Title
Search for the decay $B_s^0 \to \mu^+\mu^-$ with the ATLAS detector

Permalink
https://escholarship.org/uc/item/77x05006

Journal
Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics, 713(4-5)

ISSN
0370-2693

Authors
Aad, G
Abbott, B
Abdallah, J
et al.

Publication Date
2012-07-18

DOI
10.1016/j.physletb.2012.06.013

License
CC BY 4.0

Peer reviewed
Search for the decay $B_s^0 \to \mu^+ \mu^-$ with the ATLAS detector

ATLAS Collaboration

1. Introduction

Flavour changing neutral current processes are highly suppressed in the Standard Model (SM), and therefore their study is of particular interest in the search for new physics. The SM predicts the branching fraction for the decay $B^0 \to \mu^+ \mu^-$ to be extremely small: $(3.5 \pm 0.3) \times 10^{-9}$ [1–4]. This process might be substantially enhanced by coupling to non-SM heavy particles, such as those predicted by the Minimal Supersymmetric Standard Model [5–11] and other extensions [12]. Upper limits on this branching fraction, in the range $(0.45–5.1) \times 10^{-8}$, have been reported by the D0 [13], CDF [14], CMS [15,16] and LHCb [17,18] Collaborations. This Letter reports the result of a search performed with $pp$ collisions corresponding to an integrated luminosity of 2.4 fb$^{-1}$, collected in the first half of the 2011 data-taking period using the ATLAS detector at the LHC.

The analysis is based on events selected with a di-muon trigger and reconstructed in the ATLAS inner tracking detector and muon spectrometer [19]. Details of the detector, trigger and datasets are discussed in Section 2, together with the preselection criteria.

The $B^0 \to \mu^+ \mu^-$ branching fraction is measured with respect to a prominent reference decay ($B^\pm \to J/\psi K^\pm$) in order to minimize systematic uncertainties in the evaluation of the efficiencies and acceptances, while still providing small statistical uncertainties. The branching fraction can be written as

$$\text{BR}(B_s^0 \to \mu^+ \mu^-) = \text{BR}(B^\pm \to J/\psi K^\pm \to \mu^+ \mu^- K^\pm) \times \frac{f_u}{f_s} \times \frac{N_{\mu^+ \mu^-}}{N_{J/\psi K^\pm}} \times \frac{A_{J/\psi K^\pm}}{A_{\mu^+ \mu^-}} \times \epsilon_{\mu^+ \mu^-},$$

(1)

where the right-hand side includes the $B^\pm \to J/\psi K^\pm \to \mu^+ \mu^- K^\pm$ branching fraction, the relative production probability of $B^\pm$ and $B_s^0$, $f_u/f_s$, taken from previous measurements [20–22], the event yields after background subtraction, and the acceptance and efficiency ratios. The event yields for both signal and reference channels were obtained from signal and sideband (background) regions defined in the invariant-mass spectrum (see Table 1).

The Single Event Sensitivity (SES) corresponds to the $B_s^0 \to \mu^+ \mu^-$ branching fraction which would yield one observed signal event in the data sample:

$$\text{BR}(B_s^0 \to \mu^+ \mu^-) = N_{\mu^+ \mu^-} \times \text{SES},$$

(2)

where $N_{\mu^+ \mu^-}$ is the number of observed events.

This Letter describes the results of a blind analysis in which the di-muon mass region 5066 to 5666 MeV was removed from the analysis until the procedures for event selection, signal and limit extraction were fully defined. Sections 3.1 to 3.3 discuss the variables used in the event selection, Monte Carlo (MC) tuning and background studies. The final sample of candidates was selected with a multivariate classifier, trained on a fraction of the events from the di-muon invariant-mass sidebands, as discussed in Section 3.4. The relative efficiency and event yields in the reference channel are discussed in Sections 4.1 and 4.2, respectively. The signal extraction is discussed in Section 5 and the corresponding limit on the branching fraction is presented in Section 6.
According to the SM, the branching fraction $BR(B^0 \rightarrow \mu^+ \mu^-)$ is predicted to be about 30 times smaller than $BR(B^{+} \rightarrow \mu^+ \mu^-)$ [1,2]. Therefore, despite the increased SES of approximately a factor four due to the absence of the factor $f_u/f_d$ and possible enhancements due to new physics, the sensitivity to this channel is beyond the reach of the current analysis. Hence only a limit on $BR(B^0 \rightarrow \mu^+ \mu^-)$ was derived by assuming $BR(B^0 \rightarrow \mu^+ \mu^-)$ to be negligible.

2. ATLAS detector, data and simulation samples

The ATLAS detector\(^1\) consists of three main components: an Inner Detector tracking system (ID) immersed in a 2 T magnetic field, a system of electromagnetic and hadronic calorimeters, and an outer Muon Spectrometer (MS). A full description can be found in [19]. The detector performance characteristics most relevant to this analysis are the vertex-finding and the overall track reconstruction in the ID and MS, together with the ability of the trigger system to record events containing pairs of muons.

The ID provides precise track reconstruction within the pseudorapidity range $|\eta| < 2.5$. It employs a Pixel detector close to the beam pipe, a silicon microstrip detector (SCT) at intermediate radii and a Transition Radiation Tracker (TRT) at outer radii. The innermost Pixel layer is located at a radius of 50.5 mm and plays a key role in precise vertex determination.

The MS comprises separate trigger and high-precision tracking chambers that measure the deflection of muons in a toroidal magnetic field. The precision chambers cover the region $|\eta| < 2.7$ and measure the coordinate in the bending plane. The trigger chambers cover the range $|\eta| < 2.4$ and provide fast coarser measurements in both the bending and non-bending plane.

This analysis is based on a sample of $pp$ collisions at $\sqrt{s} = 7$ TeV, recorded by ATLAS in the period April–August 2011. Trigger and pile-up conditions changed for data taken after this period: the remainder of the 2011 dataset will be included in a future analysis. Data used in the analysis were recorded during stable LHC beam periods. Further data quality requirements were also imposed, notably on the performance of the MS and ID systems. The total integrated luminosity amounts to 2.4 fb\(^{-1}\). This sample has an average of about five primary vertices per event from multiple proton–proton interactions.

A muon trigger [23] was used to select events. In particular, the sample contains events seeded by a Level-1 di-muon trigger which required a transverse momentum $p_T > 4$ GeV for both muon candidates. A full track reconstruction of the muon candidates was performed at the second and third trigger levels, where additional cuts on the di-muon invariant mass $m_{\mu^+ \mu^-} > 1.5$ GeV were applied, loosely selecting events compatible with $J/\psi$ (2500 to 4300 MeV) or $Br_0^+ (4000$ to 8500 MeV) decays into a muon pair.

Events containing candidates for $B_0^+ \rightarrow \mu^+ \mu^-$, $B^+ \rightarrow J/\psi K^+ \rightarrow \mu^+ \mu^- K^+$ and, as discussed in Sections 3.2 and 3.3, $B_0^+ \rightarrow J/\psi \phi \rightarrow \mu^+ \mu^- K^+ K^-$ were retained for this analysis. After cutting on the mass of the intermediate resonances (1009 MeV $< m_{\psi} < 1031$ MeV, 2915 MeV $< m_{J/\psi} < 3175$ MeV) a preselection was applied, based on track properties and the quality of the reconstructed $B$ decay vertex. All charged particle tracks reconstructed in the ID were required to have at least one Pixel, six SCT and eight TRT hits. Tracks were required to have $|\eta| < 2.5$ and $p_T > 4$ GeV (> 2.5 GeV) for muon (kaon) candidates. No particle identification was used to distinguish $K^\pm$ and $\pi^\pm$ candidates. ID tracks that were matched to reconstructed MS tracks were selected as candidate muons. Decay vertices were formed by combining two, three or four tracks, according to the specific decay process [24]. All $B$-meson properties were computed based on the result of the fit of the tracks to the $B$ decay vertex. In order to reject fake track combinations, the fit $\chi^2$ per degree of freedom was required to be less than 2.0 (85% efficient) for $B^+ \rightarrow \mu^+ \mu^-$ and less than 6.0 (99.5% efficient) for the other channels. All reconstructed $B$ candidates were required to satisfy $p_T > 8$ GeV and $|\eta| < 2.5$ in order to define our efficiencies and acceptances within a fiducial phase-space volume with as little as possible reliance on MC extrapolations. Signal and sideband regions were defined according to Table 1.

The primary vertex position was obtained from a fit of charged tracks not used in the decay vertex and constrained to the interaction region of the colliding beams. If multiple candidate primary vertices were present, the one closest in $z$ to the decay vertex was chosen. After preselection, approximately $2 \times 10^5 B_0^+ \rightarrow \mu^+ \mu^-$ and $1.4 \times 10^5 B^+ \rightarrow J/\psi K^+$ candidates were obtained in the signal regions.

Samples of Monte Carlo (MC) events were used for the extraction of acceptance and efficiency ratios. MC samples were produced for the signal channel $B_0^+ \rightarrow \mu^+ \mu^-$, the reference channel $B^+ \rightarrow J/\psi K^+ (-J/\psi \rightarrow \mu^+ \mu^-)$ and the control channel $B_0^+ \rightarrow J/\psi \phi (-\phi \rightarrow K^+ K^-)$. These samples were generated with Pythia 6.4 [25] using the 2010 ATLAS [24,26] tune. MC events were filtered before detector simulation to ensure the presence of at least one decay of interest, with $B$ decay products all satisfying $|\eta| < 2.5$ and $p_T > 2.5(0.5)$ GeV for muons (kaons). An additional sample was generated with a fictitious value of the $B_0^+$ mass (6500 MeV) and the same parameters as the standard $B_0^+ \rightarrow \mu^+ \mu^-$ sample, allowing a check of the full analysis on a signal-free region before unblinding. The ATLAS detector and its response were simulated using Geant4 [27]. Additional pp interactions in the same and nearby bunch crossings (pile-up) were included in the simulation.

3. Event selection

This section describes the expected background composition, the discriminating variables used as input to the multivariable classifier, the tuning of the simulation for the determination of the signal efficiency, the data samples used to estimate the background rejection and the optimization procedure. The signal efficiency was determined from MC samples, re-weighted to account for differences between data and MC simulation of the $B$-meson kinematics. The rejection power was tested using a sub-sample of background events from the sidebands in the di-muon mass spectrum.

3.1. Background composition

Two categories of background were considered: a continuum with a smooth dependence on the di-muon invariant mass, and sources of resonant contributions from mis-reconstructed decays.

---

\(^1\) ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point. The z-axis is along the beam pipe, the x-axis points to the centre of the LHC ring and the y-axis points upward. Cylindrical coordinates ($r, \phi$) are used in the transverse plane, $\phi$ being the azimuthal angle around the beam pipe. The pseudorapidity $\eta$ is defined as $\eta = -\ln \tan(\theta/2)$, where $\theta$ is the polar angle.
Comparisons of data and MC have shown that the combinatorial background from $bb \to \mu^+\mu^-X$ decays provides a reasonable description for the distributions of the discriminating variables for the events found in the sidebands. The $bb \to \mu^+\mu^-X$ MC sample used is equivalent to about 12 pb$^{-1}$ of integrated luminosity. Such studies support the procedure of modeling the continuum background through interpolation of the di-muon yield in the sidebands, but do not reach a sufficient statistical precision. Half of the data events in the sidebands (those with odd event numbers) were used to optimize the selection procedure. The remaining events were used for the measurement of the background and for interpolation to the signal region.

Resonant background is due to B decay candidates containing either one or two hadrons erroneously identified as muons. Mis-identification may be due to punch-through of a hadron to the MS or to decays in flight where the muon carries most of the hadron momentum. In either case the hadron fakes the muon signature for the purpose of this analysis. Single-fake events are due to, e.g. $B_0^0 \to K^+\mu^-\nu$, the charged K meson being mis-identified as a muon. Double-fake events are due to two-body hadronic B decays ($B \to h\mu$), e.g. $B_0^0 \to K^+\pi^-$, where both hadrons are mis-identified as muons. MC studies have shown that double-fake events are the main source of resonant background after the selection criteria used in this analysis. The main contribution is from $B_0^0 \to K^+K^-$, followed by $B_0^0 \to \pi^+\pi^-$ and $B^0 \to K^+\pi^\mp$ [20,28].

The simulation determined the probability for a hadron to be misidentified as a muon to be equal to 2(4)% for $\pi^{\pm}$ ($K^{\pm}$), with a relative uncertainty of 20%, validated against control samples in data [29]. The value for charged K mesons was averaged over $K^+$ and $K^-$ and was found consistent with the preliminary results of data-driven studies based on the decay $D^{\pm} \to D_0^\mp \pi$. The expected event yield for $B \to h\mu$ was obtained from an estimation of the integrated luminosity, acceptance and efficiency. This constitutes a nearly irreducible background in this analysis, due to its resemblance to the actual signal.

3.2. Discriminating variables

Table 2 describes the discriminating variables used in the multivariate classifier. The $B_0^0 \to \mu^+\mu^-\mu^-$ signal is characterized by the separation between the production (primary) and decay (secondary) vertices, as well as the two-body decay topology. These variables exploit such features to discriminate against potential backgrounds: pairs of prompt charged tracks (e.g. $L_{xy}$, ct significance, $x_0^2$), as well as pairs of displaced muons originating from $bb \to \mu^+\mu^-X$ processes (e.g. $d_0^{\text{max}}, d_0^{\text{min}}$), secondary vertices with additional particles in the final state (e.g. $\sigma_{xy}, \Delta R, D_{xy}^{\text{min}}, D_2^{\text{min}}$) and non-$bb$ processes (e.g. $I_{0.7}$, $p_T^\mu, p_T^{\text{max}} p_T^{\text{min}}$).

Fig. 1 shows how the discriminating variables are distributed for signal and background. Among the discriminating variables, isolation ($I_{0.7}$) is expected to have the largest pile-up dependence. In order to minimize this dependence, the definition of $I_{0.7}$ was restricted to only include tracks originating from the primary vertex associated with the B decay. This specification makes the selection independent of pile-up, as shown in Fig. 2, where the efficiency of the selection for $B^+ \to J/\psi K^+$ is shown for events with different numbers of reconstructed primary vertices, both in sideband-subtracted data and MC.

The variable $I_{0.7}$ might also be subject to differences between $B_0^0$ and $B^0$ in the distributions of the surrounding hadrons, e.g. with harder $p_T$ spectra for kaons produced in association with the $B_0^0$ in the b-quark fragmentation. As predicted by MC, significant differences were observed between $B^+ \to J/\psi K^+$ and the control channel $B^0_\gamma \to J/\psi \phi$ in the $I_{0.7}$ distribution from data. Within statistical uncertainties, the $I_{0.7}$ distribution from the MC simulation of the control channel $B_0^0 \to J/\psi \phi$ was verified to be consistent with the corresponding sideband-subtracted signal in data.

3.3. MC re-weighting and comparison to data

Monte Carlo samples were produced for the signal, reference and control channels, with specific requirements on the B-meson decay products as described above in Section 2. In order to ensure that the data are reproduced as closely as possible, the simulation was tuned by an iterative re-weighting procedure: a generator-level (GL) re-weighting based on simulation, followed by a data-driven (DD) re-weighting.

For the GL re-weighting, additional MC samples were generated without selection on the final states and over a wider range in the b-quark kinematics: $|\eta^B| < 4$ and $p_T^B > 2.5$ GeV. These samples allowed a binned ($p_T^B, \eta^B$) map of the efficiencies of the generator-level selections to be derived for both the signal and the reference MC. The inverse of such efficiencies was then used to weight events individually, thus correcting the GL biases. These corrections were applied independently to the simulated reference and signal channel samples to correct for the biases in the relative $B^+_0/B^{\pm}$ acceptance induced by the generator-level selection. Possible residual biases were found to be negligible within the fiducial region $|\eta^B| < 2.5$ and $p_T^B > 8.0$ GeV.

Residual ($p_T^B, \eta^B$) differences between data and MC were observed after GL re-weighting. These were addressed with the DD re-weighting procedure, based on the comparison of MC events to the large sample of $B^+ \to J/\psi K^+$ decays in collision data. In order not to correlate the re-weighting procedure with the yield measurement, only candidates with odd event numbers in the ATLAS dataset were used in this procedure, while the remaining sample was used for the yield measurement.

DD weights were determined by an iterative method, comparing re-weighted MC events with sideband-subtracted $B^+_0 \to J/\psi K^+$ events in data. The procedure was applied separately to the B-meson variables $p_T^B$ and $\eta^B$ due to the limited num-
Fig. 1. Signal (filled histogram) and sideband (empty histogram) distributions for the selection variables described in Table 2. The $B_0^0 \rightarrow \mu^+ \mu^-$ signal (normalized to the background histogram) is from simulation and the background is from data in the invariant-mass sidebands.
Correction yields compatible weighting before the DD re-weighting, and by verifying that this when compared to statistical uncertainties.

be smaller than 1 the finite resolution in the measured variables were estimated to found to converge to the expected distributions. Effects related to the finite resolution in the measured variables. Fig. 3 shows comparisons for preselection. Agreement between MC and data was found for most dis- cases in Section 3.3 (corresponding to the signal efficiency defined for $|p_T^B| < 2.5$ and $p_T^B > 8.0$ GeV) and the background yield for a given set of cuts. The extraction of $N_{\text{bkg}}$ is performed by sideband interpolation as described in Section 5. The coefficient $a$ was determined by the confidence level (CL) sought in the analysis, with $a = 2$ for a 95% CL limit. This quantity is specifically designed to optimize the performance of a frequentist limit determination in a counting analysis [30].

First, a simplified optimization procedure was performed on a small set of variables that includes: $|z_{2D}|$, $I_{0.7}$, $ct$, and width $\Delta m$ of the search window centred around the $B^0$ mass (rounded to 5366 MeV). A four-dimensional scan was performed on the four variables, using odd-numbered events in the sidebands. The optimal selection cuts are shown in Table 3, where the signal efficiency $A_{\text{sig}}$, $\epsilon_{\text{sig}}$, and the background estimated from sidebands interpolation and the value of $P$ are also given. This selection serves as a benchmark for the optimization of the multivariate analysis described in Section 3.4.2.
4. Single event sensitivity ingredients

4.1. Relative acceptance and efficiency

The ratio of the acceptance times efficiency products for the charged and neutral decays

\[ R_{\text{AE}} = \frac{\langle A_{\lambda \psi K} \epsilon_{J/\psi K} \rangle}{\langle A_{\mu^+\mu^-} \epsilon_{\mu^+\mu^-} \rangle} \]

is required for the determination of the SES (Eq. (1)). The same BDT, trained on the \( B^0 \) signal MC sample and di-muon data sidebands, was used to select both decay modes.

The uncertainty on \( R_{\text{AE}} \) is affected by differences between data and MC in the distributions of the discriminating variables. Such differences are reduced by the data-driven corrections applied to the MC \( B \)-meson kinematics. Furthermore, only deviations that act incoherently between the signal and the reference channel contribute to the uncertainty on \( R_{\text{AE}} \). These effects were studied by observing the change in the relative efficiency of the BDT selection when the simulated events were re-weighted by the data-to-MC ratio of the distributions of the most sensitive variables in \( B^\pm \rightarrow J/\psi K^\pm \) events. The procedure was performed with the cut on the BDT output fixed at the optimal value for each of the three event categories. Conservatively, the corresponding variations in \( R_{\text{AE}} \) were combined linearly and taken as systematic uncertainties.

Due to large correlations between \( L_{xy}, \chi^2_{xy} \) and \( ct \)-significance, correcting for the differences in \( L_{xy} \) between data and simulation was found to also effectively remove differences in the other two variables. Therefore only \( L_{xy} \) was considered, since it induced the largest deviation in \( R_{\text{AE}} \). Differences in the \( \eta \) and \( p_T \) distributions of the final state particles, the hit multiplicity in the Pixel detector, and the multiplicity of reconstructed primary vertices were included in the systematic uncertainty evaluation.

Fig. 5 shows the distribution of the BDT output for MC samples of \( B^0 \rightarrow \mu^+\mu^- \) and \( B^\pm \rightarrow J/\psi K^\pm \) decays, with a signal–background comparison for \( B^0 \rightarrow \mu^+\mu^- \) and a sideband-subtracted data–MC comparison for \( B^\pm \rightarrow J/\psi K^\pm \). As shown in Table 4, BDT output and \( \Delta m \) cuts for each mass-resolution category, optimized according to the method described in the text.

| \( |\eta|_{\text{max}} \) range | 0–1.0 | 1.0–1.5 | 1.5–2.5 |
|-------------------------|-------|-------|-------|
| Invariant-mass window [MeV] | ±116  | ±133  | ±171  |
| BDT output threshold     | 0.234 | 0.245 | 0.270 |

Residual correlations in the BDT output were studied through the search for a fictitious decay \( X \rightarrow \mu^+\mu^- \) with \( m_X = 6500 \text{MeV} \). A Monte Carlo sample was used to provide reference signal events, while data in the mass intervals 5900 to 6200 MeV and 6800 to 7000 MeV were used as background. The BDT training and selection optimization were consistently performed on odd-numbered events. Fig. 4 shows the BDT output as a function of the di-muon mass, over the sideband regions and the fictitious signal region (6200 to 6800 MeV), which was not used in the optimization. No significant mass dependence was observed.

The optimization of the multivariate analysis was performed in the six-dimensional space of \( \Delta m \) and the BDT output cuts for each of the mass-resolution categories. The independence of the BDT output on \( m_{\mu^+\mu^-} \) and the complementarity of the samples allow the factorization of the individual cut efficiencies. Each efficiency curve was interpolated with analytical models, allowing the numerical maximization of \( P \) and yielding the optimal cuts reported in Table 4.

3.4.1. Categories of invariant-mass resolution

The ability to resolve a small \( B^0 \rightarrow \mu^+\mu^- \) signal from the continuum background depends on the width \( \Delta m \) of the search region and is therefore affected by the resolution. The latter varies considerably over different sub-samples of muon pairs measured by ATLAS, due to the increase in multiple scattering and the decrease in the magnetic field integral at large values of \( |\eta| \). The non-resonant background invariant-mass distribution was observed to be relatively independent of \( \eta \). As a consequence, different mass-resolution categories correspond to different signal-to-background conditions.

In the statistical analysis, regions of different mass resolution and hence signal-to-background ratio were separated in order to optimize them independently. The sample was separated into three categories, defined by the larger pseudorapidity value \( |\eta|_{\text{max}} \) of the two muons in each event. The three categories were defined by the intervals \( |\eta|_{\text{max}} = 0–1, 1–1.5 \) and \( 1.5–2.5 \). The corresponding average values of the mass resolution are approximately 60, 80 and 110 MeV, respectively. The relative population of each interval, in \( B^0 \rightarrow \mu^+\mu^- \) signal MC, amounts to 51%, 24% and 25%.

The same classification, based on \( |\eta|_{\text{max}} \), was used for the reference channel \( B^\pm \rightarrow J/\psi K^\pm \), and separate values of the acceptance-times-efficiency ratio were obtained for each category, as discussed in Section 4.1.

3.4.2. Multivariate selection

The selection with optimal cuts was used to validate the multivariate analysis tool used for the final results. The TMVA package [31] implementation of Boosted Decision Trees (BDT) was found to have the best performance and was selected for this analysis. As a first step, it was verified that for fixed values of \( \Delta m \), the optimal BDT corresponds to selections in the variables \( \alpha_{2D}, c_t \) and \( I_{0.7} \) directly comparable to those obtained with the cuts shown in Table 3. Next, the discriminating variables of Table 2 were introduced one-by-one into the BDT, verifying that the multivariate optimization increased the signal efficiency and the value of \( P \). With the BDT approach the \( P \) estimator improved from \( P = 0.010 \) found in the simplified optimization to \( P = 0.016 \).

In order to avoid biases in the background interpolation, the BDT selection should be insensitive to the mass of the muon pair. The BDT inputs have no correlation with the invariant mass.
ficiency as derived from simulation of the vertex reconstruction efficiency due to the data–MC discrepancy in vertex reconstruction efficiency. The change is highly correlated between the two channels: the BDT cut varying between 10% and 20% depending on the mass-resolution category. The systematic uncertainties in Table 4 were derived separately in the three mass-resolution categories. The MC-based efficiency was compared with that from tuned MC samples (triangles) and sideband-subtracted data (stars).

| $|\eta|_{\text{max}}$ range | $R_{\text{eff}}$ | $\Delta$ % stat. | $\Delta$ % syst. |
|---------------------------|-----------------|----------------|-----------------|
| 0–1.0                     | 0.274           | 3.1            | 3.1             |
| 1.0–1.5                   | 0.202           | 4.8            | 5.5             |
| 1.5–2.5                   | 0.143           | 5.3            | 5.9             |

Table 5: Values of the acceptance-times-efficiency ratio $R_{\text{eff}}$ between reference and search channel, shown separately for the different categories in mass resolution.

The selection required the BDT output to exceed 0.23–0.27, depending on the mass-resolution category. The systematic uncertainties induce a fractional change in the number of events passing the BDT cut varying between 10% and 20% depending on the category. This change is highly correlated between the two channels: the corresponding variation on the efficiency ratio is 0.6%, which was taken as a systematic uncertainty and is smaller than the ±2.3% error due to the finite MC statistics.

The value of $R_{\text{eff}}$ and its systematic uncertainties (shown in Table 5) were derived separately in the three mass-resolution categories. The MC-based efficiency was compared with that from $B^{\pm} \rightarrow J/\psi K^{\pm}$ data, computing the efficiency of the BDT cut relative to the preselection. The results are of similar precision and fully consistent: 0.258 ± 0.013(stat) for the data and 0.234 ± 0.014(stat) ± 0.011(syst) for MC.

Additional smaller contributions to the uncertainty on $R_{\text{eff}}$ are due to the data–MC discrepancy in vertex reconstruction efficiency (±2%) [24], the uncertainty on the absolute $K^{\pm}$ reconstruction efficiency as derived from simulation of the $B^{\pm} \rightarrow J/\psi K^{\pm}$ reference channel (±5%) and asymmetry differences in detector response to $K^{+}$ and $K^{-}$ mesons (±1%).

4.2. $B^{\pm} \rightarrow J/\psi K^{\pm}$ event yield

The reference channel yield $N_{J/\psi K^{\pm}}$ was determined from a binned likelihood fit to the invariant-mass distribution of the $\mu^{+}\mu^{-}K^{\pm}$ system, performed in the mass range 4930–5630 MeV. To avoid any bias induced by the DD re-weighting of the MC samples discussed in Section 3.3, only even-numbered events were used in the extraction of the $B^{\pm} \rightarrow J/\psi K^{\pm}$ event yield. The $B^{\pm}$ signal was modelled with two Gaussian distributions of equal mean value. The background was modelled with the sum of: (a) an exponential function for the continuum combinatorial background; (b) an exponential function multiplied by a complementary error function describing the low-mass ($m < 5200$ MeV) contribution for partially reconstructed decays (such as $B \rightarrow J/\psi K^{*}$, $B \rightarrow J/\psi K$ (1270) and $B \rightarrow \chi_c K$); and (c) a Gaussian function for the background from $B^{\pm} \rightarrow J/\psi \pi^{\pm}$. Fig. 6 shows the invariant-mass distribution and the result of the fit for the selected data sample.

All parameters describing the signal and background were determined from the fit, with the exception of the mass and the width of the last component (c), which were obtained from simulation. The fit was performed for each of the three categories of mass resolution.

Systematic uncertainties affecting the extracted reference yield were estimated by varying the fit model: use of different bin sizes (10 or 25 MeV and unbinned), different models for signal and continuum background, inclusion of event-wise di-muon mass resolution. The resulting $B^{\pm}$ yields are given with their statistical and systematic uncertainties in Table 6.
5. Inputs to the limit extraction

The evaluation of the SES requires as input the combined branching fraction for the reference channel \( B^0 \rightarrow J/\psi K^0 \rightarrow \mu^+ \mu^- K^0 \), which is \((6.01 \pm 0.21) \times 10^{-5}\) [20]. The relative production rate of \( B^0 \) relative to \( B^+ \) \( f_B/f_B \) is \(0.267 \pm 0.021\) [22], assuming \( f_B = f_B \) (following Ref. [21]) and no kinematic dependence of \( f_B/f_B \). The ratio of acceptance-times-efficiency is discussed in Section 4 and presented in Table 5. The branching fractions uncertainties, those on \( f_B/f_B \), together with those mentioned in the last paragraph of Section 4.1, were treated coherently in the three categories of mass resolution.

In each mass-resolution category the \( B^0 \rightarrow \mu^+ \mu^- \) signal yield \( N_{\mu^+ \mu^-} \) was obtained from the number of events observed in the search window, the number of background events in the sidebands, and the small amount of resonant background discussed in Section 3.1. The expected ratio of the background events in the sidebands to those in the search window is described by the parameter \( \lambda_{\text{bkg}} \), which depends on the width of the invariant-mass interval and on the fraction of events from the sidebands used for the interpolation. The former varies according to the mass-resolution category, and the latter is equal to 50%, corresponding to the even-numbered events in the data collection. Uncertainties in the mass dependence of the continuum background produced a ±4% systematic error in the value of \( \lambda_{\text{bkg}} \), evaluated by studying the variation of \( \lambda_{\text{bkg}} \) for different BDT output cuts and background interpolation models. The systematic variation accounts also for additional background components in the low mass sidebands (e.g. partially reconstructed \( B \) decays). This uncertainty was treated coherently in the three mass-resolution categories.

The values of the SES are given in Table 7 which also shows the values of the parameters \( \lambda_{\text{bkg}} \), the background counts in the sidebands, \( \mu \), the resonant background, and finally the observed number of events in the search region, as found after unblinding. Fig. 7 shows the invariant-mass distribution of the selected candidates in data, for the three mass categories, together with the signal projections as obtained from MC assuming \( \text{BR}(B^0 \rightarrow \mu^+ \mu^-) = 3.5 \times 10^{-8} \) (i.e. approximately 10 times the SM expectation).

6. Branching fraction limits

The upper limit on the \( B^0 \rightarrow \mu^+ \mu^- \) branching fraction was obtained by means of an implementation [32] of the CLs method [33]. The extraction was based on the likelihood:

\[
L = \text{Gauss}(\epsilon_{\text{obs}}|\epsilon, \sigma_{\epsilon}) \times \text{Gauss}(R_{\text{obs}}|R, \sigma_{R}) \times \prod_{i=1}^{N_{\text{bin}}} \text{Poisson}(N_{\text{obs}}|\epsilon_i \BR + N_{\text{bkg}}^{\text{bbk}} + N_{\text{S}}^{\theta \rightarrow \eta \eta}) \\
	imes \text{Poisson}(N_{\text{bkg}}^{\text{bbk}}| \epsilon_i R_{\text{bkg}} N_{\text{bkg}}) \\
	imes \text{Gauss}(\epsilon_{\text{obs}}|\epsilon_i, \sigma_{\epsilon_i}).
\]

For each mass-resolution category, the likelihood contains Poisson distributions for the event counts in the search and sideband regions and a Gaussian distribution for the relative efficiency \( \epsilon_i \). Two additional Gaussians describe the coherent systematic uncertainties in \( R_{\text{bkg}} \) and in the SES. The mean of the Poisson distribution in the search region is equal to the sum of the \( B^0 \) branching fraction (scaled by the normalization and relative efficiency parameters), the continuum background and the resonant background. The mean of the Poisson distribution in the sidebands is equal to the background scaled by \( R_{\text{bkg}} \). The parameters \( \sigma_{\epsilon} (\sigma_{\epsilon_i}), \sigma_{R_{\text{bkg}}} (\sigma_{R_{\text{bkg}}}) \) account for the correlated (uncorrelated) uncertainties in the SES and the background scaling factor. In this analysis the uncertainties on \( R_{\text{bkg}} \) are negligible, with \( R_{\text{bkg}} = 1.00 \pm 0.04 \). All input parameters are summarized in Table 7.

The expected limits were obtained by setting the counts in the search region equal to the interpolated background plus the
Table 7

| $|q|_{\text{max}}$ range | 0–1.0 | 1.0–1.5 | 1.5–2.5 |
|-----------------------|-------|-------|-------|
| SES $= (\epsilon_{\text{obs}})^{-1}$ ($10^{-8}$) | 0.71 | 1.6 | 1.4 |
| $\epsilon = (f_{\text{BS}}/f_{\text{FS}})\times BR(B^0 \rightarrow J/\psi K^{\pm} \rightarrow \mu^+\mu^- K^{\mp})$ ($10^4$) | 4.45 ± 0.38 | 1.40 ± 0.15 | 1.58 ± 0.26 |
| $\epsilon_{ij} = N^{B^0 \rightarrow J/\psi K^{\pm}}/R_{ij}^{B^0}$ ($10^4$) | 3.14 ± 0.17 | 1.40 ± 0.15 | 1.58 ± 0.26 |
| bkg. scaling factor $K_{\text{bg}}^{i\text{th}}$ | 1.29 | 1.14 | 0.88 |
| sideband count $N^{\text{th}}_i$ (even numbered events) | 5 | 0 | 2 |
| expected resonant bkg. $N^{B^0 \rightarrow J/\psi K^{\pm}}_i$ | 0.10 | 0.06 | 0.08 |
| search region count $N^{\text{obs}}_i$ | 2 | 1 | 0 |

Fig. 8. Observed CLs (circles) as a function of $\text{BR}(B^0 \rightarrow \mu^+\mu^-)$. The 95% CL limit is indicated by the horizontal (red) line. The dark (green) and light (yellow) bands correspond to ±1σ and ±2σ fluctuations on the expectation (dashed line), based on the number of observed events in the signal and sideband regions.

small resonant background, before the unblinding of the signal region. A median expected limit of 2.3 ± 1.3 × 10^−8 at 95% CL was obtained, where the range encloses 68% of the background-only pseudo-experiments.

For comparison the mass-resolution categories were merged and the selection optimization was performed on the merged sample. In this case eight events were found in the sidebands, resulting in a branching fraction limit of 2.9 ± 1.3 × 10^−8 at 95% CL. This test confirms the expectation of a more sensitive analysis when separate mass-resolution categories are used.

The background counts found in odd-numbered events were used to assess the magnitude of the bias that would be caused by using the same sample for selection optimization and the estimation of $N^{\text{th}}_i$. The expected limit obtained using the same sample for optimization and signal extraction is 1.7 × 10^−8, about 30% smaller than the limit presented in this Letter, for which independent samples were used for optimization and for signal extraction. The observed bias is consistent with simulation-based assessments of this effect.

Fig. 8 shows the behaviour of the observed CLs for different tested values of the $B^0 \rightarrow \mu^+\mu^-$ branching fraction, computed with 300,000 toy MC simulations per point. The observed limit is < 2.2(1.9) × 10^−8 at 95% (90%) CL. The $p$-values for the background-only hypothesis and for background plus SM prediction are 44% and 35%, respectively.

Despite the difference between the total numbers of observed and interpolated background events (equal to 3 and 6.5, respectively), the interplay of the event counts observed in the three mass-resolution categories produced an observed CLs limit close to the expected value.

7. Conclusions

A limit on the branching fraction $\text{BR}(B^0 \rightarrow \mu^+\mu^-)$ is set using 2.4 fb^−1 of integrated luminosity collected in 2011 by the ATLAS detector. The process $B^0 \rightarrow J/\psi K^{\pm}$, with $J/\psi \rightarrow \mu^+\mu^-$, is used as a reference channel for the normalization of integrated luminosity, acceptance and efficiency. The final selection is based on a multivariate analysis performed on three categories of events determined according to their mass resolution, yielding a limit of $\text{BR}(B^0 \rightarrow \mu^+\mu^-) < 2.2(1.9) \times 10^{-8}$ at 95% (90%) CL.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF, DNNRC and Lundbeck Foundation, Denmark; EPLANET and ERC, European Union; IN2P3-CNRS, CEA-DSM/IRFU, France; GNAS, Georgia; BMBF, DFG, MPG and AvH Foundation, Germany; GSRT, Greece; ISF, MINERVA, GIF, DIP and Bennoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; RCN, Norway; NSWN, Poland; GRICES and FCT, Portugal; MEYS (MECTS), Czech Republic; MES of Russia and ROSATOM, Russian Federation; JINR; MSTD, Serbia; MSSR, Slovakia; ARRS and MVZT, Slovenia; DST/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SER, SNSF and Cantons of Bern and Geneva, Switzerland; NSC, Taiwan; TAEK, Turkey; STFC, the Royal Society and Leverhulme Trust, United Kingdom; DOE and NSF, United States of America.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.

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.
ATLAS Collaboration
