Measurement of the $B^0_s \to J/\psi K^0_S$ branching fraction\*  

LHCb Collaboration

**Abstract**  
The $B^0_s \to J/\psi K^0_S$ branching fraction is measured in a data sample corresponding to 0.41 fb$^{-1}$ of integrated luminosity collected with the LHCb detector at the LHC. This channel is sensitive to the penguin contributions affecting the $\sin2\beta$ measurement from $B^0 \to J/\psi K_S^0$. The time-integrated branching fraction is measured to be $\mathcal{B}(B^0_s \to J/\psi K^0_S) = (1.83 \pm 0.28) \times 10^{-3}$. This is the most precise measurement to date.

© 2012 CERN. Published by Elsevier B.V. All rights reserved.

1. Introduction

In the Standard Model (SM) CP violation arises through a single phase in the quark mixing matrix [1]. In decays of neutral $B$ mesons to a final state which is accessible to both $B$ and $\bar{B}$, the interference between the amplitude for the direct decay and the amplitude for decay via oscillation leads to a time-dependent CP-violating asymmetry between the decay time distributions of the two mesons. The mode $B^0 \to J/\psi K^0_S$ allows for the measurement of such an asymmetry, which is parametrised by the $B^0\to\bar{B}^0$ mixing phase $\phi_d$. In the SM this phase is equal to $2\beta$ [2], where $\beta$ is one of the angles of the unitarity triangle of the mixing matrix. This phase is already measured by the $B$ factories [3] but an improved measurement is necessary to resolve conclusively the present tension in the unitarity triangle fits [4] and determine possible small deviations from the SM value. To achieve the required precision, knowledge of the doubly Cabibbo-suppressed higher order perturbative corrections, known as penguin diagrams, becomes mandatory. The contributions of these penguin diagrams are difficult to calculate reliably and therefore need to be extracted directly from experimentally accessible observables. Due to SU(3) flavour symmetry, these penguin diagrams can be studied in other decay modes where they are not suppressed relative to the tree level diagram. The $B^0 \to J/\psi K^0_S$ mode is the most promising candidate from the theoretical perspective since it is related to the $B^0 \to J/\psi K_S^0$ mode through the interchange of all $d$ and $s$ quarks ($U$-spin symmetry, a subgroup of SU(3)) [5] and there is a one-to-one correspondence between all decay topologies in these modes, as illustrated in Fig. 1. A further discussion regarding the theory of this decay and its potential at LHCb is given in Ref. [6].

To extract the parameters related to penguin contributions in these decays, a time-dependent CP violation study of the $B^0\to J/\psi K_S^0$ mode is required. The measurement of its branching fraction is an important first step, allowing to test the $U$-spin symmetry assumption that lies at the basis of the proposed approach. The CDF Collaboration reported the first observation of the $B^0_s \to J/\psi K^0_S$ decay [7]. This Letter presents a more precise measurement of this branching fraction at the LHCb experiment.

The strategy of the analysis is to measure the ratio of $B^0_s \to J/\psi K^0_S$ and $B^0 \to J/\psi K^0_S$ event yields, which is then converted into a $B^0_s \to J/\psi K^0_S$ branching fraction. We make use of the $B^0_s \to J/\psi K^0_S$ branching fraction and of the ratio of $B^0_s$ to $B^0$ meson production at the LHC, denoted $f_s/f_d$ [8].

We use an integrated luminosity of 0.41 fb$^{-1}$ of $pp$ collision data recorded at a centre-of-mass energy of 7 TeV during 2010 and the first half of 2011. The detector [9] is a single-arm spectrometer designed to study particles containing $b$ or $c$ quarks. It includes a high precision tracking system consisting of a silicon-strip vertex detector surrounding the $pp$ interaction region, a large-area silicon-strip detector located upstream of a dipole magnet with a bending power of about 4 Tm, and three stations of silicon-strip detectors and straw drift-tubes placed downstream. The combined tracking system has a momentum resolution $\Delta p/p$ that varies from 0.4% at 5 GeV/$c$ to 0.6% at 100 GeV/$c$, and an impact parameter resolution of 20 $\mu$m for tracks with high transverse momentum. Charged hadrons are identified using two ring-imaging Cherenkov (RICH) detectors. Muons are identified by a muon system composed of alternating layers of iron and multiwire proportional chambers.

The signal simulation sample used for this analysis was generated using the PYTHIA 6.4 generator [10] configured with the parameters detailed in Ref. [11]. The EVTGEN [12], PHOTOS [13] and GEANT4 [14] packages were used to decay unstable particles,
generate QED radiative corrections and simulate interactions in the detector, respectively.

2. Data samples and selection

We search for $B \rightarrow J/\psi K^0_S$ decays\(^1\) where $J/\psi \rightarrow \mu^+\mu^-$ and $K^0_S \rightarrow \pi^+\pi^-$. Events are selected by a trigger system consisting of a hardware trigger, which requires muon or hadron candidates with high transverse momentum with respect to the beam direction, $p_T$, followed by a two stage software trigger [15]. In the first stage a simplified event reconstruction is applied. Events are required to have either two oppositely charged muons with combined mass above 2.7 GeV/c\(^2\), or at least one muon or one high-$p_T$ track ($p_T > 1.8$ GeV/c) with a large impact parameter with respect to any primary vertex. In the second stage a full event reconstruction is performed and only events containing $J/\psi \rightarrow \mu^+\mu^-$ candidates are retained.

In order to reduce the data to a manageable level, very loose requirements are applied to suppress background while keeping the signal efficiency high. $J/\psi$ candidates are created from pairs of oppositely charged muons that have a common vertex and a mass in the range 3030–3150 MeV. The latter corresponds to about eight times the $\mu^+\mu^-$ mass resolution at the $J/\psi$ mass and covers part of the $J/\psi$ radiative tail. The $K^0_S$ selection requires two oppositely charged particles reconstructed in the tracking stations on either side of the magnet, both with hits in the vertex detector (long $K^0_S$ candidate) or not (downstream $K^0_S$ candidate). The $K^0_S$ candidates must be made of tracks forming a common vertex and have a mass within eight standard deviations of the $K^0_S$ mass and must not be compatible with the $\Lambda$ mass under the mass hypothesis that one of the two tracks is a proton and the other a pion.

We select $B$ candidates from combinations of $J/\psi$ and $K^0_S$ candidates with mass $m_{J/\psi K^0_S}$ in the range 5200–5500 MeV/c\(^2\). The latter is computed with the masses of the $\mu^+\mu^-$ and $\pi^+\pi^-$ pairs constrained to the $J/\psi$ and $K^0_S$ masses, respectively. The mass and decay time of the $B$ are obtained from a decay chain fit [16] that in addition constrains the $B$ candidate to originate from the primary vertex. The $\chi^2$ of the fit, which has eight degrees of freedom, is required to be less than 128 and the estimated uncertainty on the $B$ mass must not exceed 30 MeV/c\(^2\). $B$ candidates are required to have a decay time larger than 0.2 ps and $K^0_S$ candidates to have a flight distance larger than five times its uncertainty. The offline selected signal candidate is required to be that used for the trigger decision at both software trigger stages. About 1% of the selected events have several candidates sharing some final state particles. In such cases one candidate per event is selected randomly.

3. Measurement of event yields

Following the selection described above, a neural network (NN) classifier [17] is used to further discriminate between signal and background. The NN is trained entirely on data, using samples that are independent of those used to make the measurements. The training maximises the separation of signal and background events using weights determined by the sPlot technique [18]. We use the $B^0 \rightarrow J/\psi K^0_S$ signal in the data as a proxy for the $B^0 \rightarrow J/\psi K^0_S$ decay. The background events are taken from mass sidebands in the region 5390–5500 MeV/c\(^2\), thus avoiding the $B^0$ signal region. A normalisation sample of one quarter of the candidates randomly selected is left out in the NN training to allow an unbiased measurement of the $B^0$ yield.

We perform an unbinned maximum likelihood fit to the mass distribution of the selected candidates, shown in Fig. 2, and use it to assign background and signal weights to each candidate. The probability density function (PDF) is defined as the sum of a $B^0$ signal component, a combinatorial background and a small contribution from partially reconstructed $B \rightarrow J/\psi K^0_S X$ decays at masses below the $B^0$ mass. The mass lineshape of the $B^0 \rightarrow J/\psi K^0_S$ signal in both data and simulation exhibits non-Gaussian tails on both sides of the signal peak due to detector resolutions depending on angular distributions in the decay. We model the signal shape by an empirical model composed of two Crystal Ball

---

\(^1\) $B$ stands for $B^0$ or $B^0_c$. 

---

Fig. 1. Decay topologies contributing to the $B^0 \rightarrow J/\psi K^0_S$ and $B^0_c \rightarrow J/\psi K^0_S$ channel: tree diagram to the left and penguin diagram to the right.

Fig. 2. Mass distribution of the $B \rightarrow J/\psi K^0_S$ candidates used to determine the PDF. The solid line is the total PDF composed of the $B^0 \rightarrow J/\psi K^0_S$ signal shown in grey and the combinatorial background represented by the dotted line.
(CB) functions [19], one of which has the tail extending to high masses. The two CB components are constrained to have the same peak and width, which are allowed to vary in the fit. The parameters describing the CB tails are taken from $B^+ \rightarrow J/\psi K^+$ events which exhibit the same behaviour as $B \rightarrow J/\psi K^0_S$. The combinatorial background is described by a second order polynomial. The $B^0 \rightarrow J/\psi K^0_S$ signal is not included in this fit. We extract $(14.4 \pm 0.2) \times 10^2$ $B^0$ events from the fit.

The NN uses information about the candidate kinematics, vertex and track quality, impact parameter, particle identification information from the RICH and muon detectors, as well as global event properties like track and primary vertex multiplicities. The variables that are used in the NN are chosen not to induce a correlation with the mass distribution. This was verified using simulated events.

To maximise the separation power, a first NN classifier using only the five most discriminating variables is used to remove 80% of the background events while keeping 95% of the $B^0$ signal. These variables are the $\chi^2$ of the decay chain fit, the angle between the $B$ momentum and the vector from the primary vertex to the decay vertex, the $p_T$ of the $K^0_S$, the estimated uncertainty on the $B$ mass and the impact parameter $\chi^2$ of the $J/\psi$.

The weighting procedure is then repeated on the remaining candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained. A cut is then made on the second NN output in order to optimise the expected sensitivity to the $B^0$ and $B^0$ candidates and a second NN classifier containing 31 variables is trained.

The trigger, reconstruction and selection efficiencies also depend on the transverse momentum of the final state particles. Applying the trigger transverse momentum cuts on simulated $B^0$ and $B^0$ decays we find differences of up to 1%, which is taken as systematic uncertainty.

Due to the selection cuts and the correlation of the neural network with the decay time, a decay time acceptance function results in different selection efficiencies for the $B^0$ and $B^0$. We determine the lifetime acceptance of the whole selection chain using simulated events, and find that the ratio of the time-integrated decay time distributions for $B^0$ and $B^0$ is $0.975 \pm 0.007$. The uncertainties on the parametrisation of the lifetime acceptance cancel almost perfectly in the ratio, while the ones related to the $B^0$ and $B^0$ lifetimes and the $B^0$ decay width difference $\Delta \Gamma$ do not.

The largest systematic uncertainty comes from the assumed mass PDF, in particular the fraction of the positive tail of the $B^0$ extending below the $B^0$ signal. We have studied the magnitude of this effect by leaving both tails of the CB shapes free in the fit, or by allowing the two CB shapes to have different widths. The maximal deviation we observe in the ratios of downstream or long candidates is 5%, which we take as systematic uncertainty. The effect of the uncertainty on the $B^0 - B^0$ mass difference is found to be 0.4%.

The corrections and systematic uncertainties affecting the branching fraction ratio are listed in Table 2. The total uncertainty is obtained by adding all the uncertainties in quadrature.

---

**Table 1**

<table>
<thead>
<tr>
<th>Yields</th>
<th>Downstream $K^0_S$</th>
<th>Long $K^0_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^0$ in normalisation sample</td>
<td>1502 ± 39</td>
<td>970 ± 31</td>
</tr>
<tr>
<td>$B^0$ in normalisation sample (scaled to full)</td>
<td>6007 ± 157</td>
<td>3879 ± 124</td>
</tr>
<tr>
<td>$B^0$ in full sample</td>
<td>72 ± 11</td>
<td>44 ± 8</td>
</tr>
<tr>
<td>Ratio of $B^0$ to $B^0$</td>
<td>0.0120 ± 0.0018</td>
<td>0.0112 ± 0.0020</td>
</tr>
<tr>
<td>Ratio of $B^0$ to $B^0$ (weighted average, $r$)</td>
<td>0.0117 ± 0.0014</td>
<td></td>
</tr>
</tbody>
</table>

---

**Fig. 3.** Fit to full sample after the optimal NN cut has been applied with downstream $K^0_S$ to the left and long $K^0_S$ to the right.
Table 2
Summary of corrections and systematic uncertainties on the ratio of branching fractions.

<table>
<thead>
<tr>
<th>Source</th>
<th>Correction factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometrical acceptance ($\epsilon_{\text{geom}}$)</td>
<td>0.987 ± 0.018</td>
</tr>
<tr>
<td>Trigger and reconstruction</td>
<td>1.000 ± 0.010</td>
</tr>
<tr>
<td>Decay time acceptance ($\epsilon_{\text{time}}$)</td>
<td>1.975 ± 0.007</td>
</tr>
<tr>
<td>Mass shape</td>
<td>1.000 ± 0.050</td>
</tr>
<tr>
<td>$B^0_B$ - $B^0_S$ mass difference</td>
<td>1.000 ± 0.004</td>
</tr>
<tr>
<td>Total</td>
<td>0.962 ± 0.053</td>
</tr>
</tbody>
</table>

We verify that the global event variable distributions, like the number of primary vertices and the hit multiplicities, are the same for $B^0_B$ and $B^0_S$ initial states using the $B^0_S \to J/\psi K^0_S$ and $B^0 \to J/\psi K^0_S$ yields, the geometrical ($\epsilon_{\text{geom}}$) and lifetime ($\epsilon_{\text{time}}$) acceptance ratios, and assuming $f_s/f_d = 0.267^{+0.021}_{-0.020}$ [8] we measure the ratio of branching fractions

$$
\frac{B(B^0_B \to J/\psi K^0_S)}{B(B^0 \to J/\psi K^0_S)} = r \times \epsilon_{\text{geom}} \times \epsilon_{\text{time}} \times \frac{f_s}{f_d} 
$$

$$
= 0.0420 \pm 0.0049(\text{stat}) \pm 0.0023(\text{syst}) \pm 0.0033(f_s/f_d) \quad (1)
$$

where the quoted uncertainties are statistical, systematic, and due to the uncertainty in $f_s/f_d$, respectively. Using the $B^0 \to J/\psi K^0_S$ branching fraction of $(8.71 \pm 0.32) \times 10^{-4}$ [22], we determine the time-integrated $B^0\to J/\psi K^0_S$ branching fraction

$$
B(B^0 \to J/\psi K^0_S) = \left[1.83 \pm 0.21(\text{stat}) \pm 0.10(\text{syst}) \right] \times 0.14(f_s/f_d) \pm 0.07(B(B^0 \to J/\psi K^0_S)) \times 10^{-5}
$$

where the last uncertainty comes from the $B^0_B \to J/\psi K^0_S$ branching fraction. This result is compatible with, and more precise than, the previous measurement [7].

6. Comparison with SU(3) expectations

It was pointed out in Ref. [23] that because of the sizable decay width difference between the heavy and light eigenstates of the $B^0_S$ system, there is an ambiguity in the definition of the branching fractions of $B^0_S$ decays. Due to $B^0_S$ mixing, a branching fraction defined as the ratio of the time integrated number of $B^0_S$ decays to a final state and the total number of $B^0_S$ mesons, is not equal to the CP-average of the decay rates in the flavour eigenstate basis

$$
B(B^0_S \to f)_{\text{theo}} = \frac{\tau_{B^0_S}}{2} \left[ \Gamma(B^0_B \to f) + \Gamma(B^0_S \to f) \right]_{l=0} 
$$

used in the theoretical predictions; the restriction to $l = 0$ removes the effects due to the non-zero $B_s$ decay width. To obtain the latter quantity from the time-integrated decay rates the following correction factor

$$
1 - \frac{y_s^2}{1 + A_{\Delta f} \, y_s} = 0.936 \pm 0.015, \quad (3)
$$

is applied, where $y_s = \Delta f_s/2f_s$ is the normalised decay width difference between the light and heavy states and $A_{\Delta f} = 3$ is the final-state dependent asymmetry of the $B^0_S$ decay rates to the $J/\psi K^0_S$ final state. In calculating this correction factor we use $y_s = 0.075 \pm 0.010$ [24] and the SM expectation $A_{\Delta f} = 0.84 \pm 0.18$ [23].

With this correction, and assuming $B(B^0_B \to J/\psi K^0_S)_{\text{theo}} = \frac{1}{2} B(B^0 \to J/\psi K^0_S)_{\text{theo}}$ we get the $B^0 \to J/\psi K^0_S$ branching fraction at $t = 0$ for $B^0 \to J/\psi K^0_S$

$$
B(B^0_B \to J/\psi K^0_S)_{\text{theo}} = (3.42 \pm 0.40(\text{stat}) \pm 0.19(\text{syst}) \pm 0.27(f_s/f_d) \pm 0.13(B(B^0 \to J/\psi K^0_S))) \times 0.05(y_s, A_{\Delta f}) \cdot 10^{-5}.
$$

This branching fraction can be compared to theoretical expectations from SU(3) symmetry, which implies an equality of the $B^0_B \to J/\psi K^0_S$ and $B^0 \to J/\psi \pi^0$ decay widths [6]

$$
\Xi_{\text{SU(3)}} = \frac{B(B^0_B \to J/\psi K^0_S)_{\text{theo}} \, \tau_{B^0_S} \, m_{B^0} \, \Phi(B^0 \to J/\psi \pi^0)^3 \, 2B(B^0 \to J/\psi \pi^0) \, m_{B^0} \, \Phi(B^0 \to J/\psi K^0_S)^3}{2 B(B^0 \to J/\psi K^0_S)} \rightarrow 1, \quad (4)
$$

where the factor two is associated with the wave function of the $\pi^0$, $\tau_{B^0_S}$ is the mean $B^0_S$ lifetime and $\Phi$ refers to the two-body phase-space factors; see e.g. Ref. [5].

Taking the measured $B(B^0_B \to J/\psi K^0_S)_{\text{theo}}$ and using the world average [22,21] for all other quantities, this ratio becomes

$$
\Xi_{\text{SU(3)}} = 0.98 \pm 0.18
$$

and is consistent with theoretical expectation of unity under SU(3) symmetry.

7. Conclusion

The branching fraction of the Cabibbo-suppressed decay $B^0 \to J/\psi K^0_S$ is measured in a 0.41 fb$^{-1}$ data sample collected with the LHCb detector. We determine the ratio of the $B^0_B \to J/\psi K^0_S$ and $B^0 \to J/\psi K^0_S$ branching fractions to be $\frac{B(B^0_B \to J/\psi K^0_S)}{B(B^0 \to J/\psi K^0_S)} = 0.0420 \pm 0.0049(\text{stat}) \pm 0.0023(\text{syst}) \pm 0.0033(f_s/f_d)$. Using the world-average $B^0 \to J/\psi K^0_S$ branching fraction we get the time-integrated branching fraction $\frac{B(B^0_B \to J/\psi K^0_S)}{B(B^0 \to J/\psi K^0_S)} = [1.83 \pm 0.21(\text{stat}) \pm 0.10(\text{syst}) \pm 0.14(f_s/f_d) \pm 0.07(B(B^0 \to J/\psi K^0_S))] \times 10^{-5}$. The total uncertainty of 16% is dominated by the statistical uncertainty. This branching fraction is compatible with expectations from SU(3).

With larger data samples, a time dependent CP-violation measurement of this decay will be possible, allowing the experimental determination of the penguin contributions to the sin2$\beta$ measurement from $B^0 \to J/\psi K^0_S$.

Acknowledgements

We express our gratitude to our colleagues in the CERN accelerator departments for the excellent performance of the LHC. We thank the technical and administrative staff at CERN and at the
LHCb institutes, and acknowledge support from the National Agencies: CAPES, CNPq, FAPERJ and FINER (Brazil); CERN; NSFC (China); CNRS/IN2P3 (France); BMBF, DFG, HGF and MPG (Germany); SFI (Ireland); INFN (Italy) and NWO (The Netherlands); SCSR (Poland); ANCS (Romania); MinES of Russia and Rosatom (Russia); MICINN, XuntaGal and GENCAT (Spain); SNSF and SER (Switzerland); NAS Ukraine (Ukraine); STFC (United Kingdom); NSF (USA). We also acknowledge the support received from the ERC under FP7 and the Region Auvergne.

Open access
This article is published Open Access at sciencedirect.com. It is distributed under the terms of the Creative Commons Attribution License 3.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

References

LHCb Collaboration