Study of the rare decays of $B_s^0$ and $B^0$ mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

Abstract: A study of the decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ has been performed using 26.3 fb$^{-1}$ of 13 TeV LHC proton-proton collision data collected with the ATLAS detector in 2015 and 2016. Since the detector resolution in $\mu^+\mu^-$ invariant mass is comparable to the $B_s^0$-$B^0$ mass difference, a single fit determines the signal yields for both decay modes. This results in a measurement of the branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.2^{+1.3}_{-1.1}) \times 10^{-9}$ and an upper limit $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 4.3 \times 10^{-10}$ at 95% confidence level. The result is combined with the Run 1 ATLAS result, yielding $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (2.8^{+0.8}_{-0.7}) \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 2.1 \times 10^{-10}$ at 95% confidence level. The combined result is consistent with the Standard Model prediction within 2.4 standard deviations in the $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$-$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ plane.

Keywords: Hadron-Hadron scattering (experiments)

ArXiv ePrint: 1812.03017
Contents

1 Introduction ........................................... 1

2 ATLAS detector, data and simulation samples .......... 3

3 Data selection ......................................... 5

4 Background composition ................................ 6

5 Hadron misidentification ................................ 8

6 Continuum background reduction ......................... 9

7 Data-simulation comparisons ............................ 11

8 $B^+ \to J/\psi K^+$ yield extraction ...................... 13

9 Evaluation of the $B^+ \to J/\psi K^+$ to $B^0(s) \to \mu^+\mu^-$ efficiency ratio 15

10 Extraction of the signal yield .......................... 17

  10.1 Signal and background model ....................... 17

  10.2 Relative signal efficiency between BDT bins ....... 19

  10.3 Systematic uncertainties in the fit ................ 19

  10.4 Results of the signal yield extraction ............ 22

11 Branching fraction extraction .......................... 22

12 Combination with the run 1 result ...................... 24

13 Conclusions ........................................... 26

The ATLAS collaboration .................................... 30

1 Introduction

Flavour-changing neutral-current processes are highly suppressed in the Standard Model (SM). The branching fractions of the decays $B^0(s) \to \mu^+\mu^-$ are, in addition, helicity suppressed in the SM, and are predicted to be $\mathcal{B}(B^0(s) \to \mu^+\mu^-) = (3.65 \pm 0.23) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.06 \pm 0.09) \times 10^{-10}$ [1]. The smallness and precision of these predicted branching fractions provide a favourable environment for observing contributions from new physics: significant deviations from SM predictions could arise in models involving non-SM heavy particles, such as those predicted in the Minimal Supersymmetric
Standard Model [2–6] and in extensions such as Minimal Flavour Violation [7, 8], Two- Higgs-Doublet Models [6], and others [9, 10]. The CMS and LHCb collaborations reported the observation of $B_s^0 \rightarrow \mu^+ \mu^-$ [11, 12] and evidence of $B^0 \rightarrow \mu^+ \mu^-$ with combined values $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8 \pm 0.7 \pm 0.5) \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (3.9_{-1.2}^{+1.6}) \times 10^{-10}$ [13]. The LHCb Collaboration updated its Run 1 result with part of the data collected in Run 2, measuring $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0 \pm 0.6_{-0.2}^{+0.3}) \times 10^{-9}$ (where the first uncertainty is statistical and the second systematic) and determining $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.4 \times 10^{-10}$ at 95% confidence level (CL) [14]. ATLAS has measured, with the Run 1 dataset [15], $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (0.9_{-0.4}^{+1.1}) \times 10^{-9}$, while setting an upper limit on $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ of 4.2 $\times 10^{-10}$ at 95% CL.

This paper reports the result of a search for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays performed using $pp$ collision data corresponding to an effective integrated luminosity of 26.3 fb$^{-1}$, collected at 13 TeV centre-of-mass energy during the first two years of the LHC Run 2 data-taking period using the ATLAS detector. The analysis strategy follows the approach employed in the previous ATLAS measurement [15], but uses data collected with one rather than three separate sets of trigger thresholds, applies standard ATLAS muon selection criteria (section 3) rather than a dedicated muon multivariate discriminant, and employs an improved statistical treatment of the result (section 10).

The notation used throughout the paper refers to the combination of processes and their charge-conjugates, unless otherwise specified. The $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ branching fractions are measured relative to the reference decay mode $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^+$ which is abundant and has a well-measured branching fraction $\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$. The $B^0 \rightarrow \mu^+ \mu^-$ ($B_s^0 \rightarrow \mu^+ \mu^-$) branching fraction can be extracted as:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \frac{N_d(s)}{\varepsilon_{\mu^+ \mu^-}} \times \left[ \mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) \right] \frac{\varepsilon_{J/\psi K^+}}{N_{J/\psi K^+}} \times \frac{f_u}{f_d(s)}$$

$$= N_d(s) \frac{\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)}{D_{\text{ref}}} \times \frac{f_u}{f_d(s)},$$

(1.1)

where $N_d (N_s)$ is the $B^0 \rightarrow \mu^+ \mu^-$ ($B_s^0 \rightarrow \mu^+ \mu^-$) signal yield, $N_{J/\psi K^+}$ is the $B^+ \rightarrow J/\psi K^+$ reference channel yield, $\varepsilon_{\mu^+ \mu^-}$ and $\varepsilon_{J/\psi K^+}$ are the corresponding values of acceptance times efficiency (measured in fiducial regions defined in section 9), and $f_u/f_d (f_s/f_u)$ is the ratio of the hadronisation probabilities of a $b$-quark into $B^+$ and $B^0$ ($B_s^0$). In the quantity $D_{\text{ref}} = N_{J/\psi K^+} \times (\varepsilon_{\mu^+ \mu^-}/\varepsilon_{J/\psi K^+})$, the $\varepsilon$ ratio takes into account relative differences in efficiencies, integrated luminosities and the trigger selections used for the signal and the reference modes. Signal and reference channel events are selected with similar dimuon triggers. One half of the reference channel sample is used to determine the normalisation and the other half is used to tune the kinematic distributions of simulated events.

The event selection uses variables related to the $B$ candidate decay time, thus introducing a dependence of the efficiency on the signal lifetime. The relation between the measured branching fraction and the corresponding value at production is established assuming the decay time distribution predicted in the SM, where the decay occurs mainly through the heavy eigenstate $B_{(s), H}^0$ of the $B_{(s), H}^0$ system. Some models of new physics [16, 17] predict modifications to the decay time distribution of $B_s^0 \rightarrow \mu^+ \mu^-$ and a comparison with
the experimental result requires a correction to the ratio of the time-integrated efficiencies entering $D_{\text{ref}}$.

The ATLAS inner tracking system, muon spectrometer and, for efficient identification of muons, also the calorimeters, are used to reconstruct and select the event candidates. Details of the detector, trigger, data sets, and preliminary selection criteria are discussed in sections 2 and 3. A blind analysis was performed in which data in the dimuon invariant mass range from 5166 to 5526 MeV were removed until the procedures for event selection and the details of signal yield extraction were completely defined. Section 4 introduces the three main categories of background. Section 5 describes the strategy used to reduce the probability of hadron misidentification. The final sample of candidates is selected using a multivariate classifier, designed to enhance the signal relative to the dominant dimuon background component, as discussed in section 6. Checks on the distributions of the variables used in the multivariate classifier are summarised in section 7. They are based on the comparison of data and simulation for dimuon events, for $B^+ \to J/\psi K^+$ candidates and for events selected as $B_s^0 \to J/\psi \phi \to \mu^+ \mu^- K^+ K^-$, which provide an additional validation of the procedures used in the analysis. Section 8 details the fit procedure used to extract the yield of $B^+ \to J/\psi K^+$ events. The determination of the ratio of efficiencies in the signal and the reference channels is presented in section 9. Section 10 describes the extraction of the signal yield, obtained with an unbinned maximum-likelihood fit performed on the dimuon invariant mass distribution. In this fit, events are separated into classifier intervals to maximise the fit sensitivity. The results for the branching fractions $B(B_s^0 \to \mu^+ \mu^-)$ and $B(B^0 \to \mu^+ \mu^-)$ are reported in section 11 and combined with the full Run 1 results in section 12.

2 ATLAS detector, data and simulation samples

The ATLAS detector\(^1\) consists of three main components: an inner detector (ID) tracking system immersed in a 2 T axial magnetic field, surrounded by electromagnetic and hadronic calorimeters and by the muon spectrometer (MS). A full description can be found in ref. [18], complemented by ref. [19] for details about the new innermost silicon pixel layer that was installed for Run 2.

This analysis is based on the Run 2 data recorded in 2015 and 2016 from $pp$ collisions at the LHC at $\sqrt{s} = 13$ TeV. Data used in the analysis were recorded during stable LHC beam periods. Data quality requirements were imposed, notably on the performance of the MS, ID and calorimeter systems. The total integrated luminosity collected by ATLAS in this period is $36.2\,\text{fb}^{-1}$ with an uncertainty of 2.1%. These values are determined using a methodology similar to that detailed in ref. [20], based on calibration of the luminosity scale using $x$-$y$ beam-separation scans, and use the LUCID-2 detector [21] for the baseline luminosity measurement. The total effective integrated luminosity used in this analysis —

\(^1\)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, $r$ being the distance from the origin and $\phi$ 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.
accounting for trigger prescales — amounts to 26.3 fb$^{-1}$ for the signal and 15.1 fb$^{-1}$ for the reference channel.

Samples of simulated Monte Carlo (MC) events are used for training and validation of the multivariate analyses, for the determination of the efficiency ratios, and for developing the procedure used to determine the signal. Exclusive MC samples were produced for the signal channels $B_0^s \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$, the reference channel $B^+ \rightarrow J/\psi K^+$ ($J/\psi \rightarrow \mu^+ \mu^-$), and the control channel $B_s^0 \rightarrow J/\psi \phi$ ($J/\psi \rightarrow \mu^+ \mu^-$, $\phi \rightarrow K^+ K^-$). In addition, background studies employ MC samples of inclusive $b \bar{b} \rightarrow \mu^+ \mu^- X$ decays, exclusive samples of $B_0^s \rightarrow K^- \mu^+ \nu$, $B^0 \rightarrow \pi^- \mu^+ \nu$, $\Lambda_b \rightarrow p \mu^- \bar{\nu}$, $B^0(s) \rightarrow hh'$ decays with $h^0$ being a charged pion or kaon, and inclusive decays $B \rightarrow J/\psi X$ as well as the exclusive $B^+ \rightarrow J/\psi \pi^+$ decay.

Most of the dimuon candidates in the data sample originate from the decays of hadrons produced in the hadronisation of $b \bar{b}$ pairs. The inclusive $b \bar{b} \rightarrow \mu^+ \mu^- X$ MC sample used to describe this background requires the presence of two muons in the final state, with both muons originating from the $b \bar{b}$ decay chain. The size of this sample is equivalent to roughly three times the integrated luminosity of the data.

The MC samples were generated with Pythia 8 [22]. The ATLAS detector and its response were simulated using Geant4 [23, 24]. Additional $pp$ interactions in the same and nearby bunch crossings (pile-up) are included in the simulation. Muon reconstruction and triggering efficiencies are corrected in the simulated samples using data-driven scale factors. The scale factors for the trigger efficiencies are obtained by comparing data and simulation efficiencies determined with a tag-and-probe method on inclusive prompt and non-prompt $J/\psi$ candidates. This procedure yields scale factors as a function of the muon transverse momentum and pseudorapidity, which are applied throughout the analysis [25]. Reconstruction and selection efficiencies are obtained from simulation and similarly corrected according to data-driven comparisons. In addition to these efficiency corrections, simulated events are reweighted to reproduce the pile-up multiplicity observed in data, and according to the equivalent integrated luminosity associated with each trigger selection.

Using the iterative reweighting method described in ref. [26], the simulated samples of the exclusive decays considered are adjusted with two-dimensional data-driven weights (DDW) to correct for the differences between simulation and data observed in the $B$ meson transverse momentum and pseudorapidity distributions. DDW obtained from $B^+ \rightarrow J/\psi K^+$ decays are used to correct the simulation samples in the signal and reference channels. DDW obtained from the $B_s^0 \rightarrow J/\psi \phi$ control channel are found to agree with those from $B^+ \rightarrow J/\psi K^+$, showing that the same corrections are applicable to $B_s^0$ and $B^0$ decays.

Residual differences between data and simulation studied in the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ signals are treated as sources of systematic uncertainty in the evaluation of the signal efficiency, as discussed in section 9. The only exception to this treatment is the $B$ meson isolation ($I_{0.7}$ in section 6 and table 1), where residual differences are used to reweight the signal MC events and the corresponding uncertainties are propagated to account for residual systematic uncertainty effects.

Similarly to the exclusive decays, the kinematic distributions of the inclusive $b \bar{b} \rightarrow \mu^+ \mu^- X$ MC sample are reweighted with corrections obtained from the dimuon invariant mass sidebands in data.
3 Data selection

For data collected during the LHC Run 2, the ATLAS detector uses a two-level trigger system, consisting of a hardware-based first-level trigger and a software-based high-level trigger. A first-level dimuon trigger [27] selects events requiring that one muon has $p_T > 4 \text{ GeV}$ and the other has $p_T > 6 \text{ GeV}$. A full track reconstruction of the muon candidates is performed by the high-level trigger, where an additional loose selection is imposed on the dimuon invariant mass $m_{\mu\mu}$, accepting candidates in the range 4 GeV to 8.5 GeV. Due to the increased pile-up in 2016 data, an additional selection was added at this trigger stage, requiring the vector from the primary vertex to the dimuon vertex to have a positive component ($L_{xy}$) along the dimuon’s transverse momentum direction. The effect of this selection is accounted for in the analysis but has no consequence since stricter requirements are applied in the full event selection (see section 6).

The signal channel, the reference channel $B^+ \rightarrow J/\psi K^+$ and the control channel $B_s^0 \rightarrow J/\psi \phi$ were selected with trigger prescale factors that vary during the data-taking period. In the 36.2 fb$^{-1}$ of data analysed, the prescaling of the trigger approximately averages to a reduction by a factor 1.4, giving an effective integrated luminosity for the signal sample of 26.3 fb$^{-1}$, while for the reference and control channels 15.1 fb$^{-1}$ were collected due to an effective prescale of 2.4. These effects are taken into account in the extraction of the signal branching fraction, through the $\varepsilon$ factors in eq. (1.1).

Using information from the full offline reconstruction, a preliminary selection is performed on candidates for $B^0_{(s)} \rightarrow \mu^+\mu^-$, $B^+ \rightarrow J/\psi K^+ \rightarrow \mu^+\mu^-K^+$ and $B^0_s \rightarrow J/\psi \phi \rightarrow \mu^+\mu^-K^+K^-$ decays. In the ID system, muon candidates are required to have at least one measured hit in the pixel detector and two measured hits in the semiconductor tracker. They are also required to be reconstructed in the MS, and to have $\Delta R > 2.5$. The offline muon pair must pass the $p_T > 4 \text{ GeV}$ and $p_T > 6 \text{ GeV}$ requirements imposed by the trigger. Furthermore, the muon candidates are required to fulfil tight muon quality criteria [28]; this requirement is relaxed to loose for the hadron misidentification studies in section 5. Kaon candidates must satisfy similar requirements in the ID, except for a looser requirement of $p_T > 1 \text{ GeV}$.

The computed $B$ meson properties are based on a decay vertex fitted to two, three or four tracks, depending on the decay process to be reconstructed. The $B$ candidates are required to have a $\chi^2$ per degree of freedom below 6 for the fit to the $B$ vertex, and below 10 for the fit to the $J/\psi \rightarrow \mu^+\mu^-$ vertex. The selections $2915 < m(\mu^+\mu^-) < 3275 \text{ MeV}$ and $1005 < m(K^+K^-) < 1035 \text{ MeV}$ are applied to the $J/\psi \rightarrow \mu^+\mu^-$ and the $\phi \rightarrow K^+K^-$ vertices, respectively. In the fits to the $B^+ \rightarrow J/\psi K^+$ and $B^0_s \rightarrow J/\psi \phi$ channels, the reconstructed dimuon mass is constrained to the world average $J/\psi$ mass [29].

Reconstructed $B$ candidates are retained if they satisfy $p_T^B > 8.0 \text{ GeV}$ and $|\eta^B| < 2.5$. The invariant mass of each $B$ candidate is calculated using muon trajectories measured by combining the information from the ID and MS to improve upon the mass resolution obtained from ID information only [30].

The invariant mass range considered for the $B^0_{(s)} \rightarrow \mu^+\mu^-$ decay starts at 4766 MeV and is 1200 MeV wide. Within this range a 360-MeV-wide signal region is defined, starting
at 5166 MeV. The remainder of the range defines the upper and lower mass sidebands of the analysis.

For the reference and control channels, the mass range considered is 4930–5630 (5050–5650) MeV for \( B^+ \rightarrow J/\psi K^+ \) (\( B^0_s \rightarrow J/\psi \phi \)), where 5180–5380 (5297–5437) MeV is the peak region and higher and lower mass ranges comprise the mass sidebands used for background subtraction.

The coordinates of primary vertices (PV) are obtained from charged-particle tracks not used in the decay vertices, and that are constrained to the luminous region of the colliding beams in the transverse plane. The matching of a \( B \) candidate to a PV is made by extrapolating the candidate trajectory to the point of closest approach to the beam axis, and choosing the PV with the smallest distance along \( z \). Simulation shows that this method matches the correct vertex with a probability above 99% for all relevant pile-up conditions.

To reduce the large background in the \( B^0_s \rightarrow \mu^+\mu^- \) channel before applying the final selection based on multivariate classifiers, a loose collinearity requirement is applied between the momentum of the \( B \) candidate (\( \vec{p}^B \)) and the vector from the PV to the decay vertex (\( \vec{x} \)). The absolute value of the azimuthal angle \( \alpha_{2D} \) between these two vectors is required to be smaller than 1.0 radians. The combination \( R_{\text{light}} = \sqrt{\alpha_{2D}^2 + (\Delta \eta)^2} \), where \( \Delta \eta \) is the difference in pseudorapidity, is required to satisfy \( R_{\text{light}} < 1.5 \).

After the preliminary selection, approximately \( 3.5 \times 10^6 \) candidates are found in the \( B^0_{(s)} \rightarrow \mu^+\mu^- \) fit region, with about \( 1.0 \times 10^6 \) falling in the blinded range [5166, 5526] MeV.

### 4 Background composition

The background to the \( B^0_{(s)} \rightarrow \mu^+\mu^- \) signal originates from three main sources:

- **Continuum** background, the dominant combinatorial component, which consists of muons originating from uncorrelated hadron decays and is characterised by a weak dependence on the dimuon invariant mass;
- **Partially reconstructed** decays, where one or more of the final-state particles (\( X \)) in a \( b \)-hadron decay is not reconstructed, causing these candidates to accumulate in the low dimuon invariant mass sideband (this background includes a significant contribution from semileptonic decays where one of the muons is a misidentified hadron, discussed below);
- **Peaking** background, due to \( B^0_{(s)} \rightarrow hh' \) decays, with both hadrons misidentified as muons.

The continuum background consists mainly of muons produced independently in the fragmentation and decay chains of a \( b \)-quark and a \( b \)-quark. It is studied in the signal mass sidebands, and it is found to be well described by the inclusive \( b \bar{b} \rightarrow \mu^+\mu^- X \) MC sample.

The partially reconstructed decays consist of several topologies: (a) **same-side** combinatorial background from decay cascades (\( b \rightarrow c\mu^-\bar{\nu} \rightarrow s(d)\mu^+\mu^-\nu\bar{\nu} \)); (b) **same-vertex** background from \( B \) decays containing a muon pair (e.g. \( B^0 \rightarrow K^{*0}\mu^+\mu^- \) or \( B \rightarrow J/\psi\mu^+X \rightarrow \mu^+\mu^-\mu^+X \)); (c) **\( B_c^+ \)** decays (e.g. \( B_c^+ \rightarrow J/\psi\mu^+\nu \rightarrow \mu^+\mu^-\mu^+\nu \)); (d) semileptonic \( b \)-hadron decays where a final-state hadron is misidentified as a muon. The
Figure 1. (a) Dimuon invariant mass distribution for the partially reconstructed background (as categorised in section 4), from simulation, before the final selection against continuum is applied but after all other requirements. The different components are shown as stacked histograms, normalised according to world-averaged measured branching fractions. The SM expectations for the $B^0_s \rightarrow \mu^+\mu^-$ signals are also shown for comparison. Continuum background is not included here. (b) Invariant mass distribution of the $B^0 \rightarrow hh'$ peaking background components after the complete signal selection is applied. The $B^0_s \rightarrow \pi^+\pi^-$ and $B^{0*} \rightarrow K^+K^-$ contributions are negligible on this scale. In both plots the vertical dashed lines indicate the blinded analysis region. Distributions are normalised to the expected yield for the integrated luminosity of 26.3 fb$^{-1}$.

remainder of this paper implicitly excludes categories (c) and (d) when referring to partially reconstructed or $b \rightarrow \mu^+\mu^-X$ decays, since these categories are treated separately.

The $b\bar{b} \rightarrow \mu^+\mu^-X$ MC sample is used to investigate the background composition after the analysis selection. All backgrounds in this sample have a dimuon invariant mass distribution mainly below the mass range considered in this analysis, with a high-mass tail extending through the signal region. The simulation does not contemplate sources other than muons from $b\bar{b}$ decays: $c\bar{c}$ and prompt contributions are not included$^2$. All possible origins of two muons in the $b\bar{b}$ decay tree are, however, analysed, after classification into the mutually exclusive continuum and partially reconstructed categories described above. This sample is used only to identify suitable functional models for the corresponding background components, and as a benchmark for these models. No shape or normalisation constraints are derived from this simulation. This makes the analysis largely insensitive to mismatches between background simulation and data.

The semileptonic decays with final-state hadrons misidentified as muons consist mainly of three-body charmless decays $B^0 \rightarrow \pi^-\mu^+\nu$, $B^0_s \rightarrow K^-\mu^+\nu$ and $\Lambda_b \rightarrow p\mu^-\bar{\nu}$ in which the tail of the invariant mass distribution extends into the signal region. Due to branching fractions of the order of $10^{-6}$, this background is not large, and is further reduced by

$^2$These sources are suppressed by the final analysis selections introduced in section 6. Potential residual contributions are found to be consistent with the continuum background models used in the final fit (section 10.1), where systematic uncertainties for background model inconsistencies between data and MC are taken into account (section 10.3).
the muon identification requirements, discussed in section 5. The MC invariant mass distributions of these partially reconstructed decay topologies are shown together with the SM signal predictions in figure 1(a) after applying the preliminary selection criteria described in section 3.

Finally, the peaking background is due to $B^0_{(s)} \rightarrow hh'$ decays containing two hadrons misidentified as muons. The distributions in figure 1(b), obtained from simulation, show that these decays populate the signal region. This component is further discussed in section 5.

## 5 Hadron misidentification

In the preliminary selection, muon candidates are formed from the combination of tracks reconstructed independently in the ID and MS. The performance of the muon reconstruction in ATLAS is presented in ref. [28]. Additional studies were performed to evaluate the amount of background related to hadrons erroneously identified as muons.

Detailed simulation studies were performed for the $B^0_{(s)} \rightarrow hh'$ channel with a full GEANT4-based simulation [23] of all systems of the ATLAS detector. The vast majority of background events from particle misidentification are due to decays in flight of kaons and pions, in which the muon receives most of the energy of the parent meson. Hence this background is generally related to true muons measured in the MS, but not produced promptly in the decay of a $B$ meson.

The muon candidate is required to pass \textit{tight} muon requirements in the preliminary selection, which are based on the profile of energy deposits in the calorimeters as well as on tighter ID-MS matching criteria than those used for the \textit{loose} requirements. Two-body $B$ decays in control regions show that \textit{tight} selections have, relative to the \textit{loose} counterpart, an average hadron misidentification probability reduced by a factor 0.39 with a muon reconstruction efficiency of 90%. The resulting final value of the misidentification probability is 0.08% for kaons and 0.1% for pions. Efficiencies and fake rates are relative to the analysis preselections, including tracking but excluding any muon requirement.

The background due to $B^0_{(s)} \rightarrow hh'$, with double misidentification of $hh'$ as $\mu^+\mu^-$, has a reconstructed invariant mass distribution that peaks at 5240 MeV, close to the $B^0$ mass, and is effectively indistinguishable from the $B^0$ signal (see figure 1(b)). The expected number of peaking-background events can be estimated in a way analogous to that for the signal, from the number of observed $B^+ \rightarrow J/\psi K^+$ events using eq. (1.1), after taking into account the expected differences from muon identification variables and trigger selections. World average [29] values for the branching fractions of $B^0$ and $B^0_{s}$ into $K\pi$, $KK$ and $\pi\pi$ are used, together with the hadron misidentification probabilities obtained from simulation. This results in $2.7 \pm 1.3$ total expected peaking-background events, after the reference multivariate selection.$^3$

$^3$This selection, corresponding to 54% signal efficiency, was also applied to derive all other quantities quoted in this section and includes a selection against the $\mu^+\mu^-$ continuum background based on the BDT discussed in section 6.
When selecting loose muons and inverting the additional requirements imposed in the tight muon selection, the number of events containing real muons is substantially reduced, while the number of peaking-background events is approximately two times larger than in the sample obtained with the nominal selection. A fit to data for this background-enhanced sample returns $6.8 \pm 3.7$ events, which translates into a peaking-background yield in the signal region of $2.9 \pm 2.0$ events when taking into account the relative rejection of the muon quality selections. The predicted yield is in good agreement with the simulation.

Besides the peaking background, the tight muon selection also reduces the semileptonic contributions with a single misidentified hadron. Simulation yields $30 \pm 3$ events expected from $B^0 \to \pi^- \mu^+ \nu$ and $B^0_{s} \to K^- \mu^+ \nu$ in the final sample, with a distribution kinematically constrained to be mostly below the signal region. The $\Lambda_b \to \mu^- \pi^+$ contribution is negligible due to the smaller production cross section and the low rate at which protons fake muons.

## 6 Continuum background reduction

A multivariate analysis, implemented as a boosted decision tree (BDT), is employed to enhance the signal relative to the continuum background. This BDT is based on the 15 variables described in table 1. The discriminating variables can be classified into three groups: (a) $B$ meson variables, related to the reconstruction of the decay vertex and to the collinearity between $\vec{p}^B$ and the flight vector between the production and decay vertices $\Delta \vec{x}$; (b) variables describing the muons that form the $B$ meson candidate; and (c) variables related to the rest of the event. The selection of the variables aims to maximise the discrimination power of the classifier without introducing significant dependence on the invariant mass of the muon pair.

The same discriminating variables were used in the previous analysis based on the full Run 1 dataset [15]. The removal of individual variables was explored to simplify the BDT input, however, this results inevitably in a significant reduction of the BDT separation power. To minimise the dependence of the classifier on the effects of pile-up, the additional tracks considered to compute the variables $I_{0,7}$, DOCA$_{\text{strk}}$ and $N_{\text{close}}$ are required to be compatible with the primary vertex matched to the dimuon candidate.

The correlations among the discriminating variables were studied in the MC samples for signal and continuum background discussed in section 2, and in data from the sidebands of the $\mu^+ \mu^-$ invariant mass distribution. There are significant linear correlations among the variables $\chi^2_{PV, \Delta V, x}$, $L_{xy}$, $|d_0|_{\text{max-sig}}$, $|d_0|_{\text{min-sig}}$ and $\chi^2_{\mu, \Sigma PV}$. The variables IP$^3_{B}$, DOCA$_{\mu\mu}$ and $I_{0,7}$ have negligible correlation with any of the others used in the classifier.

The simulated signal sample and the data from the dimuon invariant mass sideband regions are used for training and testing the classifier. As discussed in section 2, simulated signal samples are corrected for muon reconstruction efficiency differences between simulation and data, and reweighted according to the distributions of $p_T$ and $|\eta|$ of the dimuon and of the pile-up observed in data. The BDT training is done using the TMVA toolkit [31].

Sideband data are used for the BDT training and optimisation. The sample is subdivided into three randomly selected separate and equally populated subsamples used in turn to train and validate the selection efficiency of three independent BDTs. The resulting
Variable | Description
---|---
P_B^T | Magnitude of the \( B \) candidate transverse momentum \( \vec{p}_T^B \).
\( \chi^2 \)_{PV,DV xy} | Compatibility of the separation \( \Delta \vec{x} \) between production (i.e. associated PV) and decay (DV) vertices in the transverse projection: \( \Delta \vec{x}_T \cdot \sum_{\Delta \vec{x}_T}^{-1} \cdot \Delta \vec{x}_T \), where \( \sum_{\Delta \vec{x}_T}^{-1} \) is the covariance matrix.
\( \Delta R_{right} \) | Three-dimensional angular distance between \( \vec{p}_B \) and \( \Delta \vec{x} \): \( \sqrt{\alpha_{2D}^2 + (\Delta \eta)^2} \)
| | Absolute value of the angle in the transverse plane between \( \vec{p}_T^B \) and \( \Delta \vec{x}_T \).
| | Projection of \( \Delta \vec{x}_T \) along the direction of \( \vec{p}_T^B \): \( \langle \Delta \vec{x}_T \cdot \vec{p}_T^B \rangle / |\vec{p}_T^B| \).
| | Three-dimensional impact parameter of the \( B \) candidate to the associated PV.
DOCA_{\mu\mu} | Distance of closest approach (DOCA) of the two tracks forming the \( B \) candidate (three-dimensional).
\( \Delta \phi_{\mu\mu} \) | Azimuthal angle between the momenta of the two tracks forming the \( B \) candidate.
| | Significance of the larger absolute value of the impact parameters to the PV of the tracks forming the \( B \) candidate, in the transverse plane.
| | Significance of the smaller absolute value of the impact parameters to the PV of the tracks forming the \( B \) candidate, in the transverse plane.
| | The smaller of the projected values of the muon momenta along \( \vec{p}_T^B \).
I_{0.7} | Isolation variable defined as ratio of \( |\vec{p}_T^B| \) to the sum of \( |\vec{p}_T^B| \) and the transverse momenta of all additional tracks contained within a cone of size \( \Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2} = 0.7 \) around the \( B \) direction. Only tracks matched to the same PV as the \( B \) candidate are included in the sum.
DOCA_{xtrk} | DOCA of the closest additional track to the decay vertex of the \( B \) candidate. Only tracks matched to the same PV as the \( B \) candidate are considered.
\( \chi^2_{xtrk,DV} \) | Number of additional tracks compatible with the decay vertex (DV) of the \( B \) candidate with \( \ln(\chi^2_{xtrk,DV}) < 1 \). Only tracks matched to the same PV as the \( B \) candidate are considered.
\( \chi^2_{\mu,\pi PV} \) | Minimum \( \chi^2 \) for the compatibility of a muon in the \( B \) candidate with any PV reconstructed in the event.

Table 1. Description of the 15 input variables used in a BDT classifier to discriminate between signal and continuum background. When the BDT classifier is applied to \( B^+ \to J/\psi K^+ \) and \( B^0_s \to J/\psi \phi \) candidates, the variables related to the decay products of the \( B \) mesons refer only to the muons from the decay of the \( J/\psi \). Horizontal lines separate the classifications into groups (a), (b) and (c) respectively, as described in the text. For category (c), additional tracks are required to have \( p_T > 500 \) MeV.
BDTs are found to produce results that are statistically compatible, and are combined in one single classifier in such a way that each BDT is applied only to the part of the data sample not involved in the BDT training.

Figure 2 shows the distribution of the BDT output variable for simulated signal and backgrounds, separately for continuum background and partially reconstructed events. Also shown is the BDT distribution for dimuon candidates from the sidebands of the invariant mass distribution in data. The BDT output was found to not have any significant correlation with the dimuon invariant mass. The final selection requires a BDT output value larger than 0.1439, corresponding to signal and continuum background efficiencies of 72% and 0.3% respectively. The analysis uses all candidates after this selection; however, accepted events with BDT values close to the selection threshold are effectively only constraining the background models.\footnote{The $B^{0} \rightarrow \mu^{+} \mu^{-}$ signal fit was found to be insensitive to the signal for candidates with BDT output value smaller than 0.2455 (corresponding to 54% and 0.03% efficiencies for signal and background respectively).} For this reason, signal and reference channel yields and efficiencies are measured relative to the signal reference selection discussed in section 9, while the events in the final selection with lower BDT values are used to improve the background modelling.

7 Data-simulation comparisons

Despite the lack of correlation between the BDT variable and the candidates invariant mass, the change in relative contribution of different background components as a function
Figure 3. Data and continuum MC distributions of the (a) $|\alpha_{2D}|$ and (b) $\ln\left(\chi^2_{\mu,xPV}\right)$ variables (defined in table 1). The points correspond to the sideband data, while the continuous-line histogram corresponds to the continuum MC distribution, normalised to the number of data events. The filled-area histogram shows the signal MC distribution for comparison. The bottom insets report the data/MC ratio, zoomed-in in order to highlight discrepancies in the region that is most relevant for the analysis.

of the BDT output can produce a variation of the background shape in separate BDT bins. The shape parameters and normalisations of the backgrounds are for this reason determined purely from data: the $b\bar{b} \rightarrow \mu^+\mu^- X$ MC simulation is not used for BDT training, computation of efficiencies or normalisation purposes. The only role of the $b\bar{b} \rightarrow \mu^+\mu^- X$ MC is the validation of the functional forms employed to parameterise the backgrounds on data, as discussed in section 10.1. The dependence of the background parameterisation on the BDT output variable was investigated in data and shown to be consistent with simulation, even for lower BDT values where prompt backgrounds contribute significantly. This observation is supported by the fair agreement of the BDT input variables distribution between the simulated $b\bar{b} \rightarrow \mu^+\mu^- X$ background and what observed on data sidebands.

Figure 3 compares the distributions of two of the most discriminating variables in the continuum background MC sample with data in the dimuon mass sidebands. Agreement with the sideband data is fair.

The distributions of the discriminating variables are also used to compare simulation and data in the $B^+ \rightarrow J/\psi K^+$ and $B^0_s \rightarrow J/\psi \phi$ samples. To perform these comparisons, for each variable the contribution of the background is subtracted from the $B^+ \rightarrow J/\psi K^+$ ($B^0_s \rightarrow J/\psi \phi$) signal. For this purpose, a maximum-likelihood fit is performed to the invariant mass distribution, separately in bins of rapidity and transverse momentum. The fit model used is simpler than the one employed for the extraction of the $B^+$ signal for normalisation as described in section 8, but is sufficient for the purpose discussed here.

Figure 4 shows examples of the distributions of the discriminating variables obtained from data and simulation for the reference samples. Observed differences are used to estimate systematic uncertainties on the efficiency ratio $R_e = \varepsilon(B^+ \rightarrow J/\psi K^+)/\varepsilon(B^0_{(s)} \rightarrow \mu^+\mu^-)$ with the procedure described in section 9. The discrepancy visible for the isolation
Figure 4. Data and MC distributions in $B^+ \to J/\psi K^+$ events for the discriminating variables: (a) $|\alpha_{2D}|$, (b) $\ln(\chi^2_{PV,DV,xy})$ and (c) $I_{0.7}$. The variable $I_{0.7}$ is also shown in (d) for $B^0_s \to J/\psi \phi$ events. The points correspond to the sideband-subtracted data, while the line corresponds to the MC distribution, normalised to the number of data events. The highest bin in (c) and (d) accounts for the events with $I_{0.7} = 1$. The bottom insets report the data/MC ratio, zoomed-in in order to highlight discrepancies in the region that is most relevant for the analysis.

variable $I_{0.7}$ in the $B^+ \to J/\psi K^+$ channel is the most significant among all variables and both reference channels.

8 $B^+ \to J/\psi K^+$ yield extraction

The reference channel yield is extracted with an unbinned extended maximum-likelihood fit to the $J/\psi K^+$ invariant mass distribution. The functional forms used to model both the signal and the backgrounds are obtained from studies of MC samples. All the yields are extracted from the fit to data, while the shape parameters are determined from a simultaneous fit to data and MC samples. Free parameters are introduced for the mass scale and mass resolution to accommodate data-MC differences. The best-fit values indicate a negligibly poorer resolution and a mass shift at the level of 2 MeV.
Figure 5. Result of the fit to the $J/\psi K^+$ invariant mass distribution for all $B^+$ candidates in half of the data events. The various components of the spectrum are described in the text. The inset at the bottom of the plot shows the bin-by-bin pulls for the fit, where the pull is defined as the difference between the data point and the value obtained from the fit function, divided by the error from the fit.

The fit includes four components: $B^+ \rightarrow J/\psi K^+$ decays, Cabibbo-suppressed $B^+ \rightarrow J/\psi \pi^+$ decays in the right tail of the main peak, partially reconstructed $B$ decays (PRD) where one or more of the final-state particles are missing, and the non-resonant background composed mostly of $b\bar{b} \rightarrow J/\psi X$ decays. All components other than the last one have shapes constrained by MC simulation as described below, with the data fit including an additional Gaussian convolution to account for possible data-MC discrepancies in mass scale and resolution. The shape of the $B^+ \rightarrow J/\psi K^+$ mass distribution is parameterised using a Johnson $S_U$ function \cite{32, 33}. The final $B^+ \rightarrow J/\psi K^+$ yield includes the contribution from radiative effects (i.e. where photons are emitted from the $B$ decay products). The $B^+ \rightarrow J/\psi \pi^+$ decays are modelled by the sum of a Johnson $S_U$ function and a Gaussian function, where all parameters except the normalisation are determined from the simulation. The decay modes contributing to the PRD are classified in simulation on the basis of their mass dependence. Each of the three resulting categories contributes to the overall PRD shape with combinations of Fermi-Dirac and exponential functions, contributing differently in the low-mass region. Their shape parameters are determined from simulation. Finally, the non-resonant background is modelled with an exponential function with the shape parameter extracted from the fit. The normalisation of each component is unconstrained in the fit, which is therefore mostly independent of external inputs for the branching fractions. The residual dependence of the PRD model shapes on the relative branching fractions of the contributing decays is considered as a source of systematic uncertainty. The resulting fit, shown in figure 5, yields $334 \pm 351$ $B^+ \rightarrow J/\psi K^+$ decays with a statistical uncertainty of 0.3%. The ratio of yields of $B^+ \rightarrow J/\psi \pi^+$ and $B^+ \rightarrow J/\psi K^+$ is $(3.71 \pm 0.09)\%$ (where the uncertainty reported is statistical only), in agreement with the expectation from the world average \cite{29} of $(3.84 \pm 0.16)\%$. 
Some systematic uncertainties are included by design in the fit. For example, the effect of the limited MC sample size is included by performing a simultaneous fit to data and MC samples. Scaling factors determined in the fit to data account for the differences in mass scale and resolution between data and simulation. Additional systematic uncertainties are evaluated by varying the default fit model described above. They take into account the kinematic differences between data and the MC samples used in the fit, differences in efficiency between $B^+$ and $B^-$ decays and uncertainties in the relative fractions and shapes of PRD and in the shape of the various fit components. The stability of this large sample fit is verified by repeating the fit with different initial parameter values. In each case, the change relative to the default fit is recorded, symmetrised and used as an estimate of the systematic uncertainty. The main contributions to the systematic uncertainty come from the functional models of the background components, the composition of the PRD and the signal charge asymmetry. The total systematic uncertainty in the $B^+$ yield amounts to 4.8%.

9 Evaluation of the $B^+ \to J/\psi K^+$ to $B^{0}_{(s)} \to \mu^+\mu^-$ efficiency ratio

The ratio of efficiencies $R_e = \varepsilon(B^+ \to J/\psi K^+)/\varepsilon(B^{0}_{(s)} \to \mu^+\mu^-)$ enters the $D_{\text{ref}}$ term defined in section 1: $D_{\text{ref}} = N_{J/\psi K^+}/R_e$. Both channels are measured in the fiducial acceptance for the $B$ meson, defined as $p_T^B > 8.0$ GeV and $|\eta_B| < 2.5$. Correspondingly, $\varepsilon(B^+ \to J/\psi K^+)$ and $\varepsilon(B^{0}_{(s)} \to \mu^+\mu^-)$ are measured within the $B$ meson fiducial acceptance and include additional final state particles acceptance as well as trigger, reconstruction and selection efficiencies. The final state particles acceptance is defined by the selection placed on the particles in the final state: $|\eta| < 2.5$ and $p_T^\mu > 6.0$ (4.0) GeV for the leading (trailing) muon $p_T$, $p_T^K > 1.0$ GeV and $|\eta_K| < 2.5$ for kaons. The signal reference BDT selection, defined as $\text{BDT} > 0.2455$, has an efficiency of about 54% (51%) in the signal (reference) channel. The overall efficiency ratio $R_e$ is $0.1176 \pm 0.0009$ (stat.) $\pm 0.0047$ (syst.), with uncertainties determined as described below.

The ratio $R_e$ is computed using the mean lifetime of $B^0_{(s)}$ [29, 34] in the MC generator. The same efficiency ratios apply to the $B^{0}_{(s)} \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays, within the MC statistical uncertainty of 0.8%. The statistical uncertainties in the efficiency ratios come from the finite number of events available for the simulated samples. The systematic uncertainty affecting $R_e$ comes from five sources.

The first contribution is due to the uncertainties in the data-driven weights introduced in section 2, and amounts to 0.8%. This term is assessed by creating alternative datasets using correction factors that are randomly sampled in accord with their nominal values and uncertainties. The RMS value of the distribution of $R_e$ obtained from these datasets is taken as the systematic uncertainty.

A second contribution of 1.0% is related to the muon trigger and reconstruction efficiencies. The effect of the uncertainties in the data-driven efficiencies is evaluated using random sampling, as above.

A 3.2% systematic uncertainty contribution arises from the differences between data and simulation observed in the modelling of the discriminating variables used in the BDT classifier (table 1). For each of the 15 variables, the MC samples for $B^{0}_{(s)} \to \mu^+\mu^-$ and
$B^+ \to J/\psi K^+$ are reweighted with the ratio of the $B^+ \to J/\psi K^+$ event distributions in sideband-subtracted data and the MC simulation. The isolation variable $I_{0,7}$ is computed using charged-particle tracks only, and differences between $B^+$ and $B^0_s$ are expected and were observed in previous studies [26]. Hence for this variable the reweighting procedure for the $B^0_s \to \mu^+\mu^-$ MC sample is based on $B^0_s \to J/\psi \phi$ data. For all discriminating variables except $I_{0,7}$, the value of the efficiency ratio is modified by less than 2% by the reweighting procedure and each variation is taken as an independent contribution to the systematic uncertainty in the efficiency ratio. For $I_{0,7}$ the reweighting procedure changes the efficiency ratio by about 6%. Because of the significant mis-modelling, the MC samples obtained after reweighting on the distribution of $I_{0,7}$ are taken as a reference, thus correcting the central value of the efficiency ratio. The 1% uncertainty in the $I_{0,7}$ correction is added to the sum in quadrature of the uncertainties assigned to the other discriminating variables. The total uncertainty in the modelling of the discriminating variables is the dominant contribution to the systematic uncertainty in $R_e$.

A fourth source of systematic uncertainty arises from differences between the $B^0_s \to \mu^+\mu^-$ and the $B^+ \to J/\psi K^+$ channel related to the reconstruction efficiency of the kaon track and of the $B^+$ decay vertex. These uncertainties are mainly due to inaccuracy in the modelling of passive material in the ID. The corresponding systematic uncertainty is estimated by varying the detector model in simulations, which results in changes between 0.4% and 1.5% depending on the $\eta$ range considered. The largest value is used in the full eta range.

Finally, the uncertainty associated with reweighting the simulated events as a function of the pile-up multiplicity distribution contributes 0.6%.

Table 2 summarises these systematic uncertainties.

The efficiency ratio enters in eq. (1.1) with the $D_{\text{ref}}$ term defined in section 1, multiplied by the number of observed $B^\pm$ candidates. The total uncertainty in $D_{\text{ref}}$ is $\pm 6.3\%$.

A correction to the efficiency ratio for $B^0_s \to \mu^+\mu^-$ is needed because of the width difference $\Delta \Gamma_s$ between the $B^0_s$ eigenstates. According to the SM, the decay $B^0_s \to \mu^+\mu^-$ proceeds mainly through the heavy state $B_s^H$ \cite{1,16}, which has width $\Gamma_{s,H} = \Gamma_s - \Delta \Gamma_s/2$, which is 6.6% smaller than the average $\Gamma_s$ \cite{29}. The variation in the value of the $B^0_s \to \mu^+\mu^-$ mean lifetime was tested with simulation and found to change the $B^0_s$ efficiency, and

| Source               | Contribution [%] |
|----------------------|------------------|
| Statistical          | 0.8              |
| Kinematic reweighting (DDW) | 0.8              |
| Muon trigger and reconstruction | 1.0              |
| BDT input variables  | 3.2              |
| Kaon tracking efficiency | 1.5              |
| Pile-up reweighting  | 0.6              |

Table 2. Summary of the uncertainties in $R_e$. 
consequently the $B^0_s$ to $B^+$ efficiency ratio, by $+3.3\%$. This correction is applied to the central value of $D_{\text{ref}}$ used in section 11 for the determination of $\mathcal{B}(B^0_s \to \mu^+\mu^-)$.

Due to the small value of $\Delta \Gamma_d$, no correction needs to be applied to the $B^0 \to \mu^+\mu^-$ decay.

10 Extraction of the signal yield

Dimuon candidates passing the preliminary selection and the selections against hadron misidentification and continuum background are classified according to four intervals (with boundaries at 0.1439, 0.2455, 0.3312, 0.4163 and 1) in the BDT output. Repeating the Run 1 analysis approach, each interval is chosen to give an equal efficiency of $18\%$ for signal MC events, and they are ordered according to increasing signal-to-background ratio.

An unbinned extended maximum-likelihood fit is performed on the dimuon invariant mass distribution simultaneously across the four BDT intervals, each including models for the respective signal and background contributions. The first two bins (covering the lowest BDT values considered) contribute mostly to background modelling; it has been verified with MC pseudo-experiments that they have negligible impact on the signal extraction.

The result of the fit is parameterised with the total yield of $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ events in the three highest intervals of BDT output. Section 10.1 describes the signal and background fit models. The parameters describing the background are allowed to vary freely and are determined by the fit. The normalisations of the individual fit components, including the signals, are completely unconstrained and allowed to take negative values. The ratios of the signal yields in different BDT bins are constrained to equal the ratios of the signal efficiencies in those same bins. The systematic uncertainties due to variations in the relative signal and background efficiencies between BDT intervals, to the signal parameterisation and to the background model are discussed in sections 10.2 and 10.3. Each is modelled in the likelihood as a multiplicative Gaussian distribution whose width is equal to the corresponding systematic uncertainty.

10.1 Signal and background model

The signal and background models are derived from simulations and from data collected in the mass sidebands of the search region.

The invariant mass distribution of the $B^0_{(s)} \to \mu^+\mu^-$ signal is described with two double-Gaussian distributions, centred respectively at the $B^0$ or $B^0_s$ mass. The shape parameters are extracted from simulation, where they are found to be uncorrelated with the BDT output. Systematic uncertainties in the mass scale and resolutions are considered separately. Figure 6 shows the invariant mass distributions for $B^0_s$ and $B^0$, obtained from MC events and normalised to the SM expectations.

---

Footnote: The decay time distribution of $B^0_s \to \mu^+\mu^-$ is predicted to be different from the one of $B_s$ in scenarios of new physics, with the effect related to the observable $A_{\Delta \Gamma}^\mu$ [16, 17]. The maximum possible deviation from the SM prediction of $A_{\Delta \Gamma}^\mu = +1$ is for $A_{\Delta \Gamma}^\mu = -1$, for which the decay time distribution of $B^0_s \to \mu^+\mu^-$ corresponds to the distribution of the $B_{s(L)}$ eigenstate. In the comparison with new-physics predictions, the value of $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ obtained from this analysis should be corrected by $+3.6\%$ or $+7.8\%$ respectively for $A_{\Delta \Gamma}^\mu = 0$ and $-1$. 
Figure 6. Dimuon invariant mass distribution for the $B^0_s$ and $B^0$ signals from simulation. The results of the double-Gaussian fits are overlaid. The two distributions are normalised to the SM prediction for the expected yield with an integrated luminosity of 26.3 fb$^{-1}$.

The background in the signal fit is composed of the types of events described in section 4: (a) the continuum background; (b) the background from partially reconstructed $b \to \mu^+\mu^-X$ events, which is present mainly in the low mass sideband; (c) the peaking background.

The non-peaking contributions have a common mass shape model, with parameters constrained across the BDT bins in the fit as described below, and independent yields across BDT bins and components. Systematic uncertainties arising from model assumptions will be discussed in section 10.3, including effects due to the presence of $B_c^+ \to J/\psi\mu^+\nu$ and semileptonic $B^{0}_{(s)}/A^{0}_{(s)} \to h\mu\nu$ decays.

Both in simulation and sideband data, the continuum background has a small linear dependence on the dimuon invariant mass. In the simulation, the slope parameter has a roughly linear dependence versus BDT interval; the mass sidebands in data confirm this trend, albeit with large statistical uncertainty. This dependence is included in the fit model. The small systematic uncertainties due to deviations from this assumption are discussed below in section 10.3.

The $b \to \mu^+\mu^-X$ background has a dimuon invariant mass distribution that falls monotonically with increasing dimuon mass. The mass dependence is derived from data in the low mass sideband, and described with an exponential function with the same shape in each BDT interval. The value of the shape parameter is extracted from the fit to data.

The invariant mass distribution of the peaking background is very similar to the $B^0$ signal, as shown in figure 1(b). The description of this component is obtained from MC simulation, which indicates that the shape and normalisation are the same for all BDT bins. In the fit, this contribution is included with fixed mass shape and with a normalisation of $2.9 \pm 2.0$ events, as discussed in section 5. This contribution is equally distributed among the three highest intervals of the BDT output.

The fitting procedure is tested with MC pseudo-experiments, as discussed in section 10.3.
10.2 Relative signal efficiency between BDT bins

Section 9 explains how systematic uncertainties affect the overall selection efficiency for signal candidates. The separation of the candidates according to BDT bins introduces an additional dependence on the relative efficiencies in each BDT bin, and systematic uncertainties in these relative efficiencies must be accounted for. Two different procedures are explored. First, the distribution of the BDT output is compared between MC simulation and background-subtracted data for the reference and control channels. The differences observed in the ratio of data to simulation are described with a linear dependence on the BDT output. The linear dependencies observed for \( B^+ \to J/\psi K^+ \) and \( B^{0}_{s} \to J/\psi \phi \) are in turn used to reweight the BDT-output distribution in the \( B^{0}_{(s)} \to \mu^+\mu^- \) MC sample. The maximum corresponding absolute variations in the efficiencies are equal to +1.7% and −2.3% respectively in the second and fourth BDT intervals, with the third interval basically unaffected. A second assessment of the systematic uncertainties in the relative efficiency of the BDT intervals is obtained with a procedure similar to the one used for the event selection (section 9). For each discriminating variable, the MC sample is reweighted according to the difference between simulation and data observed in the reference channels. The variation in the efficiency of each BDT interval is taken as the contribution to the systematic uncertainty due to mis-modelling of that variable. The sum in quadrature of the variations due to all discriminating variables is found to be similar in the \( B^+ \to J/\psi K^+ \) and \( B^{0}_{s} \to J/\psi \phi \) channels. Absolute variations of ±1.0%, ±2.4% and ±4.4% are found in the second, third and fourth BDT intervals respectively. The first of these procedures is used as a baseline for inclusion of Gaussian terms in the signal extraction likelihood to account for the uncertainty in the relative signal efficiency in the three most sensitive BDT bins. Care is taken in constraining the sum of the efficiencies of the three intervals sensitive to the signal, since that absolute efficiency and the corresponding uncertainty are parameterised with the \( R_e \) term.

Figure 7 shows the distribution of the BDT output from data and simulation for the reference channels, after reweighting the MC sample. The MC distribution for \( B^{0}_{s} \to \mu^+\mu^- \) events is also shown, illustrating how the linear deviation obtained from the reference channels affects the simulated signal BDT output. When studying these effects, the linear fits to the ratios in figures 7(a) and 7(b) are performed in the range corresponding to the three BDT bins with the highest signal-to-background ratio, since the remaining bin is insensitive to the signal contribution.

10.3 Systematic uncertainties in the fit

Studies based on MC pseudo-experiments are used to assess the sensitivity of the fit to the input assumptions. Variations in the description of signal and background components are used in the generation of these samples. The corresponding changes in the average numbers, \( N_s \) and \( N_d \), of \( B^{0}_{s} \) and \( B^{0} \) events determined by the fit, run in the nominal configuration, are taken as systematic uncertainties. The size of the variations used in the generation of the MC pseudo-experiments is determined in some cases by known characteristics of the ATLAS detector (reconstructed momentum scale and momentum resolution), in others using MC evaluation (background due to semileptonic \( B^{0}_{(s)}/B^{0}_{b} \to h\mu\nu \) decays and to \( B^{+}_{c} \to J/\psi \mu^+\nu \),...
Figure 7. BDT value distributions in data and MC simulation for (a) $B^+ \rightarrow J/\psi K^+$, (b) $B^0_s \rightarrow J/\psi \phi$. The MC samples are normalised to the number of data events passing the signal reference BDT selection (section 6). Figure (c) illustrates the BDT output for the $B^0_s \rightarrow \mu^+ \mu^-$ signal, with the dashed histogram illustrating the effect of the linear reweighting on the BDT output discussed in the text. The vertical dashed lines correspond to the boundaries of the BDT intervals used in the $B^0_{(s)} \rightarrow \mu^+ \mu^-$ signal fit.

and in others from uncertainties determined from data in the sidebands or from simulation (shapes of the background components and their variation across the BDT intervals).

The MC pseudo-experiments were generated with the normalisation of the continuum and $b \rightarrow \mu \mu X$ components obtained from the fit to the data in the sidebands of the invariant mass distribution, and the peaking background from the expectation discussed in section 5. The signal was generated with different configurations, roughly covering the range between zero and twice the expected SM yield.

For all variations of the assumptions and all configurations of the signal amplitudes the distributions of the differences between fit results and generated values are used to evaluate systematic uncertainties. In addition, distributions obtained from MC pseudo-experiments generated and fitted according to the nominal fit model are used to study systematic biases deriving from the fit procedure. For both signal yields, the bias is smaller than 15% of the fit error, for true values of the $B^0_s \rightarrow \mu^+ \mu^-$ branching fraction above $5 \times 10^{-10}$. 

---

{20
The shifts in $N_s$ or $N_d$ are combined by considering separately the sums in quadrature of the positive and negative shifts and taking the larger as the symmetric systematic uncertainty. The total systematic uncertainty is found to increase with the assumed size of the signal, with a dependence $\sigma_{syst}^{N_s} = 3 + 0.05N_s$ and $\sigma_{syst}^{N_d} = 2.9 + 0.05N_s + 0.05N_d$. Most of the shifts observed have opposite sign for $N_s$ and $N_d$, resulting in a combined correlation coefficient in the systematic uncertainties of $\rho_{syst} = -0.83$.

The systematic uncertainties discussed in this section are included in the fit to the $\mu^+\mu^-$ candidates in data. The fit for the yield of $B^0_s$ and $B^0$ events is modified by including in the likelihood two smearing parameters for $N_s$ and $N_d$ that are constrained by a two-dimensional Gaussian distribution parameterised by the values of $\sigma_{syst}^{N_s}$, $\sigma_{syst}^{N_d}$ and $\rho_{syst}$. 

---

### Figure 8

Dimuon invariant mass distributions in the unblinded data, in the four intervals of BDT output. Superimposed is the result of the maximum-likelihood fit. The total fit is shown as a continuous line, with the dashed lines corresponding to the observed signal component, the $b \to \mu\mu X$ background, and the continuum background. The signal components are grouped in one single curve, including both the $B^0_s \to \mu^+\mu^-$ and the (negative) $B^0 \to \mu^+\mu^-$ component. The curve representing the peaking $B^0_{(s)} \to hh'$ background lies very close to the horizontal axis in all BDT bins.
10.4 Results of the signal yield extraction

The numbers of background events contained in the signal region (5166–5526 MeV) are computed from the interpolation of the data observed in the sidebands. This procedure yields 2685 $^{+37}_{-330}$, 330 $^{+14}_{-14}$, 51 $^{+6}_{-6}$ and 7 $^{+2}_{-2}$ events respectively in the four intervals of BDT output. For comparison, the total expected numbers of signal events according to the SM prediction are 91 and 10 for $N_s$ and $N_d$ respectively, equally distributed among the three intervals with the highest signal-to-background ratio.

In those three BDT intervals, in the unblinded signal region, a total of 1951 events in the full mass range of 4766–5966 MeV are used in the likelihood fit to signal and background. Without applying any bounds on the values of the fitted parameters, the values determined by the fit are $N_s = 80 \pm 22$ and $N_d = -12 \pm 20$, where the uncertainties correspond to likelihood variations satisfying $-2 \Delta \ln(\mathcal{L}) = 1$. The likelihood includes the systematic uncertainties discussed above, but statistical uncertainties largely dominate. The result is consistent with the expectation from simulation. The uncertainties in the result of the fit are discussed in section 11, where the measured values of the branching fractions are presented.

Figure 8 shows the dimuon invariant mass distributions in the four BDT intervals, together with the projections of the likelihood. A modified Kolmogorov-Smirnov (KS) test [35] is used to estimate the fit quality: the $p$-value is estimated by comparing the maximum of the KS distance across the four histograms of figure 8 with the distribution of the same quantity from pseudo-experiments generated with the shape resulting from the fit to data. This procedure yields a compatibility probability of 84%.

11 Branching fraction extraction

The branching fractions for the decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are extracted from data using a maximum-likelihood fit. The likelihood is obtained from the one used for $N_s$ and $N_d$ by replacing the fit parameters with the corresponding branching fractions divided by normalisation terms in eq. (1.1), and including Gaussian multiplicative factors for the normalisation uncertainties. All results are obtained profiling the fit likelihood with respect to all parameters involved other than the branching fraction(s) of interest.

The normalisation terms include external inputs for the $B^+$ branching fraction and the relative hadronisation probability. The branching fraction is obtained from world averages [29] as the product of $\mathcal{B}(B^+ \to J/\psi K^+) = (1.010 \pm 0.029) \times 10^{-3}$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033)\%$. The relative hadronisation probability is equal to one for $B^0$, while for $B_s^0$ it is taken from the latest available HFLAV average [34] $f_s/f_d = 0.256 \pm 0.013$, which assumes $f_u/f_d = 1$.

The efficiency-weighted and luminosity-weighted number of events for the reference channel enters in eq. (1.1) with the denominator $D_{\text{ref}}$. The values $D_{\text{ref}} = (5.69 \pm 0.36) \times 10^6$ for $B_s^0$ and $(5.84 \pm 0.37) \times 10^6$ for $B^0$ are obtained from the results of sections 8 and 9, and include the correction to the $B_s^0 \to \mu^+\mu^-$ efficiency due to the lifetime difference between $B_{s,H}$ and $B_s^0$. The combination of $B^+$ branching fraction, hadronisation probabilities and $D_{\text{ref}}$, i.e. the single-event sensitivity, is equal to $(4.02 \pm 0.35) \times 10^{-11}$ for $B_s^0 \to \mu^+\mu^-$ and $(1.059 \pm 0.074) \times 10^{-11}$ for $B^0 \to \mu^+\mu^-$. 
Table 3. Breakdown of the expected systematic uncertainties in $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The measurements are dominated by statistical uncertainty, followed by the systematic uncertainty from the fit. The latter is dominated by contributions from the mass scale uncertainty and the parameterisation of the $b \to \mu^+\mu^-X$ background. The statistical uncertainties reported here are obtained from the maximisation of the fit likelihood and are meant only as a reference for the relative scale uncertainties.

The values of the branching fractions that maximise the profile likelihood are $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.2 \pm 0.9) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = (-1.3 \pm 2.1) \times 10^{-10}$, where the uncertainties, both statistical and systematic combined, are estimated using a Gaussian approximation in the neighbourhood of the likelihood function maximum. Figure 9(a) shows likelihood contours for the simultaneous fit to $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$, for values of $-2\Delta \ln(L)$ equal to 2.3, 6.2 and 11.8, relative to the maximum of the likelihood.

Table 3 gives a breakdown of the estimated contributions of systematic and statistical uncertainties. The results are dominated by statistical uncertainties, with the most prominent source of systematic uncertainty coming from fit uncertainties, where the largest contributors are the mass scale and $b \to \mu^+\mu^-X$ background parameterisation.

Given the statistical regime of the analysis, the likelihood contours of figure 9(a) cannot be immediately translated into contours with the conventional coverage of one, two and three Gaussian standard deviations. Moreover, the contours extend into regions of negative branching fractions, which are unphysical. In order to address these points, a Neyman construction [36] is employed to obtain the 68.3%, 95.5% and 99.7% confidence intervals in the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)\mathcal{B}(B^0 \to \mu^+\mu^-)$ plane. This construction yields the contours shown in figure 9(b). The same construction is used to determine the 68.3% confidence interval for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ with MC pseudo-experiments, and results in

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.21^{+0.96}_{-0.91}^{+0.49}) \times 10^{-9} = (3.2^{+1.1}_{-1.0}) \times 10^{-9}.$$  

Statistical and systematic uncertainties (shown separately in the first expression, and combined in the last) are separated by repeating the likelihood fit after setting all systematic uncertainties to zero. The fitted branching fractions (obtained from the unconstrained likelihood maximum) are in all cases inputs to the Neyman construction, which, by design, results in physically allowed values for the resulting branching fractions.

| Source                  | $B_s^0$ [%] | $B^0$ [%] |
|-------------------------|-------------|-----------|
| $f_s/f_d$               | 5.1         | —         |
| $B^+$ yield             | 4.8         | 4.8       |
| $R_{e}$                 | 4.1         | 4.1       |
| $\mathcal{B}(B^+ \to J/\psi K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)$ | 2.9 | 2.9 |
| Fit systematic uncertainties | 8.7 | 65 |
| Stat. uncertainty (from likelihood est.) | 27 | 150 |
The upper limit at 95% CL on $\mathcal{B}(B^0 \to \mu^+ \mu^-)$, determined with the same Neyman procedure, is

$$\mathcal{B}(B^0 \to \mu^+ \mu^-) < 4.3 \times 10^{-10}.$$  

Using the predicted SM branching fractions from section 1, the analysis is expected to yield on average a measurement of $(3.6^{+1.1}_{-1.0}) \times 10^{-9}$ for $\mathcal{B}(B_s^0 \to \mu^+ \mu^-)$ and an upper limit at 95% CL of $7.1 \times 10^{-10}$ for $\mathcal{B}(B^0 \to \mu^+ \mu^-)$.

### 12 Combination with the run 1 result

The likelihood function from the current result is combined with the likelihood function from the Run 1 result [15]. The only common parameters in the combination are the fitted $\mathcal{B}(B_s^0 \to \mu^+ \mu^-)$ and the combination of external inputs $\mathcal{F}_{\text{ext}} = \mathcal{B}(B^+ \to J/\psi K^+) \times \mathcal{B}(J/\psi \to \mu^+ \mu^-) \times (f_u/f_s) = (2.35 \pm 0.14) \times 10^{-4}$. Except for $\mathcal{F}_{\text{ext}}$, all nuisance parameters are treated as uncorrelated between the two likelihoods, with both likelihoods including their individual parameterisations of systematic uncertainty effects. A negligible change in the results, corresponding to shifts in central values and uncertainties between 1% and 4%, is found when all sources of systematic uncertainty are assumed to be fully correlated.

The maximum of the combined likelihood corresponds to

$$\mathcal{B}(B_s^0 \to \mu^+ \mu^-) = (2.8 \pm 0.7) \times 10^{-9},$$

$$\mathcal{B}(B^0 \to \mu^+ \mu^-) = (-1.9 \pm 1.6) \times 10^{-10}.$$  

Figure 10 shows the likelihood contours for the combined Run 1 and Run 2 result for $\mathcal{B}(B_s^0 \to \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \to \mu^+ \mu^-)$, for values of $-2\Delta \ln (\mathcal{L})$ equal to 2.3, 6.2 and 11.8,
relative to the maximum of the likelihood. The contours for the result from 2015–2016 Run 2 data are overlaid for comparison.

When applying the one-dimensional Neyman construction described in section 11 to this combined likelihood, whose maximum is unconstrained and allowed to access the unphysical (negative) region, the 68.3% confidence interval obtained for $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ is

$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = (2.8^{+0.8}_{-0.7}) \times 10^{-9}.$$ 

The upper limit at 95% CL on $\mathcal{B}(B^0 \to \mu^+\mu^-)$ is determined with the same Neyman procedure, yielding

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.1 \times 10^{-10}.$$ 

Using the predicted SM branching fractions from section 1, the analysis is expected to yield on average a measurement of $(3.6^{+0.9}_{-0.8}) \times 10^{-9}$ for $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ and an upper limit of $5.6 \times 10^{-10}$ for $\mathcal{B}(B^0 \to \mu^+\mu^-)$.

The Run 1 and Run 2 results are found to be 1.2 standard deviations apart. Using both runs, the combined significance of the $B^0_s \to \mu^+\mu^-$ signal is estimated to be 4.6 standard deviations, and the combined branching fraction measurements differ by 2.4 stan-
standard deviations from the SM values in the $\mathcal{B}(B^0 \to \mu^+\mu^-) - \mathcal{B}(B^0_s \to \mu^+\mu^-)$ plane. These significances are assessed purely from the evaluation of likelihood ratios.

13 Conclusions

A study of the rare decays of $B^0_s$ and $B^0$ mesons into oppositely charged muon pairs is presented, based on 36.2 fb$^{-1}$ of 13 TeV LHC proton-proton collision data collected by the ATLAS experiment in 2015 and 2016.

For the $B^0_s$ the branching fraction is determined to be $\mathcal{B}(B^0_s \to \mu^+\mu^-) = (3.2^{+1.1}_{-1.0}) \times 10^{-9}$, where the uncertainty includes both the statistical and systematic contributions. The result is consistent with the analysis expectation of $(3.6^{+1.1}_{-1.0}) \times 10^{-9}$ in the SM hypothesis.

For the $B^0$ an upper limit $\mathcal{B}(B^0 \to \mu^+\mu^-) < 4.3 \times 10^{-10}$ is placed at the 95\% confidence level, with an expected upper bound of $7.1 \times 10^{-10}$ in the SM hypothesis. The limit is compatible with the SM prediction.

The result presented in this paper is combined with the ATLAS result from the full Run 1 dataset to obtain $\mathcal{B}(B^0_s \to \mu^+\mu^-) = (2.8^{+0.8}_{-0.7}) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) < 2.1 \times 10^{-10}$. All the results presented are compatible with the branching fractions predicted by the SM as well as with currently available experimental results.

Acknowledgments

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; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPERJ, 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; DLR and DNNRC, Germany; IN2P3-CNRS, CEA-DRF/IRFU, France; BMBF, DFG and HGF, Germany; INFN, Italy; MEXT and JSPS, Japan; SRNSFG, Greece; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MSTDF, Serbia; MSSR, Slovakia; ARRS and MIZS, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, CRC and Compute Canada, Canada; COST, ERC, ERDF, Horizon 2020, and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; CERCA Programme Generalitat de Catalunya, Spain; The Royal Society and Leverhulme Trust, United Kingdom.
The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, 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 (U.K.) and BNL (U.S.A.), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. [37].

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] C. Bobeth, M. Gorbahn, T. Hermann, M. Misiak, E. Stamou and M. Steinhauser, $B_{s,d} \to \ell^+\ell^-$ in the Standard Model with reduced theoretical uncertainty, Phys. Rev. Lett. 112 (2014) 101801 [arXiv:1311.0903] [inSPIRE].

[2] C.-S. Huang, W. Liao and Q.-S. Yan, The promising process to distinguish supersymmetric models with large $\tan \beta$ from the Standard Model: $B \to X_s\mu^+\mu^-$, Phys. Rev. D 59 (1999) 011701 [hep-ph/9803460] [inSPIRE].

[3] C. Hamzaoui, M. Pospelov and M. Toharia, Higgs mediated FCNC in supersymmetric models with large $\tan \beta$, Phys. Rev. D 59 (1999) 095005 [hep-ph/9807350] [inSPIRE].

[4] S.R. Choudhury and N. Gaur, Dileptonic decay of $B_s$ meson in SUSY models with large $\tan \beta$, Phys. Lett. B 451 (1999) 86 [hep-ph/9810307] [inSPIRE].

[5] K.S. Babu and C.F. Kolda, Higgs mediated $B_0 \to \mu^+\mu^-$ in minimal supersymmetry, Phys. Rev. Lett. 84 (2000) 228 [hep-ph/9909476] [inSPIRE].

[6] S.R. Choudhury, A.S. Cornell, N. Gaur and G.C. Joshi, Signatures of new physics in dileptonic $B$-decays, Int. J. Mod. Phys. A 21 (2006) 2617 [hep-ph/0504193] [inSPIRE].

[7] G. D'Ambrosio, G.F. Giudice, G. Isidori and A. Strumia, Minimal flavour violation: an effective field theory approach, Nucl. Phys. B 645 (2002) 155 [hep-ph/0207036] [inSPIRE].

[8] A.J. Buras, Relations between $\Delta M_{s,d}$ and $B_{s,d} \to \mu\bar{\mu}$ in models with minimal flavour violation, Phys. Lett. B 566 (2003) 115 [hep-ph/0303060] [inSPIRE].

[9] S. Davidson and S. Descotes-Genon, Minimal flavour violation for leptoquarks, JHEP 11 (2010) 073 [arXiv:1009.1998] [inSPIRE].

[10] D. Guadagnoli and G. Isidori, $B(B_s \to \mu^+\mu^-)$ as an electroweak precision test, Phys. Lett. B 724 (2013) 63 [arXiv:1302.3909] [inSPIRE].

[11] CMS collaboration, Measurement of the $B^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment, Phys. Rev. Lett. 111 (2013) 101804 [arXiv:1307.5026] [inSPIRE].

[12] LHCb collaboration, Measurement of the $B^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment, Phys. Rev. Lett. 111 (2013) 101805 [arXiv:1307.5024] [inSPIRE].
[13] CMS and LHCb collaborations, Observation of the rare $B^0_s \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data, Nature 522 (2015) 68 [arXiv:1411.4413] [inSPIRE].

[14] LHCb collaboration, Measurement of the $B^0_s \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays, Phys. Rev. Lett. 118 (2017) 191801 [arXiv:1703.05747] [inSPIRE].

[15] ATLAS collaboration, Study of the rare decays of $B^0_s$ and $B^0$ into muon pairs from data collected during the LHC run 1 with the ATLAS detector, Eur. Phys. J. C 76 (2016) 513 [arXiv:1604.04263] [inSPIRE].

[16] K. De Bruyn et al., Probing new physics via the $B^0_s \to \mu^+\mu^-$ effective lifetime, Phys. Rev. Lett. 109 (2012) 041801 [arXiv:1204.1737] [inSPIRE].

[17] A.J. Buras, R. Fleischer, J. Girrbach and R. Knegjens, Probing new physics with the $B^0_s \to \mu^+\mu^-$ time-dependent rate, JHEP 07 (2013) 77 [arXiv:1303.3820] [inSPIRE].

[18] ATLAS collaboration, The ATLAS experiment at the CERN Large Hadron Collider, 2008 JINST 3 S08003 [inSPIRE].

[19] ATLAS IBL collaboration, Production and integration of the ATLAS insertable B-layer, 2018 JINST 13 T05008 [arXiv:1803.00844] [inSPIRE].

[20] ATLAS collaboration, Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC, Eur. Phys. J. C 76 (2016) 653 [arXiv:1608.03953] [inSPIRE].

[21] T. Sjöstrand, S. Mrenna and P.Z. Skands, A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (2008) 852 [arXiv:0710.3820] [inSPIRE].

[22] GEANT4 collaboration, GEANT4: a simulation toolkit, Nucl. Instrum. Meth. A 506 (2003) 250 [inSPIRE].

[23] ATLAS collaboration, The ATLAS simulation infrastructure, Eur. Phys. J. C 70 (2010) 823 [arXiv:1005.4568] [inSPIRE].

[24] ATLAS collaboration, Measurement of quarkonium production in proton-lead and proton-proton collisions at 5.02 TeV with the ATLAS detector, Eur. Phys. J. C 78 (2018) 171 [arXiv:1709.03089] [inSPIRE].

[25] ATLAS collaboration, Search for the decay $B^0_s \to \mu\mu$ with the ATLAS detector, Phys. Lett. B 713 (2012) 387 [arXiv:1204.0735] [inSPIRE].

[26] ATLAS collaboration, Performance of the ATLAS trigger system in 2015, Eur. Phys. J. C 77 (2017) 317 [arXiv:1611.09661] [inSPIRE].

[27] ATLAS collaboration, Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C 76 (2016) 292 [arXiv:1603.05598] [inSPIRE].

[28] Particle Data Group collaboration, Review of particle physics, Phys. Rev. D 98 (2018) 030001 [inSPIRE].

[29] ATLAS collaboration, Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data, Eur. Phys. J. C 74 (2014) 3130 [arXiv:1407.3938] [inSPIRE].
[31] A. Hocker et al., *TMVA — toolkit for multivariate data analysis*, physics/0703039 [inSPIRE].

[32] N.L. Johnson, *Systems of frequency curves generated by methods of translation*, *Biometrika* **36** (1949) 149 [inSPIRE].

[33] M.C. Jones and A. Pewsey, *Sinh-arcsinh distributions*, *Biometrika* **96** (2009) 761.

[34] HFLAV collaboration, *Averages of b-hadron, c-hadron and \( \tau \)-lepton properties as of summer 2016*, *Eur. Phys. J. C* **77** (2017) 895 [arXiv:1612.07233] [inSPIRE].

[35] A. Kolmogorov, *Sulla determinazione empirica di una legge di distribuzione* (in Italian), *Giorn. Ist. Ital. Attuar.* **4** (1933) 83.

[36] J. Neyman, *Outline of a theory of statistical estimation based on the classical theory of probability*, *Phil. Trans. Roy. Soc. Lond. A* **236** (1937) 333 [inSPIRE].

[37] ATLAS collaboration, *ATLAS computing acknowledgements*, ATL-GEN-PUB-2016-002, CERN, Geneva, Switzerland (2016).
Institute for High Energy Physics of the National Research Centre Kurchatov Institute, Protvino; Russia
Department of Physics, New York University, New York NY; United States of America
Ohio State University, Columbus OH; United States of America
Faculty of Science, Okayama University, Okayama; Japan
Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman OK; United States of America
Department of Physics, Oklahoma State University, Stillwater OK; United States of America
Palacky University, RCPTM, Joint Laboratory of Optics, Olomouc; Czech Republic
Center for High Energy Physics, University of Oregon, Eugene OR; United States of America
LAL, Université Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay; France
Graduate School of Science, Osaka University, Osaka; Japan
Department of Physics, University of Oslo, Oslo; Norway
Department of Physics, Oxford University, Oxford; United Kingdom
LPNHE, Sorbonne Université, Paris Diderot Sorbonne Paris Cité, CNRS/IN2P3, Paris; France
Konstantinov Nuclear Physics Institute of National Research Centre “Kurchatov Institute”, PNPI, St. Petersburg; Russia
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh PA; United States of America
Laboratório de Instrumentação e Física Experimental de Partículas - LIP, Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Lisboa; Departamento de Física, Universidade de Coimbra, Coimbra; Centro de Física Nuclear da Universidade de Lisboa, Lisboa; Departamento de Física, Universidade do Minho, Braga; Departamento de Física Teórica y del Cosmos, Universidad de Granada, Granada (Spain); Dep Física and CEFITEC of Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica; Portugal
Institute of Physics, Academy of Sciences of the Czech Republic, Prague; Czech Republic
Czech Technical University in Prague, Prague; Czech Republic
Charles University, Faculty of Mathematics and Physics, Prague; Czech Republic
Particle Physics Department, Rutherford Appleton Laboratory, Didcot; United Kingdom
IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette; France
Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz CA; United States of America
Departamento de Física, Pontificia Universidad Católica de Chile, Santiago; Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso; Chile
Department of Physics, University of Washington, Seattle WA; United States of America
Department of Physics and Astronomy, University of Sheffield, Sheffield; United Kingdom
Department of Physics, Shinshu University, Nagano; Japan
Department Physik, Universität Siegen, Siegen; Germany
Department of Physics, Simon Fraser University, Burnaby BC; Canada
SLAC National Accelerator Laboratory, Stanford CA; United States of America
Physics Department, Royal Institute of Technology, Stockholm; Sweden
Departments of Physics and Astronomy, Stony Brook University, Stony Brook NY; United States of America
Department of Physics and Astronomy, University of Sussex, Brighton; United Kingdom
School of Physics, University of Sydney, Sydney; Australia
Institute of Physics, Academia Sinica, Taipei; Taiwan
E. Andronikashvili Institute of Physics, Iv. Javakhishvili Tbilisi State University, Tbilisi; High Energy Physics Institute, Tbilisi State University, Tbilisi; Georgia
Department of Physics, Technion, Israel Institute of Technology, Haifa; Israel
Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv; Israel
159 Department of Physics, Aristotle University of Thessaloniki, Thessaloniki; Greece
International Center for Elementary Particle Physics and Department of Physics, University of Tokyo, Tokyo; Japan
Graduate School of Science and Technology, Tokyo Metropolitan University, Tokyo; Japan
Department of Physics, Tokyo Institute of Technology, Tokyo; Japan
Tomsk State University, Tomsk; Russia
Department of Physics, University of Toronto, Toronto ON; Canada
Division of Physics and Tomonaga Center for the History of the Universe, Faculty of Pure and Applied Sciences, University of Tsukuba, Tsukuba; Japan
Department of Physics and Astronomy, Tafts University, Medford MA; United States of America
Department of Physics and Astronomy, University of California Irvine, Irvine CA; United States of America
Department of Physics and Astronomy, University of Uppsala, Uppsala; Sweden
Department of Physics, University of Illinois, Urbana IL; United States of America
Instituto de Física Corpuscular (IFIC), Centro Mixto Universidad de Valencia - CSIC, Valencia; Spain
Department of Physics, University of British Columbia, Vancouver BC; Canada
Department of Physics and Astronomy, University of Victoria, Victoria BC; Canada
Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, Würzburg; Germany
Department of Physics, University of Warwick, Coventry; United Kingdom
Waseda University, Tokyo; Japan
Department of Particle Physics, Weizmann Institute of Science, Rehovot; Israel
Department of Physics, University of Wisconsin, Madison WI; United States of America
Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal; Germany
Department of Physics, Yale University, New Haven CT; United States of America
Yerevan Physics Institute, Yerevan; Armenia

a Also at Borough of Manhattan Community College, City University of New York, NY; United States of America
b Also at California State University, East Bay; United States of America
c Also at Centre for High Performance Computing, CSIR Campus, Rosebank, Cape Town; South Africa
d Also at CERN, Geneva; Switzerland
e Also at CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille; France
f Also at Département de Physique Nucléaire et Corpusculaire, Université de Genève, Genève; Switzerland
g Also at Departament de Física de la Universitat Autonoma de Barcelona, Barcelona; Spain
h Also at Departamento de Física Teorica y del Cosmos, Universidad de Granada, Granada (Spain); Spain
i Also at Departamento de Física, Instituto Superior Técnico, Universidade de Lisboa, Lisboa; Portugal
j Also at Department of Applied Physics and Astronomy, University of Sharjah, Sharjah; United Arab Emirates
k Also at Department of Financial and Management Engineering, University of the Aegean, Chios; Greece
l Also at Department of Physics and Astronomy, University of Louisville, Louisville, KY; United States of America
m Also at Department of Physics and Astronomy, University of Sheffield, Sheffield; United Kingdom
Also at Department of Physics, California State University, Fresno CA; United States of America
Also at Department of Physics, California State University, Sacramento CA; United States of America
Also at Department of Physics, King’s College London, London; United Kingdom
Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg; Russia
Also at Department of Physics, Stanford University; United States of America
Also at Department of Physics, University of Fribourg, Fribourg; Switzerland
Also at Department of Physics, University of Michigan, Ann Arbor MI; United States of America
Also at Giresun University, Faculty of Engineering, Giresun; Turkey
Also at Graduate School of Science, Osaka University, Osaka; Japan
Also at Hellenic Open University, Patras; Greece
Also at Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest; Romania
Also at I. Physikalisches Institut, Georg-August-Universität Göttingen, Göttingen; Germany
Also at Institucio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona; Spain
Also at Institut für Experimentalphysik, Universität Hamburg, Hamburg; Germany
Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest; Hungary
Also at Institute of Physics, Academia Sinica, Taipei; Taiwan
Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku; Azerbaijan
Also at Institute of Theoretical Physics, Ilia State University, Tbilisi; Georgia
Also at Joint Institute for Nuclear Research, Dubna; Russia
Also at Louisiana Tech University, Ruston LA; United States of America
Also at LNL, Sorbonne Université, Paris Diderot Sorbonne Paris Cité, CNRS/IN2P3, Paris; France
Also at Manhattan College, New York NY; United States of America
Also at Moscow Institute of Physics and Technology State University, Dolgoprudny; Russia
Also at National Research Nuclear University MEPhI, Moscow; Russia
Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg; Germany
Also at School of Physics, Sun Yat-sen University, Guangzhou; China
Also at The City College of New York, New York NY; United States of America
Also at The Collaborative Innovation Center of Quantum Matter (CICQM), Beijing; China
Also at Tomsk State University, Tomsk, and Moscow Institute of Physics and Technology State University, Dolgoprudny; Russia
Also at TRIUMF, Vancouver BC; Canada
Also at Università di Napoli Parthenope, Napoli; Italy

* Deceased