

The raw charge asymmetry of $B^\pm \rightarrow K^\pm \mu^+ \mu^-$ decays as a function of q^2 , with a charmonium control channel

Ishaan Makkar

June 2025

Abstract

The rare non-resonant component of $B^\pm \rightarrow K^\pm \mu^+ \mu^-$ probes flavour-changing neutral currents and is strongly suppressed in the Standard Model, making its direct CP asymmetry an interesting observable: the Standard Model predicts $|A_{CP}| \sim 10^{-3}$, while some extensions allow values two orders of magnitude larger. We measure the *raw* charge asymmetry $A_{\text{raw}} = (N_{B^+} - N_{B^-}) / (N_{B^+} + N_{B^-})$ of fitted signal yields in six intervals of the dimuon invariant mass squared q^2 , using reduced LHCb Run-1 pp collision ntuples of the $B \rightarrow K \mu^+ \mu^-$ channel with a multivariate selection, and we determine the same quantity for the control mode $B^+ \rightarrow J/\psi K^+$, in which the combined production and detection asymmetry of the final state dominates. The control channel gives $A_{\text{raw}}(B^+ \rightarrow J/\psi K^+) = +0.01367 \pm 0.00100$ (statistical) from 1.21×10^6 signal decays, agreeing in magnitude with the published LHCb determination of the corresponding quantity. In the six q^2 intervals the raw asymmetries lie between -0.158 and $+0.250$ with statistical intervals of 0.07 – 0.20 ; two of the six lie within one statistical standard deviation of the control value and the largest departure is 1.6σ , so at this precision no compelling discrepancy with the control is seen. The rare-mode mass fits are supported by refitted pseudo-experiments conditional on the fitted model, giving calibrated p -values between 0.16 and 0.77 ; the high-statistics control fit is not described adequately by the same analytic shape and is used only to supply a resolution. Because the reduced ntuples do not retain the information needed for the fiducial acceptance requirement, the peaking-background vetoes, the efficiency corrections or the per-polarity fits of a dedicated CP analysis, raw asymmetries are quoted throughout and no CP asymmetry is extracted. The difference between a raw asymmetry and the control value is shown only as a conditional diagnostic; it is not a calibrated test of the Standard Model, and no such test is claimed here.

Contents

1	Introduction	3
2	Detector and data sample	4
3	Candidate selection	6
4	Mass fits and extraction of the asymmetry	7
5	Uncertainties and robustness	10
6	Results	12
7	Discussion	17
8	Conclusions	17
A	Likelihood, minimisation and resampling	19
B	Analysis variations and fit diagnostics	20

1 Introduction

1.1 Why this decay

The Standard Model (SM) describes the strong, weak and electromagnetic interactions of the known quarks, leptons and gauge bosons with great precision, yet accounts neither for the observed baryon asymmetry of the universe nor for dark matter [1]. Processes that the SM suppresses very strongly are therefore valuable: a small extra amplitude from physics beyond the SM changes a suppressed rate by a large relative amount, whereas the same amplitude would be invisible against an unsuppressed one.

The decay $B^\pm \rightarrow K^\pm \mu^+ \mu^-$ is suppressed three times over. It is a flavour-changing neutral current, $b \rightarrow s$, in which the quark flavour changes without a change of electric charge; such transitions have no tree-level amplitude in the SM and proceed only through electroweak loop (“penguin”) and box diagrams. The loop costs a factor of the weak coupling; the Glashow–Iliopoulos–Maiani mechanism makes the contributions of the internal up-type quarks cancel almost exactly; and the relevant product of Cabibbo–Kobayashi–Maskawa matrix elements, $|V_{tb}V_{ts}^*| \simeq 0.04$, is small. The result is a branching fraction of $(4.36 \pm 0.23) \times 10^{-7}$ [2]. Particles circulating in such a loop are off their mass shell — energy and momentum are conserved at every vertex — and the loop is sensitive to any heavy state coupling to the b and s quarks, whether or not that state is light enough to be produced directly.

1.2 The observable

The direct CP asymmetry of the decay is

$$A_{CP} \equiv \frac{\Gamma(B^- \rightarrow K^- \mu^+ \mu^-) - \Gamma(B^+ \rightarrow K^+ \mu^+ \mu^-)}{\Gamma(B^- \rightarrow K^- \mu^+ \mu^-) + \Gamma(B^+ \rightarrow K^+ \mu^+ \mu^-)}, \quad (1)$$

with the ordering used in the published LHCb measurement [3]. Hadronic form-factor uncertainties largely cancel in a ratio of this kind, so the SM prediction is clean and small, $|A_{CP}| = \mathcal{O}(10^{-3})$ [4], whereas CP -violating new physics in $b \rightarrow s \mu^+ \mu^-$ can reach the 10^{-1} level in related modes [5]. Testing that prediction requires a calibrated measurement of A_{CP} itself; what the present data support is set out in Section 1.4.

What an experiment counts is not A_{CP} but the raw asymmetry of reconstructed yields. Taking the positively charged final state as the reference, we define

$$A_{\text{raw}} \equiv \frac{N_{B^+} - N_{B^-}}{N_{B^+} + N_{B^-}}, \quad (2)$$

where N_{B^+} and N_{B^-} are the fitted signal yields of $B^+ \rightarrow K^+ \mu^+ \mu^-$ and $B^- \rightarrow K^- \mu^+ \mu^-$. To first order in small asymmetries,

$$A_{\text{raw}} = -A_{CP} + A_P + A_D, \quad (3)$$

where A_P is the production asymmetry of B^+ and B^- mesons inside the acceptance and A_D the difference in detection and reconstruction efficiency between the $K^+ \mu^+ \mu^-$ and $K^- \mu^+ \mu^-$ final states, both written in the ordering of Eq. (2). The minus sign on A_{CP} records the change of ordering between Eqs. (1) and (2): reversing which charge is written first reverses the sign of *one and the same* observable. It does not follow that a raw asymmetry and a published A_{CP} must have opposite signs, because they are different quantities — the raw asymmetry also contains A_P and A_D , which here are larger than A_{CP} is expected to be. What does follow is that the two can only be compared after the orderings are matched and the nuisance terms are accounted for, and Section 6.2 does both explicitly.

The two nuisance terms are not small compared with the signal being sought. The detection asymmetry of a charged kaon is of order one per cent, because K^- are absorbed more strongly

than K^+ by detector material, while $|A_{CP}|$ is expected at the per-mille level. They must therefore be measured rather than assumed. The standard device is a control channel with the same final-state particles and similar kinematics in which CP violation is known to be negligible; here that is $B^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^+$, for which

$$A_{\text{raw}}(B^+ \rightarrow J/\psi K^+) \simeq A_P + A_D. \quad (4)$$

Eq. (4) holds to the extent that the control and rare modes are produced and reconstructed with the same charge-dependent efficiencies. That is an assumption about kinematic matching, selection and detector response, not an identity; Section 5.5 states which parts of it this analysis is able to check and which it is not.

1.3 Dependence on q^2

The dimuon invariant mass squared,

$$q^2 = (p_{\mu^+} + p_{\mu^-})^2, \quad (5)$$

organises the decay. Natural units with $\hbar = c = 1$ are used throughout; masses are quoted in MeV and q^2 in GeV^2 , which in units where c is restored would read MeV/c^2 and GeV^2/c^4 . At low q^2 the amplitude is dominated by the electromagnetic penguin operator and at high q^2 by the semileptonic operators. In between sit two narrow charmonium resonances that decay to two muons: the J/ψ at 3097 MeV, that is $q^2 \simeq 9.59 \text{ GeV}^2$, and the $\psi(2S)$ at 3686 MeV, that is $q^2 \simeq 13.59 \text{ GeV}^2$. There $B^+ \rightarrow J/\psi K^+$ and $B^+ \rightarrow \psi(2S)K^+$ exceed the rare mode by three orders of magnitude, so those regions are excluded from the rare-decay measurement and used instead as controls. Because new physics can enter with a q^2 -dependent phase, the asymmetry is measured in intervals of q^2 and not only inclusively.

1.4 Scope

We measure A_{raw} in six q^2 intervals and in the two charmonium control regions. The emphasis throughout is on making each step checkable: the selection is characterised rather than assumed, rare-mode fit quality is calibrated by simulating the fitted model rather than by an asymptotic formula, and the sensitivity of the result to each modelling choice is quoted as a measured range rather than as a systematic uncertainty. We do not convert A_{raw} into A_{CP} , and we do not present any test of the Standard Model: without the acceptance, veto and efficiency corrections of a dedicated analysis, the quantities measured here are not CP asymmetries, and agreement or disagreement between them and a control value cannot be read as evidence for or against the Standard Model. Section 5 sets out which corrections a dedicated measurement applies and why they are unavailable here; an independent published LHCb analysis of Run-1 data [3] provides the external point of comparison in Section 6.

2 Detector and data sample

2.1 Features of LHCb that matter for a charge asymmetry

LHCb is a single-arm forward spectrometer at the Large Hadron Collider covering pseudorapidity $2 < \eta < 5$, designed for beauty and charm physics [6]. Four of its properties bear directly on this measurement.

The dipole magnet. A warm dipole of about 4 Tm bending power separates charges and provides the momentum measurement, and its polarity is reversed periodically during data taking. Reversal averages over the left–right asymmetry of the detector, but it does *not* remove the part of A_D that comes from the different nuclear interaction cross-sections of positive and

Table 1. Composition of the data sample. The magnet polarity and year of subsample 3 are read from its parent ntuple; the other three are inferred as described in the text. Each subsample carries its own selection score, produced by a separately trained classifier (Section 3).

Subsample	Magnet polarity, year	Candidates	Selection model
1	down, 2011 (inferred)	2,147,487	classifier A
2	down, 2012 (inferred)	4,262,377	classifier B
3	up, 2011 (from the parent ntuple)	1,498,035	classifier C
4	up, 2012 (inferred)	4,475,158	classifier D
Total		12,383,057	

negative hadrons with detector material [3]. That component survives magnet reversal, and it is the main reason a control channel is needed rather than a symmetry argument.

Particle identification. Two ring-imaging Cherenkov detectors separate kaons from pions and a muon system identifies muons. Because the charge of the kaon tags the flavour of the decaying B meson, a kaon–pion or kaon–muon misidentification can move a candidate from one charge category to the other and so bias the asymmetry directly.

Acceptance edges. Tracks near the edge of the acceptance can be swept out of the detector by the magnetic field in a charge-dependent way, producing locally very large apparent asymmetries. A dedicated analysis removes these regions with a fiducial requirement on the kaon momentum components [3].

Mass resolution. The tracking system gives a relative momentum resolution of 0.4–0.6%, corresponding to a B -candidate mass resolution of roughly 20 MeV; the value measured in the control channel below is $\sigma_1 = 16.9$ MeV for the core Gaussian.

2.2 Data sample

The analysis uses four reduced ntuples of the LHCb Run-1 $B \rightarrow K\mu^+\mu^-$ selection, supplied for this project by the project supervisors rather than retrieved by the authors from a public portal. Each file corresponds to one magnet polarity and one year of data taking and carries, for every candidate, the $K\mu^+\mu^-$ invariant mass, the dimuon mass, the sign of the reconstructed B meson, eight kinematic and geometric quantities, and a precomputed multivariate selection score. Table 1 lists them.

One of the four parent ntuples survives in full, with 1099 branches. Its luminosity accounting block sums to $397.15 \pm 0.37 \text{ pb}^{-1}$ and its polarity branch is uniformly positive, which identifies it as the magnet-up 2011 sample. The relative sizes of the four subsamples track the per-polarity Run-1 luminosities, and together with the file ordering used consistently in the original processing this supports the assignment in Table 1; for the other three files that assignment is an inference and not a label read from the data. The four files together hold 12.38×10^6 candidates. Scaling the one surviving luminosity record by the relative sample sizes is consistent with the full Run-1 dataset of about 3 fb^{-1} , but the exact luminosity and run coverage of the other three reduced files cannot be established from the surviving metadata. No luminosity-normalised quantity is therefore quoted anywhere in this paper, and no simulated sample accompanies the data.

Each subsample enters the analysis exactly once. This is worth stating explicitly because the observable is a difference between two nearly equal numbers: entering one subsample twice and omitting another would understate the statistical uncertainty, unbalance the magnet polarities, and spoil the binned fit quality, since the bin contents would no longer be Poisson distributed.

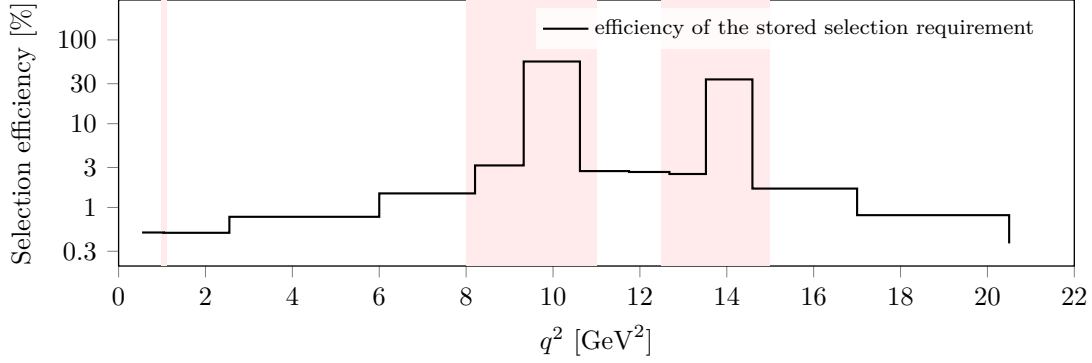


Figure 1. Efficiency of the stored selection requirement against q^2 , measured on subsample 3 as the fraction of input candidates retained, not a true-signal efficiency. It is 0.5% in the lowest interval and 55% in the J/ψ region, a factor of roughly one hundred. Shaded bands mark the $\phi(1020)$ and charmonium regions excluded from the rare-mode measurement.

3 Candidate selection

3.1 The multivariate selection

Background in this channel is dominated by random combinations of tracks that happen to form a B -like vertex. The reduced ntuples carry a precomputed score from a gradient-boosted decision-tree classifier [7], trained separately on each subsample, and the analysis applies a single requirement to that score. Because the selection shapes the sample in which a charge asymmetry is then measured, we characterise it explicitly rather than accept its original description. The stored models were read directly and re-evaluated; each reproduces the score stored in its own subsample to within 1.3×10^{-6} , so the selection is known exactly, candidate by candidate.

The classifier takes ten inputs: the transverse momenta of the B candidate, the kaon and each muon; the vertex-fit χ^2 ; the impact-parameter and flight-distance significances with respect to the primary vertex; the direction angle; the dimuon invariant mass; and the sign of the reconstructed B meson. The reconstructed B mass itself is correctly *not* an input, which matters because it is the variable fitted later, and no other input is strongly correlated with it: the largest linear correlation is +0.30, for the dimuon mass. Two aspects of the input set nevertheless need care, and both are quantified below. The dimuon mass is the variable in which the results are binned, and the reconstructed charge is the variable whose asymmetry is being measured. The training target was not a truth-matched signal sample but a mass window: candidates with $5240 < m(K\mu^+\mu^-) < 5320$ MeV were labelled positive against $m(K\mu^+\mu^-) > 5500$ MeV labelled negative. The labels are therefore noisy by construction, and 78% of the positive class lies in the J/ψ region.

3.2 What the selection does, measured

Figure 1 shows the efficiency of the selection requirement against q^2 : 0.5% in the lowest interval and 55% in the J/ψ region. These are retention fractions for the observed mixture of signal and background candidates, not estimates of true-signal efficiency. Their strong variation is consistent with a classifier trained against a charmonium-dominated mass window while receiving dimuon mass as an input. The selected sample is overwhelmingly charmonium: after the mass requirement of Section 4 it contains 1,523,573 candidates, the great majority in the two resonance regions, as Figure 2 shows. Candidate retention in the rare-mode intervals is much lower; the surviving signal yields determine the statistical precision of the fits. The archived score is retained to keep the nominal selection reproducible from the recovered models. Retraining without charge and dimuon mass would define a different selection requiring

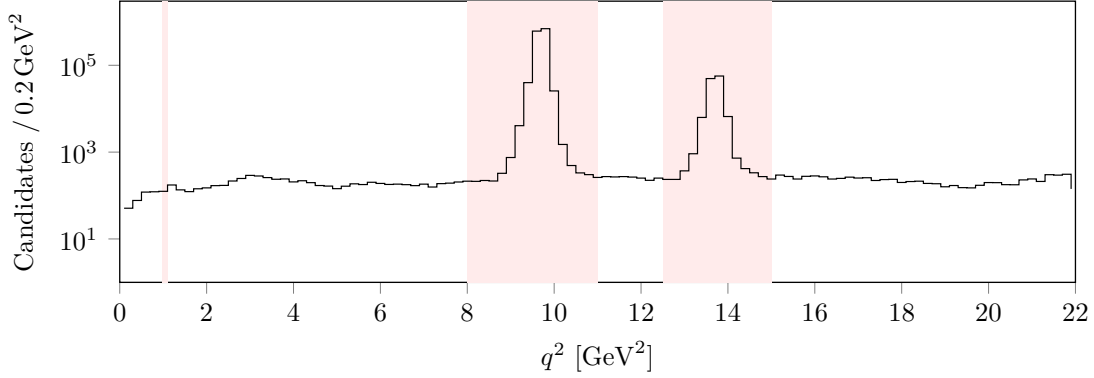


Figure 2. Distribution of q^2 for the 1,523,573 candidates passing the selection and the mass requirement $5200 < m(K\mu^+\mu^-) < 5700$ MeV. Shaded bands are the excluded $\phi(1020)$, J/ψ and $\psi(2S)$ regions. The three orders of magnitude between the resonance peaks and the rare-mode continuum reflect both the physical branching fractions and the strong q^2 dependence of the selection efficiency in Figure 1.

independent validation; it is not undertaken here. The dependence of the stored selection on these inputs is examined explicitly below.

The second point is more serious in principle: the reconstructed charge is an input, so the selection is not manifestly blind to the quantity being measured. That input carries 0.8% of the decision splits, against 13.1% for the dimuon mass and 14.2% for the kaon transverse momentum, but a small effect on the central observable still matters, so we measure it. Each candidate was scored twice with the same model, once with the charge as reconstructed and once with its sign reversed and every other input unchanged. The mean change in score is 0.0122, and 0.15% of selection decisions change in the non-resonant region. Since the efficiency there is itself only 0.82%, that corresponds to an induced efficiency asymmetry of +0.0066: real, but an order of magnitude below the statistical uncertainties reported below. Section 5 goes further and rebuilds the selection so that its explicit charge dependence is removed exactly, then repeats the whole measurement.

3.3 Intervals in q^2

Three regions are excluded from the rare-mode measurement, following Ref. [3]: the $\phi(1020)$ region $0.98 < q^2 < 1.10$ GeV², the J/ψ region $8.0 < q^2 < 11.0$ GeV² and the $\psi(2S)$ region $12.5 < q^2 < 15.0$ GeV². These windows are much wider than the natural widths of the resonances because they must also contain the radiative tails. The remaining range $0.10 < q^2 < 22.0$ GeV² is divided into the six intervals of Table 3, chosen to respect the veto boundaries and to keep of order a hundred or more signal decays in each. This is a single common scheme fixed for the final analysis, not a blinded or preregistered one: earlier, exploratory fits in other intervals were performed during the recovery of the original analysis and informed the choice. What the scheme does guarantee is that every interval it defines is reported — none was dropped, split or merged after its result was seen.

The control samples are selected from $8.41 < q^2 < 10.24$ GeV² for $B^+ \rightarrow J/\psi K^+$, as in Ref. [3], and from $12.86 < q^2 < 14.18$ GeV² for $B^+ \rightarrow \psi(2S) K^+$.

4 Mass fits and extraction of the asymmetry

4.1 Parameterisation

In each q^2 interval the two charge samples are fitted simultaneously with an extended unbinned maximum-likelihood fit to the $K\mu^+\mu^-$ invariant mass over $5200 < m(K\mu^+\mu^-) < 5700$ MeV. The asymmetry of Eq. (2) is a parameter of the fit rather than a quantity formed afterwards:

the signal yields are written

$$N_{B^+} = \frac{1}{2}N_{\text{sig}}(1 + A_{\text{raw}}), \quad N_{B^-} = \frac{1}{2}N_{\text{sig}}(1 - A_{\text{raw}}), \quad (6)$$

so that the uncertainty on A_{raw} comes directly from the likelihood with the correlation between the two yields included, and can be quoted as a profile-likelihood interval instead of a propagated Gaussian error. The negative log-likelihood is

$$-\ln \mathcal{L} = \sum_{c=+,-} \left[N_c^{\text{tot}} - \sum_{i \in c} \ln (N_{\text{sig}}^c f_{\text{sig}}(m_i) + N_{\text{bkg}}^c f_{\text{bkg}}(m_i)) \right], \quad (7)$$

with both densities normalised over the finite fit window, so that the window is accounted for exactly rather than approximately.

The signal is modelled by two Gaussians sharing a mean,

$$f_{\text{sig}}(m) = (1 - f_2) \mathcal{G}(m; \mu, \sigma_1) + f_2 \mathcal{G}(m; \mu, \sigma_2), \quad (8)$$

a narrow core plus a wider component that absorbs the resolution tail, and the combinatorial background by a single exponential with an independent slope for each charge. The mean μ is shared between charges and left free; the three shape parameters σ_1 , σ_2 and f_2 are fixed to the values measured in the J/ψ control sample, which contains four orders of magnitude more decays than any rare-mode interval. Fixing an externally measured resolution in this way is the standard procedure [3]; leaving it free in a sample of a few hundred decays does not measure the resolution, it only lets statistical noise move the split between signal and background.

4.2 Control samples

Table 2 gives the control-sample results. The J/ψ control contains $(1.210 \pm 0.002) \times 10^6$ signal decays and yields

$$A_{\text{raw}}(B^+ \rightarrow J/\psi K^+) = +0.01367_{-0.00100}^{+0.00099} \text{ (stat)}, \quad (9)$$

and moves by at most 0.0004 over the alternative signal-shape, background and window configurations of Section 5: of the 7 tried on this sample, 1 converged with a valid interval; the other 6 did not converge with a valid interval and are excluded, and are listed in the saved output. By Eq. (4) this is, under the assumptions stated there, a measurement of $A_P + A_D$ for the $K^\pm \mu^+ \mu^-$ final state. Its magnitude is close to the published LHCb determination of the corresponding raw control asymmetry, quoted there as -0.015 and -0.013 for the two magnet polarities in the opposite charge ordering [3]. Because the two analyses apply different selections to differently reconstructed samples, exact numerical agreement is not expected; what the comparison does establish is that the sign convention of Eq. (2) is the one assumed here, since reversing the charge ordering reverses the sign of this single observable. The $\psi(2S)$ control, with 97.7×10^3 decays and $A_{\text{raw}} = +0.00968 \pm 0.00359$, agrees with the J/ψ value within 1.1 combined standard deviations, which is what is expected if the measured quantity is mainly a property of the $K^\pm \mu^+ \mu^-$ final state rather than of the dimuon system.

The shape parameters from the J/ψ control are $\sigma_1 = 16.90 \pm 0.05 \text{ MeV}$, $\sigma_2 = 36.97 \pm 0.45 \text{ MeV}$ and $f_2 = 0.218 \pm 0.004$, with a fitted mean of 5283.7 MeV . That mean lies about 4 MeV above the known $B^\pm$ mass. Its origin is not established by these fits, and no absolute mass calibration is inferred. The mean is floated in each rare-mode interval. The quoted parameter uncertainties are the statistical errors of this fit. They are not a complete statement of how well the resolution is known, because the fitted shape does not describe the control peak adequately: the joint deviance of Eq. (10) is 1196 for 70 degrees of freedom, with a largest single-bin pull of 12.3. A large sample does not excuse this; it is what makes the inadequacy visible. The consequence is that the resolution transferred to the rare-mode fits carries a model error that these statistical uncertainties do not measure, and the alternative configurations of Section 5 are the available, partial, probe of it.

Table 2. Control-sample fits. Uncertainties on A_{raw} are profile-likelihood intervals, statistical only. D is the joint two-charge deviance of Eq. (10) over 2×40 bins with 70 degrees of freedom: neither control peak is described adequately by the analytic shape fitted to it. The controls are therefore used to supply a resolution for the much smaller rare-mode samples, and are not themselves a demonstration of model adequacy.

Control mode	$N(B^+)$	$N(B^-)$	N_{sig}	A_{raw}	D
$B^+ \rightarrow J/\psi K^+$, $8.41 < q^2 < 10.24$	698,731	683,299	1.210×10^6	$+0.01367 \pm 0.00100$	1196
$B^+ \rightarrow \psi(2S)K^+$, $12.86 < q^2 < 14.18$	60,820	59,959	97.7×10^3	$+0.00968 \pm 0.00359$	166

4.3 Is a charge-independent resolution justified?

Fixing one resolution for both charges is an assumption, and the charmonium samples are large enough to examine it. Allowing the signal-shape parameters to differ between B^+ and B^- improves the J/ψ likelihood by $2\Delta \ln \mathcal{L} = 40.9$ for four extra parameters. Referred to a χ^2_4 distribution that corresponds to $p = 2.8 \times 10^{-8}$; the free fit is started from the charge-shared optimum, so the two models are nested at the minimiser and the likelihood difference is non-negative by construction. Taken at face value the test says that, *within this model*, the reconstructed response is not exactly charge independent — which is physically unsurprising, since opposite charges bend in opposite directions and traverse different material. It should not be over-read: the same fitted shape does not describe the control peak adequately (Section 4), and a likelihood-ratio test between two misspecified models measures the data’s preference between them, not the true size of a charge effect. On this sample the core width moves from 16.90 MeV when shared to 16.80 and 17.01 MeV when free, a relative difference of about 12 per mille, and the fitted asymmetry moves by 0.0019. In the $\psi(2S)$ control the same comparison gives $2\Delta \ln \mathcal{L} = 3.6$ for four parameters ($p = 0.46$) and a shift of 0.0053.

Those two numbers describe the charmonium samples. They do not bound the corresponding effect in the rare-mode intervals, which differ in q^2 , kinematics, signal-to-background ratio and sample size, and in which the reconstructed charge resolution may well differ; no closure test demonstrating such a transfer is available from these ntuples. When the same per-charge freedom is given to an interval containing a few hundred decays, the fitted widths differ by several MeV and the asymmetry moves by up to 0.28, far more than in the controls. Two effects are confounded there and cannot be separated with these data: the statistical fluctuation of parameters the sample cannot constrain, and genuine mismatch between the fitted shape and the data. We therefore adopt the shared, externally fixed resolution as the nominal model, because it is the configuration in which the small samples are not asked to measure a resolution; and we report the per-charge free fits in full (Section 5, Table 3 and Appendix B) as the broad end of the measured sensitivity, rather than discarding them or treating the charmonium shifts as a calibrated bound on them.

4.4 Fit quality and its calibration

Fit quality is assessed with the saturated Poisson deviance of the two charge histograms taken together, 40 bins each,

$$D = 2 \sum_{c,j} \left[\nu_{cj} - n_{cj} + n_{cj} \ln(n_{cj}/\nu_{cj}) \right], \quad (10)$$

where n_{cj} is the observed and ν_{cj} the expected content of bin j for charge c , the expectation being integrated across the bin rather than evaluated at its centre [10]. This is the appropriate statistic for Poisson bin contents; a Pearson χ^2 with the expectation taken at the bin centre can be badly misleading where bins are sparse.

Referring D to a χ^2 distribution would still be an approximation here, because the parameters are estimated from the unbinned likelihood, three of them are fixed externally, and

some bins hold few entries. We therefore calibrate. For each interval, 400 pseudo-experiments are generated from the fitted model and refitted with the identical configuration. A pseudo-experiment contributes to the reference distribution only if its own refit reached a valid minimum with an accurate covariance matrix, since a deviance read off a failed minimisation is not a draw from the statistic being calibrated; 400 of the 400 attempted refits per interval were accepted on that criterion. The calibrated p -value in Table 3 is the fraction of accepted pseudo-experiments with a deviance at least as large as the data, and it carries a Monte Carlo uncertainty of at most 0.025 from the finite number of pseudo-experiments. With this number of toys a p -value could not be resolved below about 0.0025; none of the observed values is near that floor.

Three limits of this procedure should be kept in view. The reference distribution is conditional on the fitted model being the true one, so it calibrates the sampling behaviour of D under that model rather than testing the model against a wider family. It treats candidates as independent draws, so it does not account for multiple candidates from the same event (Section 5.5). And it is a statement about the rare-mode fits only; the control fits are not calibrated this way, and Section 4 reports their deviance directly. No universal acceptance threshold is imposed: the p -values are reported interval by interval, together with the largest single-bin pull, so that each fit can be judged on its own.

5 Uncertainties and robustness

5.1 Statistical

The statistical uncertainty on A_{raw} is the profile-likelihood interval from the fit of Eq. (6), obtained by scanning the likelihood rather than from the curvature at the minimum, so it is asymmetric where the likelihood is. Every interval quoted below converged with a positive-definite covariance matrix and a valid profile-likelihood interval; configurations that failed either test are reported as failures rather than quietly replaced.

5.2 Sensitivity to the analysis configuration

No complete systematic budget is quoted in this paper, because the archived material does not support one: the acceptance, veto, efficiency and polarity corrections of Section 5.5 are missing outright, and their effect on A_{raw} is unknown rather than small. What can be reported is how much the measured number moves when a modelling choice that *is* available is changed. Ten alternative configurations are used: the signal shape replaced by the sum of two Crystal Ball functions determined from the same control; the background replaced by a second-order Chebyshev polynomial and by the sum of two exponentials; the window widened to 5170–5700 MeV and narrowed to 5200–5350 MeV; the resolution taken from the $\psi(2S)$ control instead of the J/ψ control; the control resolution shifted up and down by its own marginal uncertainty; the signal mean fixed to the control value instead of floating; and the signal shape left free for each charge separately. Appendix B lists every variation and every shift.

Two ranges are reported in Table 3, and they answer different questions. The *constrained-model* range is the largest absolute shift over the nine configurations that keep the resolution shared and externally determined; it reaches 0.109 and is dominated in most intervals by the narrow window, which deliberately leaves too little sideband to constrain the background slope and is retained to show how much that choice matters. The *full* range additionally includes the per-charge free signal shape, and reaches 0.281. Neither is a calibrated uncertainty, and neither should be added to the statistical interval as though it were one: they are observed ranges over a finite, correlated and incomplete set of choices. Changing the window also changes which background shape is preferred, so combining the shifts in quadrature would imply an independence they do not have.

The $\pm 1\sigma$ control-resolution variations move all three shape parameters together by their marginal errors. Because those parameters are strongly correlated in the control fit, this is a named directional sensitivity check and not a propagation of the control covariance; the shifts it produces, below 0.001 in every interval, should be read accordingly. Variations whose own fit did not reach a valid minimum with an accurate covariance are marked as failures and left out of both ranges, and the failures are listed explicitly in Appendix B rather than being silently dropped: the two-exponential background fails in four intervals, where the two slopes become degenerate, and the per-charge free shape fails in one.

5.3 Stability against the selection requirement

The selection requirement was varied by $\pm 10\%$ about its nominal value. The resulting samples are nested: the tighter sample is a subset of the nominal one, which is a subset of the looser one, so the two fitted asymmetries share most of the same candidates and are strongly positively correlated. Judging the difference against an independent-sample error would use $\sigma_1^2 + \sigma_2^2$ in place of $\sigma_1^2 + \sigma_2^2 - 2\text{cov}$, and with a positive covariance that *overstates* the variance of the difference and so tends to *understate* the significance of a genuine shift [9]. For nonlinear likelihood estimators the sign of the effect is not guaranteed, which is a further reason to measure the relevant variance rather than assume it. We therefore resample, with replacement, the candidates passing the loosest requirement, re-apply all three requirements to each replica and refit, so that the correlation is retained by construction; 250 replicas are used. Table 4 gives the shifts and their bootstrap spreads. The largest shift is 1.6 standard deviations of its own bootstrap distribution, so no dependence on the selection requirement is resolved at this precision and no additional uncertainty is assigned on this account. Two of the twelve variations could not be fitted at the tighter requirement because too few candidates remain; they are marked as such rather than omitted. The bootstrap resamples candidates, so it is conditional on candidates being independent draws; multiple candidates from the same event (Section 5.5) would correlate them, and that clustering is not modelled here.

5.4 Blindness of the selection to the measured charge

Section 3 established that the reconstructed charge is one of the classifier inputs. To bound the consequence, the selection was rebuilt so that its explicit charge dependence is removed exactly. Each candidate was scored twice with the model belonging to its own subsample, once with the charge input set to $+1$ and once to -1 with every other input untouched, and the selection applied to the average

$$s_{\text{sym}}(x) = \frac{1}{2} [s(x, +) + s(x, -)], \quad (11)$$

which by construction assigns the same score to a B^+ and a B^- candidate with identical kinematics. The rescoring covered every candidate with a stored score above 0.3 in the fit window. That prefilter cannot remove a candidate that would have passed: the two scores lie in $[0, 1]$, so $s_{\text{sym}} > 0.9$ requires $s(x, \text{as stored}) > 0.8$, which is far above the floor. Consistently, the lowest stored score among the symmetrically selected candidates is 0.876. The whole measurement was then repeated. The shifts in A_{raw} run from -0.0390 to $+0.0227$, the largest in magnitude being 0.0390, in every case well inside the statistical interval; in the J/ψ control, rescored on a quarter of the sample to keep the computation bounded, the shift is $+0.00031$. The per-interval shift is given in Table 3. This removes the explicit dependence; it cannot remove dependence entering implicitly through inputs correlated with the charge, which remains a limitation.

5.5 Corrections that are not applied

The following elements of a dedicated CP measurement cannot be reproduced from the reduced ntuples. Their absence is why this paper reports raw asymmetries.

- *Fiducial acceptance.* Candidates whose kaon is swept out of the acceptance in a charge-dependent way are removed in Ref. [3] by a requirement on the kaon momentum components. Those components are not retained in three of the four reduced files, so the requirement cannot be applied. This is the single most important missing correction for a charge asymmetry.
- *Peaking-background vetoes.* A dedicated analysis removes $B^+ \rightarrow D^0(\rightarrow K^+\pi^-)\pi^+$ with pions misidentified as muons, and $B \rightarrow J/\psi K$ decays in which a muon and a hadron are mistaken for each other. The second is charge-sensitive, because the kaon charge tags the B flavour. Neither veto can be applied without the relevant mass combinations and particle-identification variables.
- *Efficiency corrections.* No simulated sample accompanies the reduced ntuples, so no efficiency can be computed and no efficiency-weighted or luminosity-normalised quantity is quoted.
- *Per-polarity fits.* Ref. [3] fits each magnet polarity separately and averages, which cancels the left-right part of A_D inside the fit. Only one of the four reduced files retains its polarity label, so the fits here use the polarity-summed sample. That sample is approximately balanced between polarities, but the cancellation is not enforced.
- *Multiple candidates.* No best-candidate requirement is applied. In the one surviving parent ntuple, 16.9% of rows are not the first candidate of their event, so the fraction of rows belonging to multiple-candidate events is somewhat larger than that. Ref. [3] assigns no duplicate-candidate uncertainty for this channel, but the effect is not quantified here.
- *Control-mode kinematics.* Eq. (4) assumes the control and signal modes share production and detection asymmetries. The control muons come from a J/ψ decay, so their kinematics differ somewhat. Reweighting the control sample in B transverse momentum to match each signal interval changes $A_{\text{raw}}(B^+ \rightarrow J/\psi K^+)$ by less than 10^{-4} , but that is a one-variable check, not a demonstration that the matching is adequate: a full treatment would reweight in several variables using background-subtracted signal kinematics, which these ntuples do not support. The validity of the control subtraction is therefore not established.
- *Control-shape model error.* The analytic shape fitted to the J/ψ control does not describe it adequately (Section 4), and the resolution it supplies is transferred to intervals at other q^2 without a closure test. The resulting model error on the rare-mode yields is probed only by the configuration changes of Section 5 and is not quantified.

For these reasons the difference between a measured A_{raw} and the control value appears in Section 6 only as a conditional diagnostic, no CP asymmetry is extracted, and nothing in this paper constitutes a test of the Standard Model.

6 Results

6.1 Raw charge asymmetries

Table 3 and Figure 3 give the raw charge asymmetry in the six q^2 intervals. Figures 4 and 5 show the control fits and a representative rare-mode fit with their pulls.

The six values scatter about the control value $A_{\text{raw}}(B^+ \rightarrow J/\psi K^+) = +0.0137$. Two of the six lie within one statistical standard deviation of it; the largest departure is $11.0\text{--}12.5\text{ GeV}^2$, at 1.6 standard deviations of its own statistical interval, which among six intervals is not a notable outlier. The appropriate statement is therefore that no compelling discrepancy with the control is resolved at this precision — which is weaker than saying that every interval agrees

Table 3. Raw charge asymmetry of $B^\pm \rightarrow K^\pm \mu^+ \mu^-$ in intervals of q^2 . $N(B^+)$ and $N(B^-)$ are candidate counts in the fit window and N_{sig} the fitted total signal yield. The uncertainty on A_{raw} is the profile-likelihood interval, statistical only. The two ranges are observed sensitivities, not uncertainties, and must not be added to the statistical interval: “constrained” is the largest shift over the converged variations that keep the resolution shared and externally fixed, and “all” additionally includes the per-charge free signal shape (Sections 4.3 and 5; the full list, including failed variations, is in Appendix B). “Charge-blind” is the shift when the selection is made independent of the reconstructed charge. The last column is the calibrated fit p -value from refitted pseudo-experiments, conditional on the fitted model.

q^2 [GeV ²]	$N(B^+)$	$N(B^-)$	N_{sig}	A_{raw}	Sensitivity range		Charge-blind	p
					constrained	all		
0.10–0.98	221	243	152 ± 18	$-0.158^{+0.115}_{-0.116}$	0.019	0.281	$+0.023$	0.68
1.10–4.00	1,553	1,402	485 ± 40	$+0.109^{+0.083}_{-0.082}$	0.086	0.126	-0.011	0.17
4.00–8.00	1,798	1,869	665 ± 46	$-0.057^{+0.070}_{-0.069}$	0.014	0.103	$+0.010$	0.57
11.0–12.5	984	939	209 ± 30	$+0.250^{+0.151}_{-0.145}$	0.109	0.109^f	-0.039	0.16
15.0–19.0	2,342	2,413	609 ± 47	$+0.016^{+0.077}_{-0.077}$	0.022	0.028	$+0.006$	0.31
19.0–22.0	1,480	1,645	192 ± 37	$+0.105^{+0.203}_{-0.191}$	0.098	0.098	$+0.002$	0.77

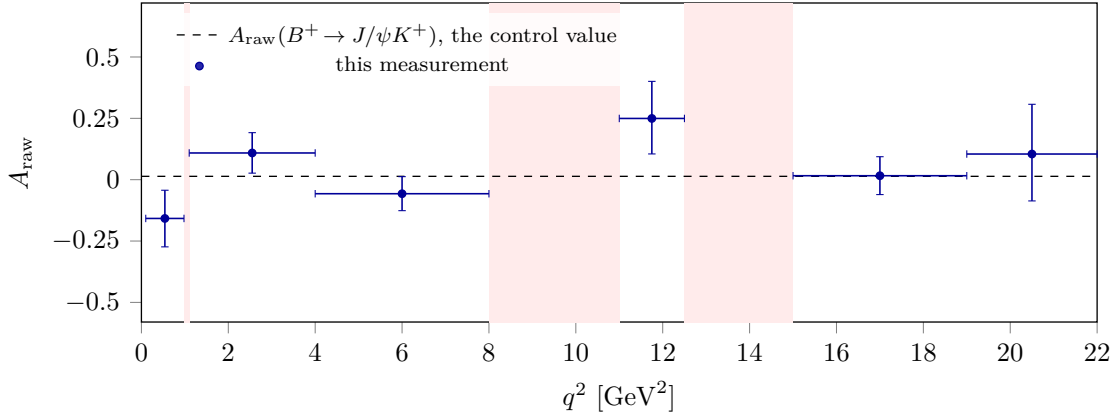


Figure 3. Raw charge asymmetry against q^2 . Vertical bars are profile-likelihood intervals (statistical only), horizontal bars the interval widths. The dashed line is the control-channel value. Under Eq. (4) that value measures $A_P + A_D$, and the rare mode would reproduce it if A_{CP} were negligible *and* if the two modes shared their production, detection and selection asymmetries — an assumption that is not established here (Section 5.5). Shaded bands are the excluded resonance regions.

with it, and much weaker than a test of A_{CP} : by Eq. (3) the rare-mode raw asymmetry would reproduce the control value if A_{CP} were negligible *and* if the production, detection and selection asymmetries of the two modes were equal, and the second condition is not demonstrated here (Section 5.5). The fitted signal mean in each interval agrees with the control value to within a few MeV except in the highest interval, where the yield is smallest and the mean is least constrained (Appendix B).

6.2 Relation to the published CP measurement

LHCb has published an independent analysis of its Run-1 $B^+ \rightarrow K^+ \mu^+ \mu^-$ sample with the full offline selection, simulation-based efficiency corrections, the fiducial requirement and the peaking-background vetoes listed in Section 5.5, giving $A_{CP}(B^+ \rightarrow K^+ \mu^+ \mu^-) = 0.012 \pm 0.017 \pm 0.001$ averaged over q^2 [3]. The reduced ntuples used here are drawn from the same Run-1 data taking, but their exact provenance, run coverage and reconstruction version cannot be established from the surviving metadata (Section 2), so this is a comparison with an independent published analysis of the same data-taking period and not a remeasurement of an identical sample. A direct comparison with Table 3 would in any case be wrong on two counts: the

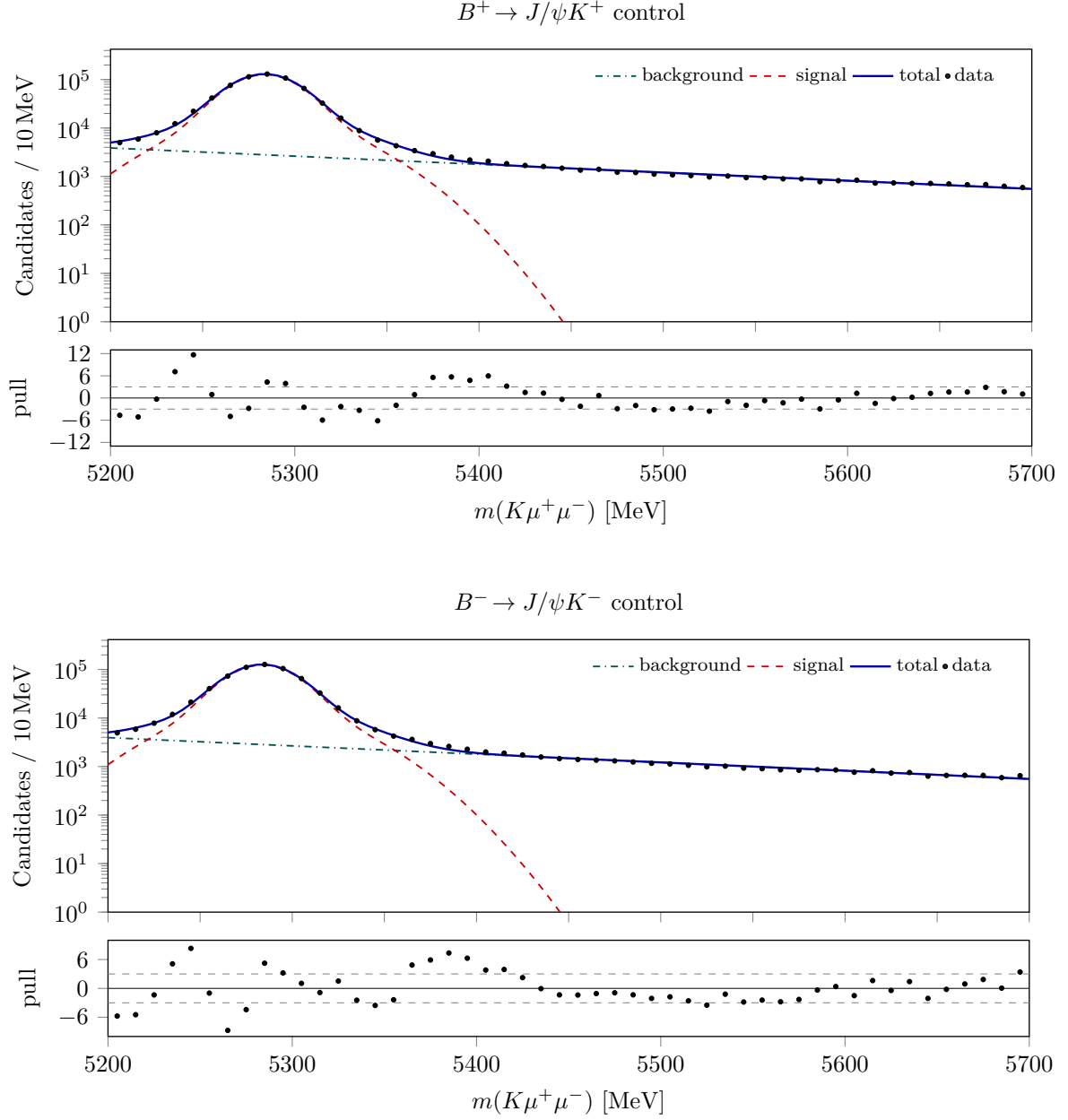


Figure 4. Invariant-mass fits to the $B^+ \rightarrow J/\psi K^+$ (top) and $B^- \rightarrow J/\psi K^-$ (bottom) control samples, on a logarithmic scale. These fits determine the signal resolution used for the rare-mode intervals. Points are data with Poisson errors, the solid curve the total model, the dashed curve the signal and the dash-dotted curve the background; the sub-panel below each fit shows the pulls $(n - \nu)/\sqrt{\nu}$ on an axis wide enough to contain them all, with dashed guides at ± 3 . The pulls reach 12.3 and the joint deviance is 1196 for 70 degrees of freedom: this analytic shape does not describe the control peak adequately, and the large sample size is what makes that visible rather than an excuse for it. The consequence for the rare-mode fits is discussed in Sections 4 and 5.5.

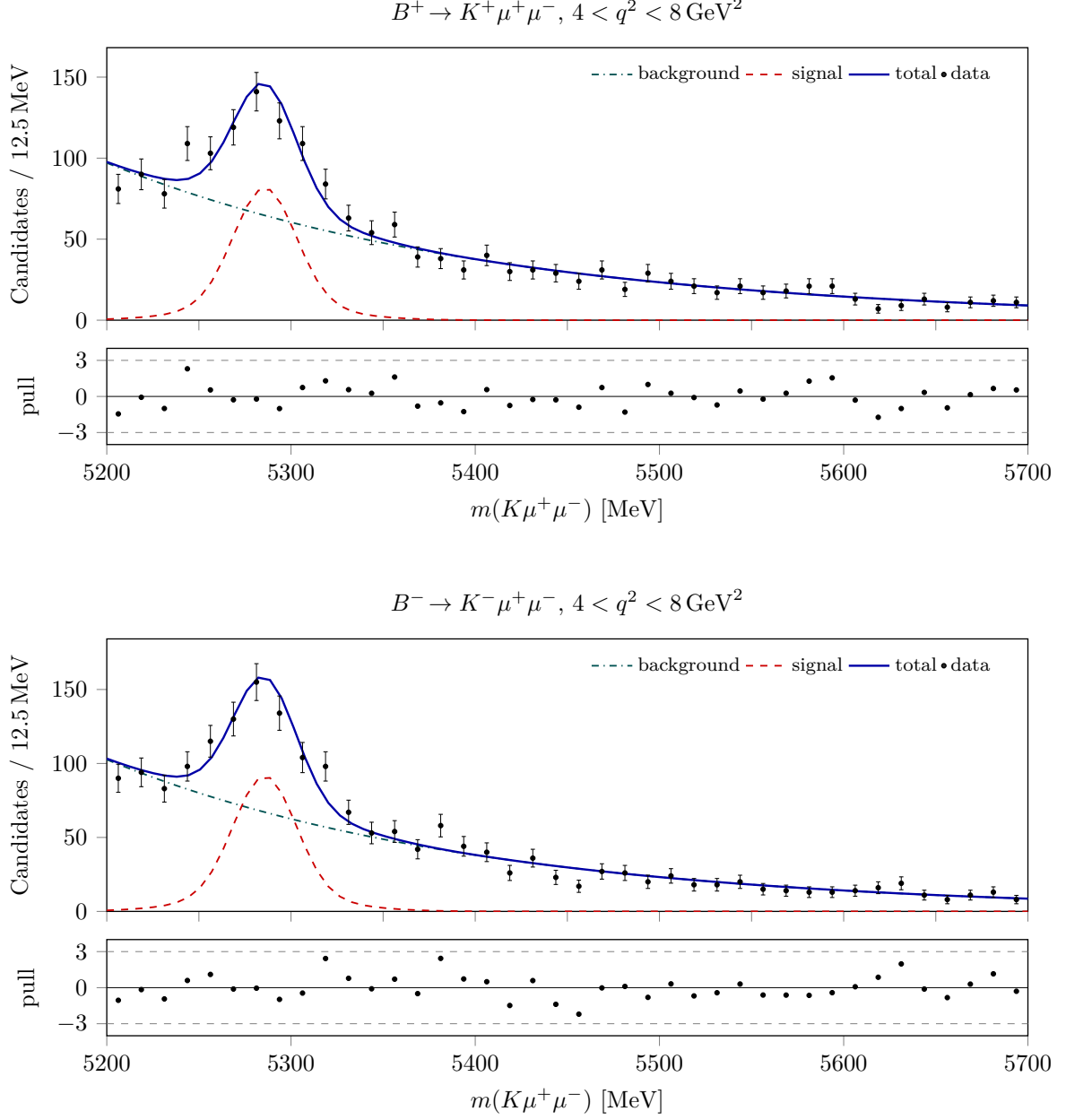


Figure 5. Invariant-mass fits to the $4 < q^2 < 8 \text{ GeV}^2$ interval, B^+ (top) and B^- (bottom), representative of the six rare-mode fits. Curves, points and pull sub-panels are as in Figure 4; the pull axis again spans the observed residuals, whose largest absolute value in this interval is 2.43. The calibrated p -value for this interval is 0.57, and Table 6 gives the corresponding diagnostics for all six intervals. The signal resolution is fixed from the control fits of Figure 4 and is shared between the charges, an assumption examined in Section 4.3.

Table 4. Stability of A_{raw} against a $\pm 10\%$ change of the selection requirement. Each entry is the shift with respect to the nominal requirement, its spread over 250 paired bootstrap replicas of the same candidates, and the shift in units of that spread. The paired bootstrap retains the strong correlation between the nested samples, so these spreads are the appropriate scale for the shifts, in place of an independent-sample error. They remain conditional on candidates being independent draws (Section 5).

q^2 [GeV ²]	looser requirement	tighter requirement
0.10–0.98	-0.001 ± 0.087 (-0.0σ)	too few candidates
1.10–4.00	-0.075 ± 0.054 (-1.4σ)	$+0.004 \pm 0.140$ ($+0.0\sigma$)
4.00–8.00	$+0.043 \pm 0.049$ ($+0.9\sigma$)	-0.088 ± 0.145 (-0.6σ)
11.0–12.5	-0.080 ± 0.117 (-0.7σ)	-0.410 ± 0.250 (-1.6σ)
15.0–19.0	$+0.033 \pm 0.069$ ($+0.5\sigma$)	$+0.020 \pm 0.117$ ($+0.2\sigma$)
19.0–22.0	$+0.034 \pm 0.233$ ($+0.1\sigma$)	too few candidates

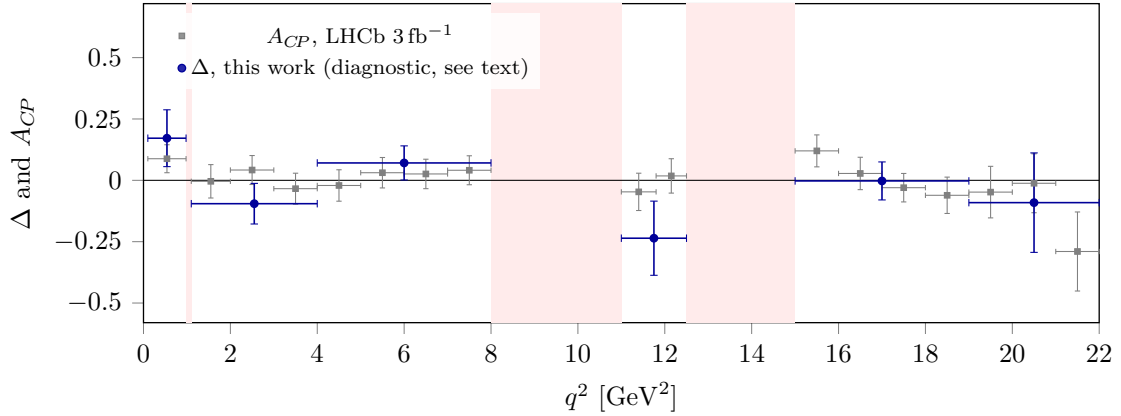


Figure 6. The control-subtracted, sign-matched diagnostic Δ of Eq. (12) compared with the published per-interval A_{CP} of Ref. [3], which analysed LHCb Run-1 data independently. The intervals are not the same, and the points are not two measurements of one quantity: Δ is *not* a CP asymmetry, because the corrections of Section 5.5 are not applied and the exactness of the control subtraction is not established. Uncertainties on Δ are statistical only, combining the interval and the control; the sensitivity ranges of Table 3 are not shown. The comparison establishes consistency of convention and of scale, not competitive precision.

quantities differ by the production and detection asymmetries, and they are defined with opposite ordering of the two charges.

Figure 6 therefore shows the diagnostic combination

$$\Delta \equiv -[A_{\text{raw}}(q^2) - A_{\text{raw}}(B^+ \rightarrow J/\psi K^+)], \quad (12)$$

which by Eqs. (3) and (4) would equal A_{CP} in the published convention if the control channel corrected the nuisance terms exactly and if the corrections of Section 5.5 were negligible. Neither condition is established here, so Δ is a conditional diagnostic and not a measurement of A_{CP} : the control matching is checked in one kinematic variable only, the rare-mode selection is not charge-symmetric by construction, and the acceptance, veto, efficiency and polarity corrections are absent. What the comparison shows is that, once the charge ordering is matched and the control asymmetry subtracted, the quantities obtained here sit on the same scale as the published per-interval A_{CP} and are compatible with them at the much coarser precision reached here. It is deliberately not averaged over q^2 into a single number: averaging raw asymmetries over different regions of phase space without efficiency weights would not produce the quantity Ref. [3] quotes.

7 Discussion

The six raw asymmetries scatter about the control-channel value without a compelling departure from it at the precision reached, and the control-channel value is of the same magnitude as the published raw control asymmetry. It is worth being precise about what that does and does not amount to. The quantities measured here are raw asymmetries containing production, detection and selection effects that have not been corrected; the Standard Model prediction $|A_{CP}| \sim 10^{-3}$ [4] is two orders of magnitude below the statistical precision of any interval, and would be invisible here even if every correction had been applied. Consistency with a control value is therefore not a test of the Standard Model, and neither supports nor constrains the enhancements some extensions predict [5]. What the result does establish is an internally consistent, traceable raw measurement with its assumptions stated.

The selection has a strong dependence on q^2 . Figure 1 measures the fraction of all input candidates retained, whereas the efficiency quoted for the dedicated classifier in Ref. [3] concerns signal. These quantities have different denominators and cannot establish how much of the precision difference between the analyses is due to the selection. The training labels are based on a mass window and an upper sideband, with dimuon mass among the inputs; the positive training class is dominated by charmonium. This design motivates the explicit selection-sensitivity checks, but a true-signal efficiency comparison would require additional validation. A classifier intended for this measurement should be trained with a truth-matched or background-subtracted positive class, and should be given neither the variable in which the results are binned nor the charge whose asymmetry is measured. It should also be evaluated on a sample held out from both the early stopping and the hyper-parameter search, which was not the case here.

A second observation concerns how a resolution should be constrained, and it is the main unresolved methodological issue in this analysis. In a sample of a few hundred decays, allowing the signal shape to float independently for the two charges moves the fitted asymmetry by up to 0.28, larger than the statistical interval; in the two charmonium samples, with 10^5 and 10^6 decays, the corresponding shifts are 0.0019 and 0.0053. It is tempting to read the charmonium numbers as a bound on the rare-mode ones and to call the difference statistical noise, and we do not do so: the charmonium samples differ from the rare-mode intervals in q^2 , kinematics and signal-to-background ratio, no closure test of the transfer is available, and the likelihood-ratio test of Section 4.3 indicates that some charge dependence of the reconstructed shape is present under this model. The defensible position is narrower. A parameter that a sample of a few hundred decays cannot determine will be driven by fluctuations and by any mismatch between model and data if it is left free, so the nominal fit constrains it externally; and because that constraint cannot be validated for these intervals, the shifts it suppresses are reported in full as the broad end of the sensitivity range rather than removed from view. Both numbers are given in Table 3 for every interval.

Finally, the comparison of Section 6.2 becomes meaningful only after two pieces of book-keeping that are easy to omit: matching the ordering of the two charges, and subtracting the production and detection asymmetries measured in a control channel. Neither step is a correction for new physics, and neither is a single universal number: the ordering flip changes the sign of whatever is being compared, and the control subtraction removes a quantity measured here as +0.0137, which is itself specific to this final state, this selection and this sample.

8 Conclusions

The raw charge asymmetry of $B^\pm \rightarrow K^\pm \mu^+ \mu^-$ decays has been measured in six intervals of q^2 between 0.10 and 22.0 GeV² using reduced LHCb Run-1 ntuples, together with the asymmetry of the $B^+ \rightarrow J/\psi K^+$ control mode. The control mode gives $A_{\text{raw}}(B^+ \rightarrow J/\psi K^+) = +0.01367 \pm 0.00100$ from 1.21×10^6 decays, consistent in magnitude with the published LHCb

determination of the corresponding quantity. Under Eq. (4) this is a measurement of the combined production and detection asymmetry of the final state. The six rare-mode asymmetries lie between -0.158 and $+0.250$ with statistical intervals of 0.07 – 0.20 ; two lie within one statistical standard deviation of the control value and the largest departure is 1.6σ , so no compelling discrepancy with the control is resolved at this precision. The rare-mode fits are supported by refitted pseudo-experiments conditional on the fitted model, with calibrated p -values between 0.16 and 0.77 ; the control fits, by contrast, are not described adequately by the analytic shape used and supply only a resolution.

Because the reduced ntuples do not retain the information needed for the fiducial acceptance requirement, the peaking-background vetoes, any efficiency correction or a per-polarity fit, these are raw asymmetries and not CP asymmetries, and no q^2 -averaged CP asymmetry is quoted. The control-subtracted diagnostic of Eq. (12) sits on the same scale as the published per-interval A_{CP} once the sign convention and the nuisance asymmetries are accounted for, which is a consistency check of conventions and scale. It is not a CP measurement, and this paper makes no statement about the Standard Model or about physics beyond it.

These limitations belong to this analysis, not only to the archive it works from, and they are what a future version would have to remove. Doing so would require a selection trained without the binning variable or the measured charge and validated on a genuinely independent sample; retention of the momentum components required by the fiducial requirement and of the particle-identification variables required by the vetoes; a simulated sample for the efficiency; separate fits per magnet polarity; a signal model that describes the high-statistics control peak; and a demonstrated transfer of the resolution across q^2 . Until then the defensible product is what is reported here: a reproducible raw-yield measurement whose assumptions, failures and sensitivities are all on the page.

Acknowledgements

I thank Adam Fontinelle for his collaboration on the original research project. I thank my project supervisors, Dr Michael McCann and Professor Mitesh Patel, for supplying the reduced data samples on which this work is based and the analysis library from which the original fitting and classification code was developed. This work uses data recorded by the LHCb experiment at CERN. The published LHCb measurement used for comparison is taken from Ref. [3], distributed under CC-BY-4.0.

Use of software assistance

Large language models were used in preparing this manuscript: to help draft and edit prose, and to help write, audit and re-run analysis code. All numerical results were produced by code that forms part of the analysis record, every number in the tables and figures is read directly from its saved output, and the author is responsible for the content.

Data and code availability

The analysis runs as a sequence of scripts from the reduced ntuples to every table and figure in this paper; the random seed, the number of pseudo-experiments and the number of bootstrap replicas are recorded in the output files. The reduced ntuples are not included with this report. LHCb has since released its Run-1 proton–proton dataset publicly [12].

References

- [1] Particle Data Group, S. Navas *et al.*, *Review of Particle Physics*, Phys. Rev. D **110** (2024) 030001.

- [2] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay*, JHEP **02** (2013) 105, arXiv:1209.4284.
- [3] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP asymmetries in the decays $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ and $B^+ \rightarrow K^+ \mu^+ \mu^-$* , JHEP **09** (2014) 177, doi:10.1007/JHEP09(2014)177, arXiv:1408.0978.
- [4] W. Altmannshofer, P. Ball, A. Bharucha, A. J. Buras, D. M. Straub and M. Wick, *Symmetries and asymmetries of $B \rightarrow K^* \mu^+ \mu^-$ decays in the Standard Model and beyond*, JHEP **01** (2009) 019, arXiv:0811.1214.
- [5] A. K. Alok, A. Datta, A. Dighe, M. Duraissamy, D. Ghosh and D. London, *New physics in $b \rightarrow s \mu^+ \mu^-$: CP-violating observables*, JHEP **11** (2011) 122, arXiv:1103.5344.
- [6] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.
- [7] T. Chen and C. Guestrin, *XGBoost: a scalable tree boosting system*, Proc. 22nd ACM SIGKDD (2016) 785, arXiv:1603.02754.
- [8] F. James and M. Roos, *Minuit: a system for function minimization and analysis of the parameter errors and correlations*, Comput. Phys. Commun. **10** (1975) 343; H. Dembinski *et al.*, *iminuit*, doi:10.5281/zenodo.3949207.
- [9] R. Barlow, *Systematic errors: facts and fictions*, Proc. Advanced Statistical Techniques in Particle Physics, Durham (2002) 134, arXiv:hep-ex/0207026.
- [10] S. Baker and R. D. Cousins, *Clarification of the use of chi-square and likelihood functions in fits to histograms*, Nucl. Instrum. Meth. **221** (1984) 437.
- [11] J. Pivarski *et al.*, *uproot: ROOT I/O in pure Python and NumPy*, doi:10.5281/zenodo.4340632.
- [12] LHCb collaboration, *LHCb releases the entire Run-1 proton–proton dataset*, CERN Open Data Portal (2023), opendata.cern.ch/docs/lhcb-releases-entire-run1-dataset.

A Likelihood, minimisation and resampling

The extended unbinned likelihood of Eq. (7) is minimised with MINUIT through the IMINUIT interface [8], using the likelihood convention $\Delta \ln \mathcal{L} = 0.5$ for one standard deviation. Both component densities are normalised over the finite window: the Gaussians analytically through the error function, the exponential analytically, and the Chebyshev alternative by fixed-order Gauss–Legendre quadrature with 512 nodes, which over a 500 MeV window places more than twenty nodes per resolution width. For the charmonium controls, where each charge holds 10^5 – 10^6 candidates, the unbinned likelihood is replaced by the extended Poisson likelihood of a 300-bin histogram with bin-integrated expectations; the 1.7 MeV bin width is an order of magnitude below the mass resolution, so the information loss is negligible while the cost becomes independent of sample size. Data are read from the ROOT files with UPROOT [11].

Parameter limits are $5260 < \mu < 5300$ MeV, $5 < \sigma_1 < 50$ MeV, $10 < \sigma_2 < 200$ MeV, $0 < f_2 < 1$, $-1 < A_{\text{raw}} < 1$ and non-negative yields. A fit is used only if MIGRAD reports a valid minimum, HESSE returns an accurate covariance matrix and the MINOS interval on A_{raw} is valid. Every free parameter of every fit entering Table 3 was checked against its bounds and none was found at a bound.

The deviance of Eq. (10) is the Poisson likelihood-ratio statistic of Ref. [10]. Its calibration uses 400 pseudo-experiments per interval: the number of candidates of each charge is drawn from a Poisson distribution with the fitted mean, masses are drawn from the fitted mixture by inverse-transform sampling on a 4001-point grid, and each pseudo-experiment is refitted with the identical configuration, including the externally fixed resolution. Pseudo-experiments whose refit does not itself reach a valid minimum with an accurate covariance are counted and excluded rather than contributing a deviance read off a failed minimisation; Table 6 reports

the accepted and rejected counts per interval. The charge-shared versus charge-free likelihood-ratio tests of Section 4.3 start the free fit from the charge-shared optimum, so the fitted likelihood of the larger model cannot be worse than that of the smaller one; the difference was confirmed to be non-negative in both control samples. The selection-stability bootstrap of Table 4 resamples, with replacement, the candidates passing the loosest requirement and then applies all three requirements to the same replica; this is the paired construction appropriate to nested selections, for which the difference of independent-sample variances is not a reliable scale [9]. All random numbers come from one generator seeded with 20260101.

B Analysis variations and fit diagnostics

Table 5 lists the shift in A_{raw} for every variation in every interval, including the per-charge free signal shape that is excluded from the constrained range but included in the full range of Table 3 for the reasons given in Section 4.3. Entries marked f denote configurations whose own fit did not reach a valid minimum with an accurate covariance matrix; they are shown rather than omitted, but contribute to neither range. Table 6 collects the fit diagnostics.

Table 5. Shift in A_{raw} for each analysis variation, by q^2 interval in GeV^2 . f : the variation’s own fit did not converge with an accurate covariance and is excluded from both ranges of Table 3.

Variation	0.10–0.98	1.10–4.00	4.00–8.00	11.0–12.5	15.0–19.0	19.0–22.0
signal shape: two Crystal Ball functions	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
background: second-order Chebyshev	−0.019	−0.004	<0.001	+0.006	−0.017	+0.018
background: two exponentials	+0.006	f	f	<0.001	f	−0.098
window 5170–5700 MeV	+0.002	+0.002	−0.002	−0.066	−0.019	+0.029
window 5200–5350 MeV	+0.002	−0.086	−0.014	+0.109	+0.022	−0.045
resolution from the $\psi(2S)$ control	−0.001	−0.004	−0.002	+0.003	+0.001	<0.001
control resolution $+1\sigma$	+0.001	<0.001	<0.001	<0.001	<0.001	<0.001
control resolution -1σ	−0.001	<0.001	<0.001	<0.001	<0.001	<0.001
signal mean fixed to the control value	+0.011	−0.003	<0.001	−0.002	+0.002	−0.019
resolution free for each charge	+0.281	−0.126	−0.103	f	+0.028	+0.042

Table 6. Fit diagnostics. The deviance of Eq. (10) is computed over 2×40 bins; the asymptotic p -value refers it to a χ^2 distribution with the quoted degrees of freedom, the calibrated p -value to the distribution from refitted pseudo-experiments, with its Monte Carlo uncertainty from the finite number of toys. “Toys” gives the number accepted and the number attempted: a pseudo-experiment is accepted only if its own refit converged with an accurate covariance matrix. The fitted signal mean is given for reference. The control peak lies about 4 MeV above the known $B^\pm$ mass, but the origin of this offset is unestablished; these fits do not provide an absolute mass calibration.

q^2 [GeV^2]	D	dof	p (asymptotic)	p (calibrated)	Toys	max. pull	μ [MeV]
0.10–0.98	73.1	73	0.474	0.677 ± 0.023	400/400	2.45	5287.3
1.10–4.00	86.3	73	0.137	0.175 ± 0.019	400/400	3.08	5286.1
4.00–8.00	71.0	73	0.546	0.575 ± 0.025	400/400	2.43	5285.7
11.0–12.5	87.2	73	0.123	0.158 ± 0.018	400/400	2.73	5288.8
15.0–19.0	78.7	73	0.303	0.312 ± 0.023	400/400	2.45	5284.8
19.0–22.0	64.1	73	0.762	0.765 ± 0.021	400/400	2.10	5275.8