Evidence for the decay $B_s^0 \rightarrow \overline{K}^{*0} \mu^+ \mu^-$

The LHCb collaboration

E-mail: andrew.christopher.crocombe@cern.ch

Abstract: A search for the decay $B_s^0 \rightarrow \overline{K}^{*0} \mu^+ \mu^-$ is presented using data sets corresponding to 1.0, 2.0 and 1.6 fb$^{-1}$ of integrated luminosity collected during $pp$ collisions with the LHCb experiment at centre-of-mass energies of 7, 8 and 13 TeV, respectively. An excess is found over the background-only hypothesis with a significance of 3.4 standard deviations. The branching fraction of the $B_s^0 \rightarrow \overline{K}^{*0} \mu^+ \mu^-$ decay is determined to be $\mathcal{B}(B_s^0 \rightarrow \overline{K}^{*0} \mu^+ \mu^-) = [2.9 \pm 1.0 \text{ (stat)} \pm 0.2 \text{ (syst)} \pm 0.3 \text{ (norm)}] \times 10^{-8}$, where the first and second uncertainties are statistical and systematic, respectively. The third uncertainty is due to limited knowledge of external parameters used to normalise the branching fraction measurement.

Keywords: B physics, FCNC Interaction, Flavor physics, Hadron-Hadron scattering (experiments), Rare decay

ArXiv ePrint: 1804.07167
1 Introduction

The decay $B^0_s \rightarrow K^\ast(892)^0 \mu^+\mu^-$, hereafter referred to as $B^0_s \rightarrow K^\ast^0 \mu^+\mu^-$, proceeds via a $b \rightarrow d$ flavour-changing neutral-current (FCNC) transition. The leading contributions to the amplitude of the decay correspond to loop Feynman diagrams and involve the off-diagonal element $V_{td}$ of the Cabibbo-Kobayashi-Maskawa (CKM) quark-mixing matrix. This process is consequently rare in the Standard Model of particle physics (SM). New particles predicted by extensions of the SM can enter in competing diagrams and can significantly enhance or suppress the rate of the decay, see for example refs. [1, 2]. Form-factor computations for the $B^0_s \rightarrow K^\ast^0$ transition have been made using light-cone sum rule [3, 4] and lattice QCD [5] techniques. Standard Model predictions for the branching fraction of the decay are in the range $3\text{–}4 \times 10^{-8}$ [6–8].

The observation of the rare $b \rightarrow d\ell^+\ell^-$ FCNC decays $B^+ \rightarrow \pi^+\mu^+\mu^-$ and $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ has been previously reported by the LHCb collaboration in refs. [9] and [10], respectively. Evidence for the decay $B^0 \rightarrow \pi^+\pi^-\mu^+\mu^-$ has also been established in ref. [11]. The decay $B^0_s \rightarrow \overline{K}^{*0}\mu^+\mu^-$ has not yet been observed. The measured ratio of the $B^+ \rightarrow \pi^+\mu^+\mu^-$ and $B^+ \rightarrow K^+\mu^+\mu^-$ branching fractions has also been used to determine the ratio of CKM elements $|V_{td}/V_{ts}|$ [12], exploiting correlations between the $B \rightarrow K$ and $B \rightarrow \pi$ form-factors in lattice computations. A similar approach could, in the future, be applied to the ratio of the $B^0_s \rightarrow \overline{K}^{*0}\mu^+\mu^-$ and $B^0 \rightarrow \overline{K}^{*0}\mu^+\mu^-$ decay rates [13].

The decay $\overline{B}^0 \rightarrow \overline{K}^{*0}\mu^+\mu^-$, which involves a $b \rightarrow s\ell^+\ell^-$ transition, has been studied extensively by BaBar, Belle, CDF and by the LHC experiments [14–19]. The rate of the
decay appears to be systematically lower than current SM predictions. Global analyses of $b \to s$ processes favour a modification of the SM at the level of 4 to 5 standard deviations [20–24]. Similar studies of $b \to d$ processes are important to understand the flavour structure of the underlying theory.

This paper presents a search for the decay $B^0_s \to \bar{K}^{*0}\mu^+\mu^-$, where the inclusion of charge-conjugate processes is implied throughout, using data collected with the LHCb experiment in $pp$ collisions during Runs 1 and 2 of the LHC. The data set used in this paper is as follows: 1.0 fb$^{-1}$ of integrated luminosity collected at a centre-of-mass energy of 7 TeV during Run 1; 2.0 fb$^{-1}$ of integrated luminosity collected at a centre-of-mass energy of 8 TeV during Run 1; and 1.6 fb$^{-1}$ of integrated luminosity collected at a centre-of-mass energy of 13 TeV during Run 2. Section 2 of this paper describes the LHCb detector and the experimental setup used for the analysis. Section 3 outlines the selection processes used to identify signal candidates. Section 4 describes the method used to estimate the number of $B^0_s \to \bar{K}^{*0}\mu^+\mu^-$ decays in the data set. Section 5 describes the determination of the $B^0 \to \bar{K}^{*0}\mu^+\mu^-$ branching fraction, normalising the number of observed signal decays to the number of $B^0 \to J/\psi K^{*0}$ decays present in the data set. Section 6 discusses sources of systematic uncertainty on the $B^0_s \to \bar{K}^{*0}\mu^+\mu^-$ branching fraction. Finally, conclusions are presented in section 7.

2 Detector and simulation

The LHCb detector [25, 26] is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$, designed for the study of particles containing $b$ or $c$ quarks. The detector includes a high-precision tracking system consisting of a silicon-strip vertex detector surrounding the $pp$ interaction region [27], 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 [28] placed downstream of the magnet. The tracking system provides a measurement of momentum, $p$, of charged particles with a relative uncertainty that varies from 0.5% at low momentum to 1.0% at 200 GeV/$c$. The minimum distance of a track to a primary vertex (PV), the impact parameter (IP), is measured with a resolution of $(15 + 29/p_T)$ µm, where $p_T$ is the component of the momentum transverse to the beam, in GeV/$c$. Different types of charged hadrons are distinguished using information from two ring-imaging Cherenkov detectors [29]. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic calorimeter and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers [30].

The online event selection is performed by a trigger [31]. The trigger consists of a hardware stage, based on information from the calorimeter and muon systems, followed by a software stage, which applies a full event reconstruction. The signal candidates are required to pass through a hardware trigger that selects events containing at least one muon with $p_T$ greater than 1 to 2 GeV/$c$, depending on the data-taking conditions. The software trigger requires a two-, three- or four-track secondary vertex with a significant
displacement from any primary \( pp \) interaction vertex. At least one charged particle must have a large transverse momentum \( p_T > 1 \text{ GeV}/c \) and be inconsistent with originating from a PV. A multivariate algorithm [32] is used for the identification of secondary vertices consistent with the decay of a \( b \) hadron.

Samples of simulated \( B^0_s \rightarrow K^{*0} \mu^+\mu^- \), \( B^0 \rightarrow K^{*0} \mu^+\mu^- \), \( B^0_s \rightarrow J/\psi \bar{K}^{*0} \) and \( B^0 \rightarrow J/\psi \bar{K}^{*0} \) decays are used to develop an offline event selection and to determine the efficiency to reconstruct the \( B^0 \) and \( B^0_s \) candidates in the different data-taking periods. In the simulation, \( pp \) collisions are generated using Pythia [33] with a specific LHCb configuration [34]. Decays of hadronic particles are described by EvtGen [35], in which final-state radiation is generated using PHOTOS [36]. The interaction of the generated particles with the detector, and its response, are implemented using the Geant4 toolkit [37, 38] as described in ref. [39]. Data-driven corrections are applied to the simulation to account for mismodelling of the detector occupancy and of the \( B^0_s \) meson production kinematics. The particle identification (PID) performance is measured from data using calibration samples [26].

3 Candidate selection

Signal candidates are formed by combining a \( K^{*0} \) candidate with two oppositely charged tracks, which are identified as muons by the muon system. The \( K^{*0} \) meson is reconstructed through its decay to the \( K^-\pi^+ \) final state with invariant mass within \( \pm 70 \text{ MeV}/c^2 \) of the known \( K^*(892)^0 \) mass [40]. The muon pair is required to have an invariant mass squared in the range \( 0.1 < q^2 < 19.0 \text{ GeV}^2/c^4 \), excluding the region \( 12.5 < q^2 < 15.0 \text{ GeV}^2/c^4 \) dominated by the \( \psi(2S) \) resonance. Candidates in the region \( 8.0 < q^2 < 11.0 \text{ GeV}^2/c^4 \), which are dominated by decays via a \( J/\psi \) resonance, are treated separately in the analysis. The remaining candidates include \( B^0_s \) meson decays that produce a dimuon pair through the decay of a light-quark resonance or a charmonium state above the open charm threshold, which are inseparable from the short-distance component of the decay. These are considered part of the signal in the analysis.

The selection process used in this analysis is similar to that described in ref. [18]. The four charged tracks are required to each have a significant IP with respect to all PVs in the event and to be consistent with originating from a common vertex. The \( B^0_s \) meson candidate is required to be consistent with originating from one of the PVs in the event and its decay vertex is required to be well separated from that PV. The kaon and pion candidates must also be identified as kaon-like and pion-like by a multivariate algorithm [26] based on information from the RICH detectors, tracking system and calorimeters. The PID requirements are chosen to maximise the sensitivity to a SM-like \( B^0_s \rightarrow K^{*0} \mu^+\mu^- \) signal.

To improve the resolution on the reconstructed \( K^-\pi^+\mu^+\mu^- \) invariant mass, \( m(K^-\pi^+\mu^+\mu^-) \), candidates with an uncertainty larger than \( 22 \text{ MeV}/c^2 \) on their measured mass are rejected. The opening angle between every pair of final-state particles is also required to be larger than \( 5 \text{ mrad} \) in the detector. This requirement removes a possible source of background that arises when the hits associated to a given charged particle are mistakenly used in more than one reconstructed track. A kinematic fit is also performed,
constraining the candidate to originate from its most likely production vertex \cite{41}. In the kinematic fit of candidates with $q^2$ in the $J/\psi$ mass window, the dimuon pair is also constrained to the known $J/\psi$ mass. This mass constraint improves the resolution in $m(K^-\pi^+\mu^+\mu^-)$ for candidates involving an intermediate $J/\psi$ resonance decay by a factor of two.

Signal candidates are further classified using an artificial neural network \cite{42}. The neural network is trained using a sample of simulated $B^0 \to K^{*0}\mu^+\mu^-$ decays as a proxy for the signal decay. Candidates in data with $m(K^-\pi^+\mu^+\mu^-) > 5670\,\text{MeV}/c^2$ are used as a background sample. This sample is predominantly comprised of combinatorial background, where uncorrelated tracks from the event are mistakenly combined. The neural network uses the following variables related to the topology of the $B^0_{(s)}$ meson decay: the angle between the reconstructed momentum vector of the $B^0_{(s)}$ meson and the vector connecting the PV and the decay vertex of the $B^0_{(s)}$ candidate; the IP, $p_T$, and proper decay time of the $B^0_{(s)}$ candidate; the vertex fit quality of the $B^0_{(s)}$ decay vertex and of the dimuon pair; the minimum and maximum $p_T$ of the final-state particles and for the Run 1 data set a measure of the isolation of the final-state particles in the detector. It has been verified that the distribution of the variables used as input to, and the output distribution from, the classifier agree between the simulation and the data. The output of the neural network is transformed such that it is uniform in the range 0–1 on the signal proxy. Candidates with neural network response below 0.05 are rejected in the subsequent analysis. This requirement removes a background-dominated part of the data sample. The neural network response is validated on simulated $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0_s \to K^{*0}\mu^+\mu^-$ decays to ensure that it does not introduce any bias in $m(K^-\pi^+\mu^+\mu^-)$.

Finally, a number of vetoes are applied to reject specific sources of background. Signal candidates are rejected if the pion candidate has a nonnegligible probability to be a kaon and if the $K^+K^-$ invariant mass, after assigning the kaon mass to the pion candidate, is consistent within $10\,\text{MeV}/c^2$ of the known $\phi(1020)$ meson mass. This veto removes 98% of $B^0_{s} \to \phi\mu^+\mu^-$ decays inside the $\phi(1020)$ mass window. Candidates are also rejected if the kaon or pion are identifiable as a muon and the $K^-\mu^+$ or $\pi^+\mu^-$ mass, after assigning the muon mass hypothesis to the kaon or pion candidate, are consistent with that of a $J/\psi$ or $\psi(2S)$ meson (within $\pm60\,\text{MeV}/c^2$ of their known masses).

4 Signal yields

In order to maximise sensitivity to a $B^0_{s} \to \overline{K}^{*0}\mu^+\mu^-$ signal, candidates are divided into regions of neural network response. The candidates are also divided based on the two data-taking periods, Run 1 and Run 2. Four regions of neural network response are selected for each data-taking period, each containing an equal amount of expected signal decays. The yield of the $B^0_{s} \to \overline{K}^{*0}\mu^+\mu^-$ decay is determined by performing a simultaneous unbinned maximum likelihood fit to the $m(K^-\pi^+\mu^+\mu^-)$ distribution of the eight resulting subsets of the data.

In the likelihood fit, the signal lineshape of both the $B^0$ and the $B^0_{s} \to \overline{K}^{*0}\mu^+\mu^-$ decays is described by the sum of three functions: a Gaussian function with a power-law tail on the
lower-side of its peak, used to describe final-state radiation and energy loss in the detector; a Gaussian function with a power-law tail on the upper-side of its peak, used to describe the non-Gaussian tails of the signal mass distribution at large masses; and an additional Gaussian function to account for differences in the per-candidate resolution of the reconstructed mass. The two functions with power-law tails share a common width and all three functions share a common peak position. The $B_0^s$ peak position is displaced from that of the $B^0$ by 87.5 MeV/$c^2$ [43]. The relative fractions of each function are fixed from fits to simulated $B^0$ and $B_0^s \to \bar{K}^0\mu^+\mu^-$ decays. The widths of the functions and all of the tail parameters are also fixed from the simulation, except for an overall scaling of the widths and of the tail parameters to allow for potential data-simulation differences. The peak position and these scale factors are obtained from a fit to candidates with the dimuon in the $J/\psi$ mass window, where the mass constraint on the dimuon mass has not been applied. The result of this fit is shown in the appendix in figure 4. In the fit to the data, the widths vary from their values in the simulation by 10 to 15%. The turn-on point of the upper tail (relative to the width of the distribution) is found to be consistent between data and simulation.

After applying the selection procedure, the background predominately comprises combinatorial background. The combinatorial background is described in the fit by a separate exponential function in each subset of the data. A number of other sources of background are accounted for in the fit. The decay $B^0 \to K^{*0}\mu^+\mu^-$ forms a source of background if the kaon is mistakenly identified as the pion and vice versa. The shape of this background is taken from the simulation. The yield of the background is constrained relative to that of the $B^0 \to \bar{K}^0\mu^+\mu^-$ decay based on measurements of the kaon-to-pion and pion-to-kaon misidentification probabilities in the PID calibration samples. The decay $A_0^0 \to p\bar{K}^{-}\mu^+\mu^-$ forms a source of background if the final-state hadrons are misidentified. This background is constrained from a control region in the data, by modifying the PID requirements on the candidates to preferentially select $p\bar{K}^-$ rather than $K^-\pi^+$ combinations. The shape of this background is modelled in the fit by Crystal Ball functions. The yield in each subset of the data is constrained using the proton and kaon identification and misidentification probabilities determined from the PID calibration samples. The decay $B^- \to K^-\mu^+\mu^-$ forms a source of background if a pion from the event is mistakenly combined with the particles coming from the $B^-$ meson decay. The background contribution from $B^- \to K^-\mu^+\mu^-$ decays is determined from a control region in the data, by selecting candidates with a $K^-\mu^+\mu^-$ invariant mass that is consistent with the known $B^-$ mass. This background is only visible for the candidates with $q^2$ in the $J/\psi$ mass region. The shape of the background in the fit is modelled by Crystal Ball functions. Several other sources of background are considered but are found to have a negligible contribution to the fit. These sources include semileptonic decays of $b$ hadrons via intermediate open-charm states and fully hadronic $b$-hadron decays. The background from semileptonic decays is predominantly reconstructed at low $m(K^-\pi^+\mu^+\mu^-)$ and does not contribute to the analysis. Fully hadronic $b$-hadron decays contribute at the level of 1 to 2 candidates at masses close to the known $B^0_s$ mass. This background is neglected in the analysis but is considered as a source of systematic uncertainty in section 6.

Figure 1 shows the fit to the candidates, where the result of the fit in the three most signal-like neural network response bins for each data-taking period has been combined.
Figure 1. Distribution of reconstructed $K^-\pi^+\mu^+\mu^-$ invariant mass of candidates outside the $J/\psi$ and $\psi(2S)$ mass regions, summing the three highest neural network response bins of each run condition. The candidates are shown (left) over the full range and (right) over a restricted vertical range to emphasise the $B_s^0 \rightarrow K^{*0}\mu^+\mu^-$ component. The solid line indicates a combination of the results of the fits to the individual bins. Components are detailed in the legend, where they are shown in the same order as they are stacked in the figure. The background from misidentified $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decays is included in the $\bar{B}^0 \rightarrow K^{*0}\mu^+\mu^-$ component.

Candidates in the least signal-like bin are not included. This bin has a much higher level of combinatorial background and would visually obscure any $B_s^0$ signal. The dominant contribution in the fit is the $\bar{B}^0 \rightarrow K^{*0}\mu^+\mu^-$ decay. Figure 2 shows the fit to the mass-constrained candidates in the $J/\psi$ mass region, also with the three highest neural network response bins for each data taking period combined. In this fit, a small background component from $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decays is included. This background has the same final state but is constrained to the wrong dimuon mass and becomes a broad component in the fit. The fit results in individual bins of neural network response are shown in the appendix in figures 5 and 6. Summing over the bins of neural network response and data-taking periods, the yields are: 627244 ± 837 for the $B^0 \rightarrow J/\psi K^{*0}$ decay, 5730 ± 94 for the $B_s^0 \rightarrow J/\psi K^{*0}$ decay, 4157 ± 72 for the $\bar{B}^0 \rightarrow K^{*0}\mu^+\mu^-$ decay, and 38 ± 12 for the $B_s^0 \rightarrow K^{*0}\mu^+\mu^-$ decay. No correction has been made to these yields to account for cases where the $K^-\pi^+$ system does not originate from a $K^*(892)^0$ decay. Contamination from non-$K^{*0}$ decays is discussed further in section 5. Using Wilks’ theorem, and a likelihood ratio test between the signal-plus-background and the background-only hypothesis, the significance of the $B_s^0 \rightarrow K^{*0}\mu^+\mu^-$ yield is determined to be $\sqrt{-2 \log(L_{S+B}/L_B)} = 3.4$ standard deviations. The signal significance has been validated using pseudoexperiments generated under the null hypothesis. This includes the systematic uncertainties on the yield discussed in section 6. Figure 3 shows the variation of the log-likelihood of the simultaneous fit as a function of the $B_s^0 \rightarrow K^{*0}\mu^+\mu^-$ yield.
Figure 2. Distribution of reconstructed $J/\psi K^{-}\pi^{+}$ invariant mass of the candidates in the $J/\psi$ mass region summing the three highest neural network response bins of each run condition, shown (left) over the full range and (right) over a restricted vertical range to emphasise the $B^{0}_{s}\rightarrow J/\psi K^{*0}$ component. The solid line indicates a combination of the results of the fits to the individual bins. Components are detailed in the legend, where they are shown in the same order as they are stacked in the figure. The background from misidentified $B^{0}\rightarrow J/\psi K^{*0}$ decays is included in the $B^{0}\rightarrow J/\psi K^{*0}$ component.

Figure 3. Change in log-likelihood from the simultaneous fit to the candidates in the two data-taking periods and the different bins of neural network response, as a function of the $B^{0}_{s}\rightarrow K^{*0}\mu^{+}\mu^{-}$ yield. Systematic uncertainties on the yield have been included in the likelihood.
5 Results

The branching fraction of the $B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-$ decay is determined with respect to that of $B^0 \to J/\psi \bar{K}^{*0}$ according to

$$
\mathcal{B}(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-) = \mathcal{B}(B^0 \to J/\psi \bar{K}^{*0}) \mathcal{B}(J/\psi \to \mu^+ \mu^-) \times \frac{f_d N(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-)}{f_s \varepsilon(B^0 \to J/\psi \bar{K}^{*0}) \varepsilon(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-) N(B^0 \to J/\psi \bar{K}^{*0})}. \tag{5.1}
$$

Here, $N$ is the yield for a given decay mode determined from the fit to $m(K^-\pi^+\mu^+\mu^-)$ or $m(J/\psi K^-\pi^+)$ and $\varepsilon$ is the efficiency to reconstruct and select the given decay mode. The ratio $f_s/f_d$ is the relative production fraction of $B_s^0$ and $B^0$ mesons in $pp$ collisions.

The efficiency to trigger, reconstruct and select each of the decay modes is determined from the simulation after applying the data-driven corrections. The efficiency for the $B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-$ decay is corrected to account for events in the vetoed $q^2$ regions following the same prescription as ref. [19]. The efficiency corrected yields are further corrected for contamination from decays with the $K^-\pi^+$ system in an S-wave configuration. For the decay $B_s^0 \to J/\psi \bar{K}^{*0}$, the S-wave fraction of $F_S(B^0 \to J/\psi \bar{K}^{*0}) = (6.4 \pm 0.3 \pm 1.0)\%$ determined in ref. [44] is used. The S-wave contamination of the $B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-$ decay is unknown but it is assumed to be at a similar level to that of the $B^0 \to \bar{K}^{*0} \mu^+ \mu^-$ decay. The full size of the S-wave correction is taken as a systematic uncertainty. The S-wave contamination of the $B^0 \to \bar{K}^{*0} \mu^+ \mu^-$ decay is determined using the model from ref. [19]. This model predicts an S-wave fraction of $F_S(B^0 \to \bar{K}^{*0} \mu^+ \mu^-) = (3.4 \pm 0.8)\%$ in the $K^-\pi^+$ mass window used in this analysis.

The ratio of production fractions, $f_s/f_d$, has been measured at 7 and 8 TeV to be $f_s/f_d = 0.259 \pm 0.015$ in the LHCb detector acceptance [45]. The production fraction at 13 TeV has been shown to be consistent with that of the 7 and 8 TeV data in ref. [47]. The production fraction at 13 TeV has also been validated in this analysis by comparing the efficiency-corrected yields of the $B^0$ and the $B_s^0 \to J/\psi \bar{K}^{*0}$ decays in bins of the $B_s^0$ meson $p_T$. Taking the branching fractions of the decays $B^0 \to J/\psi \bar{K}^{*0}$ and $J/\psi \to \mu^+ \mu^-$ to be $(1.19 \pm 0.01 \pm 0.08) \times 10^{-3}$ [48] and $(5.96 \pm 0.03)\%$ [40], respectively, results in a branching fraction for the $B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-$ decay of

$$
\mathcal{B}(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-) = [2.9 \pm 1.0 \text{ (stat)} \pm 0.2 \text{ (syst)} \pm 0.3 \text{ (norm)}] \times 10^{-8}.
$$

The first and second uncertainties are statistical and systematic, respectively. The third uncertainty is due to limited knowledge of the external parameters used to normalise the observed yield. This comprises the uncertainty on the external branching fraction measurements, on $f_s/f_d$, $F_S(B^0 \to J/\psi \bar{K}^{*0})$ and $F_S(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-)$.

A measurement of the branching fraction of the $B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-$ decay relative to that of $B_s^0 \to J/\psi \bar{K}^{*0}$ is also made. The S-wave contamination of the $B_s^0 \to J/\psi \bar{K}^{*0}$ decay is corrected for by using the measurements of $F_S$ in bins of $m(K^-\pi^+)$ from ref. [49], scaled
where the third uncertainty is due to \( F_s(B_s^0 \to J/\psi K^{*0}) \) and \( F_s(B_s^0 \to K^{*0} \mu^+ \mu^-) \).

In order to determine the ratio \( |V_{td}/V_{ts}| \) it is also useful to extract the ratio

\[
\frac{B(B_s^0 \to K^{*0} \mu^+ \mu^-)}{B(B_d^0 \to K^{*0} \mu^+ \mu^-)} = \frac{f_d \left[ N(B_s^0 \to K^{*0} \mu^+ \mu^-) \right]}{f_s \left[ N(B_d^0 \to K^{*0} \mu^+ \mu^-) \right]} \times 10^{-2},
\]

where the third uncertainty corresponds to the uncertainties on \( f_s/f_d, F_s(B_d^0 \to K^{*0} \mu^+ \mu^-) \) and \( F_s(B_s^0 \to K^{*0} \mu^+ \mu^-) \).

## 6 Systematic uncertainties

The measurements presented in section 5 are performed relative to decays that have the same final-state particles as the \( B_s^0 \to K^{*0} \mu^+ \mu^- \) decay. Consequently, many potential sources of systematic uncertainty largely cancel in the ratios. The remaining sources of systematic uncertainty are discussed below and are summarised in table 1. Only systematic uncertainties that have an effect on the measured yield are considered when evaluating the significance of the observed signal. These are systematic uncertainties related to the signal resolution, neural network binning scheme and the residual backgrounds at \( m(K^- \pi^+ \pi^-) \) close to the known \( B_s^0 \) meson mass.

The \( m(K^- \pi^+ \pi^-) \) model used to describe the decays \( B_s^0 \) and \( B_s^0 \to K^{*0} \mu^+ \mu^- \) is taken from the simulation with a simple scaling of the width and tail parameters based on the fit to the data in the \( J/\psi \) mass region. Any difference in the \( q^2 \) spectrum of the simulation and the data could result in a small mismodelling of the lineshape. To account for this possibility, the width of the \( m(K^- \pi^+ \pi^-) \) resolution model is allowed to vary within 0.5 MeV/c^2 in the fit. This covers the full variation in the simulation of the width across the allowed \( q^2 \) range and contributes 0.1% to the systematic uncertainty. A final uncertainty on the signal lineshape is evaluated based on the difference in fits to the candidates in the \( J/\psi \) mass region with and without the constraint on the dimuon mass. A systematic uncertainty of 0.5% is assigned, taken as the difference in efficiency-corrected \( B^0 \to J/\psi K^{*0} \) yields between these two fits. In addition, an alternative parameterisation with an exponential tail rather than a power-law tail is tested for the lineshape describing the \( A^0_s \) background. The difference in yields between the two models results in a systematic uncertainty of 0.1% on the \( B_s^0 \to K^{*0} \mu^+ \mu^- \) yield. The total uncertainty related to mass lineshapes is taken as the sum in quadrature of the uncertainties.

The systematic uncertainty related to the relative efficiencies in each neural network response bin is evaluated in two parts: an uncertainty due to the limited size of the simulation sample used to determine the relative fractions and an uncertainty due to differences
between simulated samples and the data. The latter is evaluated by correcting the fraction of $B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-$ decays in each neural network response bin by the measured difference between simulation and data for the $B^0 \to J/\psi \bar{K}^{*0}$ decays. The combination of these uncertainties is 0.5%.

Sources of background from hadronic $b$-hadron decays, where two of the final-state hadrons are misidentified as muons, are neglected in the final fit to the $K^{*0} \mu^+ \mu^-$ candidates. These backgrounds are estimated to contribute 1 to 2 candidates at $m(K^- \pi^+ \mu^+ \mu^-)$ close to the known $B_s^0$ mass. The resulting systematic uncertainty on the $B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-$ yield is estimated to be 2%. The background is negligible compared to the $B^0$ yield. The background yield from $\Lambda_b^0$ decays is constrained using PID efficiencies from control samples and these efficiencies have an associated systematic uncertainty. This uncertainty is accounted for in the statistical uncertainty of the fit and is negligible.

Other sources of systematic uncertainties are associated to the normalisation of the observed yield for the measurements of the branching fraction and branching-fraction ratios. The largest source of systematic uncertainty on both $B(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-)$ and the branching-fraction ratio measurements is associated to how well external parameters are known: there is a 5.8% uncertainty on the ratio of the $B_s^0$ and $B^0$ fragmentation fractions, a 1.1% systematic uncertainty due to $F_S(B^0 \to J/\psi \bar{K}^{*0})$, a 0.8% uncertainty due to $F_S(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-)$, a 4.0% uncertainty due to $F_S(B_s^0 \to J/\psi \bar{K}^{*0})$ and a 6.8% uncertainty on $B(B^0 \to J/\psi \bar{K}^{*0})$. It is assumed that these external uncertainties are uncorrelated.

<table>
<thead>
<tr>
<th>Source</th>
<th>$\mathcal{B}(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-)$</th>
<th>Uncertainties</th>
<th>$\mathcal{B}(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-)$</th>
<th>$\mathcal{B}(B_s^0 \to J/\psi \bar{K}^{*0})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass lineshapes</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Neural network response</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Residual background</td>
<td>2.0%</td>
<td>2.0%</td>
<td>2.0%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Decay models</td>
<td>4.0%</td>
<td>4.0%</td>
<td>4.0%</td>
<td>4.0%</td>
</tr>
<tr>
<td>Non-$K^{*0}$ states</td>
<td>3.4%</td>
<td>3.4%</td>
<td>3.4%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Efficiency</td>
<td>1.3%</td>
<td>1.5%</td>
<td>1.4%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Data-simulation differences</td>
<td>2.2%</td>
<td>2.2%</td>
<td>0.8%</td>
<td>0.8%</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>6.2%</td>
<td>6.3%</td>
<td>5.9%</td>
<td>5.9%</td>
</tr>
<tr>
<td>External parameters</td>
<td>8.9%</td>
<td>5.9%</td>
<td>4.0%</td>
<td>4.0%</td>
</tr>
</tbody>
</table>

Table 1. Main sources of systematic uncertainty considered on the branching fraction measurements. The first uncertainty applies to the measurement of $\mathcal{B}(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-)$, the second to $\mathcal{B}(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-)/\mathcal{B}(B^0 \to \bar{K}^{*0} \mu^+ \mu^-)$ and the third to $\mathcal{B}(B_s^0 \to \bar{K}^{*0} \mu^+ \mu^-)/\mathcal{B}(B_s^0 \to J/\psi \bar{K}^{*0})$, respectively. A description of the different contributions can be found in the text. The first three sources of uncertainty affect the measured yield of the signal decay. The total uncertainty is the sum in quadrature of the individual sources. The final row indicates the additional uncertainty arising from the uncertainties on external parameters used in the measurements.
The second largest source of uncertainty is due to how well the amplitudes for the $B^0 \to J/\psi K^{*0}$, $B^0_s \to J/\psi K^{*0}$, $B^0 \to K^{*0} \mu^+ \mu^-$, and $B^0_s \to K^{*0} \mu^+ \mu^-$ decays are known. The uncertainty on the decay structure leads to an uncertainty on the efficiencies used to correct the observed yields. The amplitude structure of the $B^0 \to J/\psi K^- \pi^+$ decay has been studied in refs. [44, 48], and the amplitude structure of the $B^0_s \to J/\psi K^- \pi^+$ decay in ref. [49]. These measurements are used to weight the simulated events used to determine $\varepsilon$ and a systematic uncertainty is assigned as the difference of $\varepsilon$ with and without the weighting. The full angular distribution of $B^0 \to K^{*0} \mu^+ \mu^-$ has been studied by the LHCb collaboration in ref. [19]. The decay structure of the $B^0_s \to K^{*0} \mu^+ \mu^-$ decay is, however, unknown. To determine a systematic uncertainty associated to the knowledge of these decay models, the simulated samples are weighted such that the coupling strengths used in the model are consistent with the results from global fits to $b \to s$ data [20–24]. Again, the systematic uncertainty is assigned as the difference of $\varepsilon$ with and without the weighting. The total systematic uncertainty due to the knowledge of decay models is 4% for all measurements. Finally, the contribution from non-$K^{*0}$ states in the $B^0_s \to K^{*0} \mu^+ \mu^-$ is also considered. This contribution is also unknown and is assumed to be at a similar level as seen in the decay $B^0 \to K^- \pi^+ \mu^+ \mu^-$ [19]. Assigning the full size of the effect as systematic uncertainty results in a 3.4% uncertainty.

The efficiency ratios used to determine the different branching fraction measurements have an uncertainty of around 1.5%. These uncertainties comprise a statistical component due to the limited size of the simulated samples and a systematic component associated to the choice of binning in kinematic variables used to evaluate PID and track reconstruction efficiencies. A separate systematic uncertainty is also considered on the ratio of efficiencies due to data-simulation differences. This systematic uncertainty is evaluated by taking the deviation between the efficiency ratio with and without corrections described in section 2 applied. This includes corrections to the $B^0_s$ meson kinematics, PID performance and track reconstruction efficiency. This results in an additional uncertainty of 1 to 2% depending on the measurement considered.

7 Summary

A search for the decay $B^0_s \to \bar{K}^{*0} \mu^+ \mu^-$ is performed using data sets corresponding to 1.0, 2.0 and 1.6 fb$^{-1}$ of integrated luminosity collected with the LHCb experiment at centre-of-mass energies of 7, 8 and 13 TeV, respectively. A yield of $38 \pm 12 B^0\to \bar{K}^{*0} \mu^+ \mu^-$ decays is obtained, providing the first evidence for this decay with a significance of 3.4 standard deviations above the background-only hypothesis. The resulting branching fraction is determined to be

$$B(B^0_s \to \bar{K}^{*0} \mu^+ \mu^-) = [2.9 \pm 1.0 \text{(stat)} \pm 0.2 \text{(syst)} \pm 0.3 \text{(norm)}] \times 10^{-8}.$$  

This measurement is consistent with existing SM predictions of the branching fraction of the decay and a SM-like value of $|V_{td}/V_{ts}|$. A detailed analysis of the $q^2$ spectrum of the $B^0_s \to \bar{K}^{*0} \mu^+ \mu^-$ decay requires a larger data set. Such a data set should be available with the upgraded LHCb experiment [50].
Acknowledgments

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 the LHCb institutes. We acknowledge support from CERN and from the national agencies: CAPES, CNPq, FAPERJ and FINEP (Brazil); MOST and NSFC (China); CNRS/IN2P3 (France); BMBF, DFG and MPG (Germany); INFN (Italy); NWO (The Netherlands); MNiSW and NCN (Poland); MEN/IFA (Romania); MinES and FASO (Russia); MinECo (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); NSF (U.S.A.). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (The Netherlands), PIC (Spain), GridPP (United Kingdom), RRCKI and Yandex LLC (Russia), CSCS (Switzerland), IFIN-HH (Romania), CBPF (Brazil), PL-GRID (Poland) and OSC (U.S.A.). We are indebted to the communities behind the multiple open-source software packages on which we depend. Individual groups or members have received support from AvH Foundation (Germany), EPLANET, Marie Skłodowska-Curie Actions and ERC (European Union), ANR, Labex P2IO and OCEVU, and Région Auvergne-Rhône-Alpes (France), Key Research Program of Frontier Sciences of CAS, CAS PIFI, and the Thousand Talents Program (China), RFBR, RSF and Yandex LLC (Russia), GVA, XuntaGal and GENCAT (Spain), Herchel Smith Fund, the Royal Society, the English-Speaking Union and the Leverhulme Trust (United Kingdom).

A Invariant-mass distributions

In these appendices, the fits to the $J/\psi K^-\pi^+$ and $K^-\pi^+\mu^+\mu^-$ invariant mass of the selected candidates in bins of neural network response for both the Run 1 and Run 2 data sets are shown. The fit to the $K^-\pi^+\mu^+\mu^-$ invariant mass of the candidates in the $J/\psi$ mass window is shown in figure 4. This fit is used to determine the resolution and tail parameters for the $B_s^0 \rightarrow K^{*0}\mu^+\mu^-$ decay. The fit to $K^-\pi^+\mu^+\mu^-$ invariant mass of the $B_s^0 \rightarrow K^{*0}\mu^+\mu^-$ candidates is shown in figure 5. The fit to the $J/\psi K^-\pi^+$ invariant mass after application of the $J/\psi$ mass constraint is shown in figure 6.
Figure 4. Distribution of reconstructed $K^-\pi^+\mu^+\mu^-$ invariant mass of candidates in the $J/\psi$ mass window in (top four figures) the Run 1 and (bottom four figures) Run 2 data sets. The candidates are divided into four independent bins of increasing neural network response per data taking period.
Figure 5. Distribution of reconstructed $K^−\pi^+\mu^+\mu^−$ invariant mass of candidates outside of the $J/\psi$ and $\psi(2S)$ mass regions in (top four figures) the Run 1 and (bottom four figures) Run 2 data sets. The candidates are divided into four independent bins of increasing neural network response per data taking period.
Figure 6. Distribution of reconstructed $J/\psi K^−\pi^+$ invariant mass after application of a $J/\psi$ mass constraint of candidates in (top four figures) the Run 1 and (bottom four figures) Run 2 data sets. The candidates are divided into four independent bins of increasing neural network response per data taking period.
Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] T.M. Aliev and M. Savci, Exclusive $B \to \pi \ell^+\ell^-$ and $B \to \rho \ell^+\ell^-$ decays in two Higgs doublet model, *Phys. Rev.* D **60** (1999) 014005 [hep-ph/9812272] [inSPIRE].


[8] B. Kindra and N. Mahajan, *Predictions of angular observables for $B_s \to K^*\ell\ell$ and $B \to \rho\ell\ell$ in Standard Model*, arXiv:1803.05876 [inSPIRE].


[10] LHCb collaboration, *Observation of the suppressed decay $\Lambda^0_c \to p\pi^- \mu^+\mu^-$*, *JHEP* **04** (2017) 029 [arXiv:1701.08705] [inSPIRE].


[16] CDF collaboration, T. Aaltonen et al., Measurement of the forward-backward asymmetry in the $B \to K^{(*)\mu^+\mu^-}$ decay and first observation of the $B^0_d \to \phi\mu^+\mu^-$ decay, Phys. Rev. Lett. 106 (2011) 161801 [arXiv:1101.1028] [SPIRE].


[18] LHCb collaboration, Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using $3 fb^{-1}$ of integrated luminosity, JHEP 02 (2016) 104 [arXiv:1512.04442] [SPIRE].

[19] LHCb collaboration, Measurements of the $S$-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^+(892)^0\mu^+\mu^-$ differential branching fraction, JHEP 11 (2016) 047 [Erratum ibid. 04 (2017) 142] [arXiv:1606.04731] [SPIRE].


[34] LHCb collaboration, Handling of the generation of primary events in Gauss, the LHCb simulation framework, J. Phys. Conf. Ser. 331 (2011) 032047 [SPIRE].


School of Physics, University College Dublin, Dublin, Ireland
Sezione INFN di Bari, Bari, Italy
Sezione INFN di Bologna, Bologna, Italy
Sezione INFN di Cagliari, Cagliari, Italy
Sezione INFN di Ferrara, Ferrara, Italy
Sezione INFN di Firenze, Firenze, Italy
Laboratori Nazionali dell’INFN di Frascati, Frascati, Italy
Sezione INFN di Genova, Genova, Italy
Sezione INFN di Milano Bicocca, Milano, Italy
Sezione INFN di Milano, Milano, Italy
Sezione INFN di Padova, Padova, Italy
Sezione INFN di Pisa, Pisa, Italy
Sezione INFN di Roma Tor Vergata, Roma, Italy
Sezione INFN di Roma La Sapienza, Roma, Italy
Henryk Niewodniczanski Institute of Nuclear Physics Polish Academy of Sciences, Kraków, Poland
AGH - University of Science and Technology, Faculty of Physics and Applied Computer Science, Kraków, Poland
National Center for Nuclear Research (NCBJ), Warsaw, Poland
Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest-Mugurele, Romania
Petersburg Nuclear Physics Institute (PNPI), Gatchina, Russia
Institute of Theoretical and Experimental Physics (ITPE), Moscow, Russia
Institute of Nuclear Physics, Moscow State University (SINP MSU), Moscow, Russia
Institute for Nuclear Research of the Russian Academy of Sciences (INR RAS), Moscow, Russia
Yandex School of Data Analysis, Moscow, Russia
Budker Institute of Nuclear Physics (SB RAS), Novosibirsk, Russia
Institute for High Energy Physics (IHEP), Protvino, Russia
ICCUB, Universitat de Barcelona, Barcelona, Spain
Instituto Galego de Física de Altas Enerxías (IGFAE), Universidade de Santiago de Compostela, Santiago de Compostela, Spain
European Organization for Nuclear Research (CERN), Geneva, Switzerland
Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Physik-Institut, Universität Zürich, Zürich, Switzerland
Nikhef National Institute for Subatomic Physics, Amsterdam, The Netherlands
Nikhef National Institute for Subatomic Physics and VU University Amsterdam, Amsterdam, The Netherlands
NSC Kharkiv Institute of Physics and Technology (NSC KIPT), Kharkiv, Ukraine
Institute for Nuclear Research of the National Academy of Sciences (KINR), Kyiv, Ukraine
University of Birmingham, Birmingham, United Kingdom
H.H. Wills Physics Laboratory, University of Bristol, Bristol, United Kingdom
Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
Department of Physics, University of Warwick, Coventry, United Kingdom
STFC Rutherford Appleton Laboratory, Didcot, United Kingdom
School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom
Imperial College London, London, United Kingdom
School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom
Department of Physics, University of Oxford, Oxford, United Kingdom
Massachusetts Institute of Technology, Cambridge, MA, United States
University of Cincinnati, Cincinnati, OH, United States
University of Maryland, College Park, MD, United States
Syracuse University, Syracuse, NY, United States
Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, associated to
University of Chinese Academy of Sciences, Beijing, China, associated to
School of Physics and Technology, Wuhan University, Wuhan, China, associated to
Institute of Particle Physics, Central China Normal University, Wuhan, Hubei, China, associated to
Departamento de Física, Universidad Nacional de Colombia, Bogotá, Colombia, associated to
Institut für Physik, Universität Rostock, Rostock, Germany, associated to
National Research Centre Kurchatov Institute, Moscow, Russia, associated to
National University of Science and Technology MISIS, Moscow, Russia, associated to
National Research Tomsk Polytechnic University, Tomsk, Russia, associated to
Instituto de Física Corpuscular, Centro Mixto Universidad de Valencia - CSIC, Valencia, Spain, associated to
Van Swinderen Institute, University of Groningen, Groningen, The Netherlands, associated to
Los Alamos National Laboratory (LANL), Los Alamos, United States, associated to
Universidade Federal do Triângulo Mineiro (UFTM), Uberaba-MG, Brazil
Laboratoire Leprince-Ringuet, Palaiseau, France
P.N. Lebedev Physical Institute, Russian Academy of Science (LPI RAS), Moscow, Russia
Università di Bari, Bari, Italy
Università di Bologna, Bologna, Italy
Università di Cagliari, Cagliari, Italy
Università di Ferrara, Ferrara, Italy
Università di Genova, Genova, Italy
Università di Milano Bicocca, Milano, Italy
Università di Roma Tor Vergata, Roma, Italy
Università di Roma La Sapienza, Roma, Italy
AGH - University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland
LIFAELS, La Salle, Universitat Ramon Llull, Barcelona, Spain
Hanoi University of Science, Hanoi, Vietnam
Università di Padova, Padova, Italy
Università di Pisa, Pisa, Italy
Università degli Studi di Milano, Milano, Italy
Università di Urbino, Urbino, Italy
Università della Basilicata, Potenza, Italy
Scuola Normale Superiore, Pisa, Italy
Università di Modena e Reggio Emilia, Modena, Italy
MSU - Iligan Institute of Technology (MSU-IIT), Iligan, Philippines
Novosibirsk State University, Novosibirsk, Russia
National Research University Higher School of Economics, Moscow, Russia
Escuela Agrícola Panamericana, San Antonio de Oriente, Honduras
† Deceased