008	0.014	-0.012
S-wave component [P]	0.012	-0.138	-0.043	0.011	-0.045	0.032	-0.097
AC Non-factorisable Down [C]	-0.012	-0.018	0.027	-0.020	0.013	-0.016	0.009
AC Non-factorisable Up [C]	0.011	0.020	-0.028	0.022	-0.012	0.015	-0.009
Tracking efficiency Down [F]	-0.001	0.006	-0.002	0.001	0.000	0.000	0.000
Tracking efficiency Up [F]	0.000	-0.006	0.001	-0.000	-0.001	0.004	0.001
Trigger efficiency Down [D]	0.004	-0.005	0.002	-0.007	-0.003	0.001	0.001
Trigger efficiency Up [D]	0.002	-0.002	0.002	-0.007	-0.003	0.002	-0.002

Table 65: Variation of A_{T}^{Im} when systematically varying fit parameters or the weights applied to the input data set.

I Weight scaling scheme

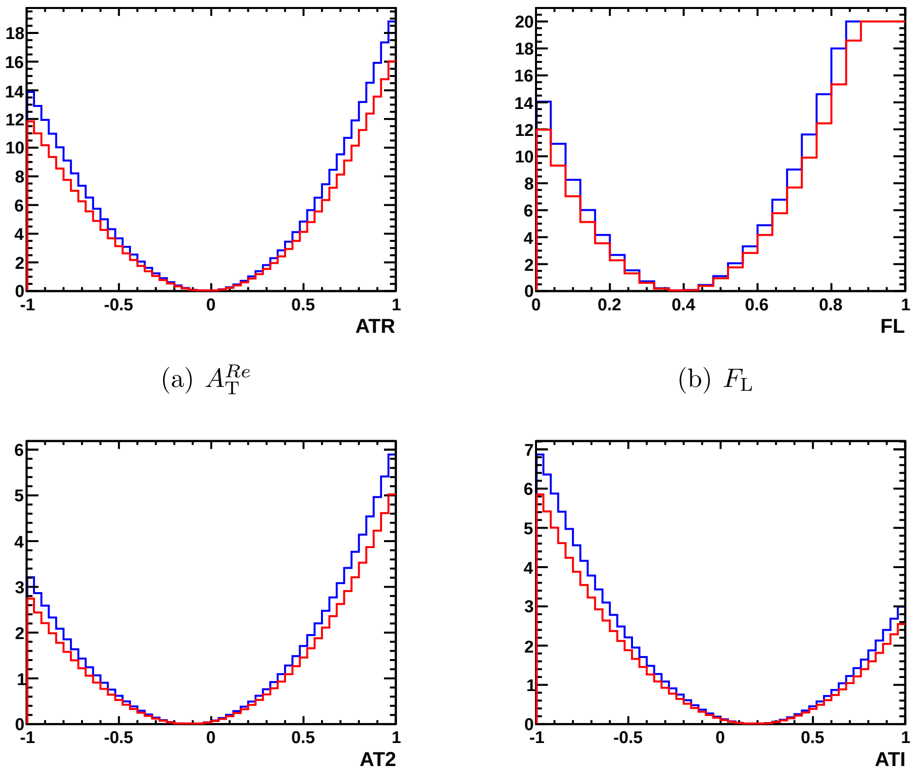
In the acceptance correction procedure, each candidate is re-weighted according to the inverse of the efficiency. As the total efficiency of each candidate is on the order of 0.5% , the weight given to each candidate is on the order of 200. In the analysis, the weights are renormalised according to

$$\alpha = \frac{N}{\sum_{i=1}^{N} w_i}, \tag{30}$$

where N is the number of candidates in the sample, and w_i is the weight of each candidate. This ensures that the sum-of-weights of the candidates is equal to the number of candidates in the sample. An alternative approach would be to scale the weights according to

$$\alpha = \frac{\sum_{i=1}^{N} w_i}{\sum_{i=1}^{N} (w_i)^2}. \tag{31}$$

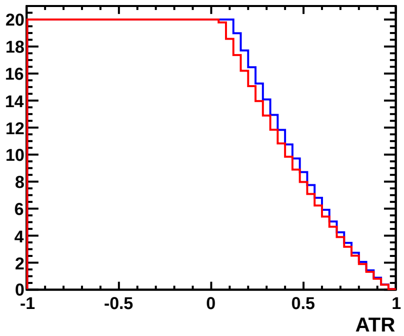
To compare the two weighting schemes, 1D likelihood scans are produced for each obervable using each of the weighting schemes, see Figs. 90 and 91. These distributions indicate that the weighting scheme given in Eq. 31 gives larger confidence intervals for each observable than that used in the analysis (Eq. 30), which are more similar to the intervals obtained from the FC procedure in Sec. 15.1.1. The same behaviour is observed in each of the q^2 bins.



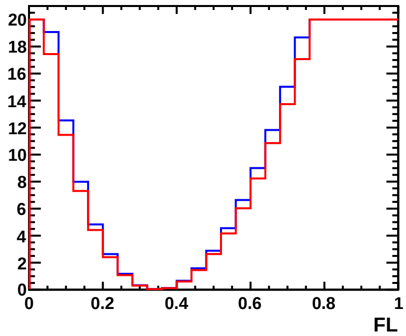
(a) A_{T}^{Re} (b) F_{L}

(c) A_{T}^{2} (d) A_{T}^{Im}

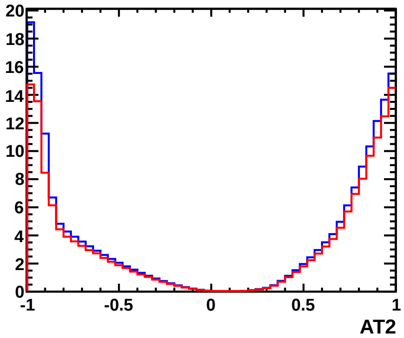
Figure 90: Comparison of likelihood scans for the observables (a) A_{T}^{Re}, (b) F_{L}, (c) A_{T}^{2} and (d) A_{T}^{Im} in the $0.10 < q^2 2.00 <$ GeV$^2/c^4$ region, if the weight of candidates from the data is renormalised according to Eq. 30 (blue histogram) and Eq. 31 (red histogram).



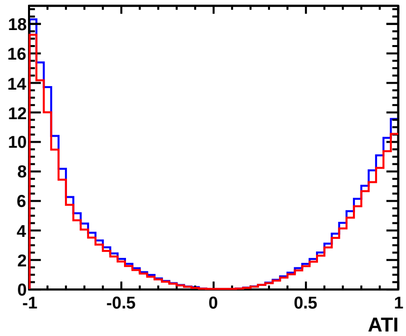
(a) A_{T}^{Re}



(b) F_{L}



(c) A_{T}^{2}



(d) A_{T}^{Im}

Figure 91: Comparison of likelihood scans for the observables (a) A_{T}^{Re}, (b) F_{L}, (c) A_{T}^{2} and (d) A_{T}^{Im} in the $14.18 < q^2 16.00 <$ GeV$^2/c^4$ region, if the weight of candidates from the data is renormalised according to Eq. 30 (blue histogram) and Eq. 31 (red histogram).

References

[1] M. Williams, *How good are your fits? Unbinned multivariate goodness-of-fit tests in high energy physics*, JINST **5** (2010) P09004, `arXiv:1006.3019`.

[2] LHCb collaboration, T. Blake *et al.*, *Analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at LHCb*, LHCb-ANA-2011-077.

[3] LHCb collaboration, *Angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$*, LHCb-CONF-2011-038.

[4] BaBar Collaboration, B. Aubert *et al.*, *Angular distributions in the decays $B \to K^*\ell^+\ell^-$*, Phys. Rev. **D79** (2009) 031102, `arXiv:0804.4412`.

[5] 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`.

[6] 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`, 8 pages, 2 figures, 4 tables. Submitted to Phys. Rev. Lett.

[7] 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`, 7 pages, 3 figures, 3 tables. Submitted to Phys. Rev. Lett.

[8] LHCb collaboration, *Differential branching fraction and angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay*, LHCb-CONF-2012-008.

[9] C. Bobeth, G. Hiller, D. van Dyk, and C. Wacker, *The Decay $B \to K\ell^+\ell^-$ at Low Hadronic Recoil and Model-Independent $\Delta B = 1$ Constraints*, `arXiv:arXiv/1111.2558`.

[10] U. Egede, T. Hurth, J. Matias, M. Ramon, and W. Reece, *New observables in the decay mode $\bar{B}_d \to \bar{K}^{*0}\ell^+\ell^-$*, JHEP **0811** (2008) 032, `arXiv:0807.2589`.

[11] W. Altmannshofer *et al.*, *Symmetries and Asymmetries of $B \to K^*\mu^+\mu^-$ Decays in the Standard Model and Beyond*, JHEP **0901** (2009) 019, `arXiv:0811.1214`.

[12] D. Becirevic and E. Schneider, *On transverse asymmetries in $B \to K^*\ell^+\ell^-$*, Nucl. Phys. **B854** (2012) 321, `arXiv:1106.3283`.

[13] T. Blake *et al.*, *Angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at LHCb with 1 fb^{-1}*, LHCb-ANA-2011-089.

[14] LHCb collaboration, T. Blake *et al.*, *Analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at LHCb*, LHCb-ANA-2011-022.

[15] J. Dickens, V. Gibson, C. Lazzeroni, and M. Patel, *A Study of the Sensitivity to the Forward-Backward Asymmetry in $B_d \to K^*\mu^+ mu^-$ Decays at LHCb*, Tech. Rep. LHCb-2007-039. CERN-LHCb-2007-039, CERN, Geneva, Jul, 2007.

[16] 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.

[17] LHCb collaboration, *Evidence for the decay $B_s^0 \to J/\psi\overline{K}^{*0}$*, LHCb-CONF-2011-025.

[18] M. Pivk and F. R. Le Diberder, *sPlot: a statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, `arXiv:physics/0402083`.

[19] C. Bobeth, G. Hiller, and D. van Dyk, *More Benefits of Semileptonic Rare B Decays at Low Recoil: CP Violation*, JHEP **1107** (2011) 067, `arXiv:1105.0376`.

[20] M. Needham *et al.*, *Flavor-untagged angular analysis of $B_d^0 \to J/\psi K^*$ and $B_s^0 \to J/\psi\phi$ decays*, Tech. Rep. LHCb-ANA-2011-002, CERN, Geneva, May, 2011.

[21] 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`.

[22] H.-Y. Cheng, Y.-Y. Keum, and K.-C. Yang, *$B \to J/\psi K^*$ decays in QCD factorization*, Phys. Rev. **D65** (2002) 094023, `arXiv:hep-ph/0111094`.

[23] G. J. Feldman and R. D. Cousins, *A Unified Approach to the Classical Statistical Analysis of Small Signals*, Phys. Rev. **D57** (1998) 3873, `arXiv:physics/9711021`.

[24] C.-D. Lu and W. Wang, *Analysis of $B \to K_J^*(\to K\pi)\mu^+\mu^-$ in the higher kaon resonance region*, `arXiv:1111.1513`, * Temporary entry *.

[25] C. Bobeth, G. Hiller, and G. Piranishvili, *CP Asymmetries in $\bar{B} \to \bar{K}^*(\to \bar{K}\pi)\bar{\ell}\ell$ and Untagged $\bar{B}_s$, $B_s \to \phi(\to K^+K^-)\bar{\ell}\ell$ Decays at NLO*, JHEP **0807** (2008) 106, `arXiv:0805.2525`.

[26] 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`, 21 pages, 16 figures. Minor typo in Eq. (4.8) corrected: version to appear in Phys. Rev. D Report-no: UAB-FT 560.

Ω