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

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation
ATLAS Collaboration. 2016. 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. The European Physical Journal C 76, no. 9. doi:10.1140/epjc/s10052-016-4338-8.

Published Version
doi:10.1140/epjc/s10052-016-4338-8

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:29362021

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP
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

The ATLAS Collaboration

Abstract

A study of the decays $B^0 \rightarrow \mu^+\mu^-$ and $B^0_s \rightarrow \mu^+\mu^-$ has been performed using data corresponding to an integrated luminosity of 25 fb$^{-1}$ of 7 TeV and 8 TeV proton–proton collisions collected with the ATLAS detector during the LHC Run 1. For $B^0$, an upper limit on the branching fraction is set at $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 4.2 \times 10^{-10}$ at 95% confidence level. For $B^0_s$, the branching fraction $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) = (0.9^{+1.1}_{-0.8}) \times 10^{-9}$ is measured. The results are consistent with the Standard Model expectation with a $p$-value of 4.8%, corresponding to 2.0 standard deviations.
1 Introduction

Flavour-changing neutral-current (FCNC) processes are highly suppressed in the Standard Model (SM), and their study is relevant to indirect searches for physics beyond the SM. The branching fractions of the decays $B^0_{(s)} \to \mu^+\mu^-$ are of particular interest because of the additional helicity suppression and since they are accurately predicted in the SM: $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = (3.65 \pm 0.23) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.06 \pm 0.09) \times 10^{-10}$ [1]. Significant deviations from these values can arise in models involving non-SM heavy particles, such as those predicted in the Minimal Supersymmetric Standard Model [2–6] in extensions such as Minimal Flavour Violation [7, 8], Two-Higgs-Doublet Models [6], and others [9, 10]. The CMS and LHCb collaborations have reported the observation of $B^0 \to \mu^+\mu^-$ [11, 12] and evidence of $B^0 \to \mu^+\mu^-$, with combined values: $\mathcal{B}(B^0 \to \mu^+\mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10}$ [13].

This paper reports the result of a search for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays performed using $pp$ collision data corresponding to an integrated luminosity of 25 fb$^{-1}$, collected at 7 and 8 TeV in the full LHC Run 1 data-taking period using the ATLAS detector. This analysis supersedes the previous result [14] based on 2011 data and exploits improved analysis techniques in addition to the larger dataset.

2 Outline

In order to minimise systematic uncertainties, the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ branching fractions are measured relative to the normalisation decay $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ that has a known branching fraction, $\mathcal{B}(B^+ \to J/\psi K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)$. The procedure needs to be corrected firstly for the ratio of the different probabilities of the $b$ quark to hadronise with a $u$ or an $s$ quark and this correction enters as the relative $b$-quark hadronisation probability $f_u/f_s$ or $f_d/f_s$. Secondly, the different selection efficiencies $\varepsilon_{\mu^+\mu^-}$ and $\varepsilon_{J/\psi K^+}$, for the signal and normalisation channels, must be taken into account. In the simplest form, the $B^0 \to \mu^+\mu^-$ ($B^0_{(s)} \to \mu^+\mu^-$) branching fraction can be extracted as:

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = \frac{N_{d(s)}}{\varepsilon_{\mu^+\mu^-}} \times \left[ \frac{\mathcal{B}(B^+ \to J/\psi K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)}{\mathcal{B}(B^0 \to J/\psi K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)} \right] \frac{f_u}{f_{d(s)}} \times \frac{1}{\mathcal{D}_{\text{norm}}},$$

where $N_d$ ($N_s$) is the $B^0 \to \mu^+\mu^-$ ($B^0_{(s)} \to \mu^+\mu^-$) signal yield and $N_{J/\psi K^+}$ is the $B^+ \to J/\psi K^+$ normalisation yield. For this study, a modified formula is used to include events selected in different data-taking periods and with different trigger selections:

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = \frac{N_{d(s)}}{\varepsilon_{\mu^+\mu^-}} \times \left[ \frac{\mathcal{B}(B^+ \to J/\psi K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)}{\mathcal{B}(B^0 \to J/\psi K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)} \right] \frac{f_u}{f_{d(s)}} \times \frac{1}{\mathcal{D}_{\text{norm}}}, \quad (1)$$

with

$$\mathcal{D}_{\text{norm}} = \sum_k N_{J/\psi K^+} \alpha_k \left( \frac{\varepsilon_{\mu^+\mu^-}}{\varepsilon_{J/\psi K^+}} \right)_k.$$  

The denominator $\mathcal{D}_{\text{norm}}$ consists of a sum whose index $k$ runs over the data-taking periods and the trigger selections. In the sum, the $\alpha_k$ parameter takes into account the different trigger prescale factors and integrated luminosities in the signal and normalisation channels, and the ratio of the efficiencies corrects for reconstruction differences in each data sample $k$. Signal and reference channel events are selected with
similar dimuon triggers and the normalisation is performed independently for each data-taking period and trigger category.

The notation used throughout the paper refers to both the stated and charge-conjugated process, unless otherwise specified. The analysis is performed without tagging of the flavour $B_0^{(*)}$ or $\bar{B}_0^{(*)}$ at production. The yield measurement in the normalisation channel is obtained by summing $J/\psi K^+$ and $J/\psi K^-$ contributions.

The ATLAS inner tracking system and muon spectrometer are used to reconstruct and select the event candidates. Details of the detector, trigger, data sets, and preliminary selection criteria are discussed in Sections 3 and 4.

A blind analysis was performed in which data in the dimuon invariant mass region from 5166 to 5526 MeV were removed until the procedures for event selection and the details of signal yield extraction were completely defined. The variables used in the event selection, Monte Carlo (MC) tuning and background studies are discussed in Sections 3 to 5. A dedicated strategy was developed to greatly reduce the background related to hadrons from $B_0^{(*)}$ decays which are misidentified as muons, as described in Section 6. The final sample of candidates is selected using a multivariate classifier, trained on MC samples, as discussed in Section 7. Checks on the distributions of the variables used in the multivariate classifier are summarised in Section 8. Section 9 details the fit procedure to extract the yield of $B^+ \to J/\psi K^+$ events. As an ancillary measurement to the $B^+ \to J/\psi K^+$ yield determination, a measurement of the ratio $B(B^+ \to J/\psi \pi^+)/B(B^+ \to J/\psi K^+)$ is performed, as presented in Subsection 9.1. The ratio of efficiencies in the signal and the normalisation channels is presented in Section 10. The extraction of the signal yield is discussed in Section 11, and the branching fractions are reported in Section 12.

3 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. [15].

This analysis is based on the Run 1 data sample recorded in 2011 and 2012 by the ATLAS detector from $pp$ collisions at the LHC at $\sqrt{s} = 7$ and 8 TeV, respectively. Data used in the analysis were recorded during stable LHC beam periods. Data quality requirements were imposed, notably on the performance of the MS and ID systems. The total integrated luminosity of good quality data used in this analysis is 4.9 fb$^{-1}$ for the 2011 sample and 20 fb$^{-1}$ for 2012. The average number of reconstructed primary vertices (PV) per event, related to multiple proton–proton interactions, is 6.2 and 11.4 in the two years respectively.

Samples of simulated MC events are used for training and validation of the multivariate analyses, for the determination of the efficiency ratios, and for guiding the signal extraction fits. Exclusive MC samples were produced for the signal channels $B_0^+ \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, the normalisation channel $B^+ \to J/\psi K^+$ ($J/\psi \to \mu^+\mu^-$), the $B^+ \to J/\psi \pi^+$ channel, and the control channel $B_0^0 \to J/\psi \phi$ ($\phi \to K^+K^-$). In addition, background studies employ MC samples of inclusive semileptonic decays $B \to \mu X$, samples of

---

$^1$ 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.
$B_s^0 \to K^-\mu^+\nu$, $B^0 \to \pi^-\mu^+\nu$, $\Lambda_b \to p\mu^-\overline{\nu}$, $B^0_{(s)} \to hh'$ decays with $h^{(*)}$ being a charged pion or kaon, and inclusive decays $B \to J/\psi X$.

Most of the muon pairs in the data sample originate from the uncorrelated decays of hadrons produced in the hadronisation of a $b$ and a $\bar{b}$ quarks. To describe this background, defined as combinatorial, a large MC sample was generated by selecting specific topologies that dominate it and producing them semi-exclusively. The strategy is to consider both the primary decays from $b$ quarks and the secondary decays from $c$ quarks. Independent samples of events with forced semileptonic decays or decays including muon pairs were generated in all combinations of the two types. The total number of events in each sample is chosen to reproduce the composition of oppositely charged muon pairs representative of our data.

The MC samples were generated with Pythia 6 [16] for studies related to data collected in 2011, and with Pythia 8 [17] and EvtGen [18] for the 2012 sample and the development of multivariate classifiers. The ATLAS detector and its response are simulated using Geant4 [19, 20]. Additional $pp$ interactions in the same and nearby bunch crossings (pile-up) are included in the simulation. All simulated samples are reweighted to have the same distribution of the number of PVs per bunch crossing found in data.

Using the iterative reweighting method described in Ref. [14], the simulated samples of the exclusive decays considered are adjusted with two-dimensional data-driven weights (DDW) to correct for the differences between MC and data observed in the $p_T^B$ and $|\eta^B|$ distributions. DDW obtained from $B^+ \to J/\psi K^+$ decays are used to correct the simulation samples in the signal and normalisation channels. DDW obtained from the $B^0_s \to J/\psi\phi$ control channel are found to agree with those from $B^+ \to J/\psi K^+$ showing the consistency of the corrections.

Similarly to the exclusive decays, the large continuum background MC sample is reweighted via DDW obtained from its comparison with the data in the sidebands of the signal region.

### 4 Data selection

For data collected during the LHC Run 1, the ATLAS detector used a three-level trigger system, consisting of a hardware-based Level-1 trigger, software-based Level-2 and Event Filter triggers.

A dimuon trigger [21, 22] is used to select events. The 2011 data sample contains events seeded by a Level-1 dimuon trigger that required a transverse momentum $p_T > 4$ GeV for both muon candidates. Due to the increased pile-up in 2012 data, this dimuon trigger was prescaled at the beginning of every fill. The effect of prescaling is mitigated by including in the analysis events selected by two additional Level-1 triggers scarcely affected by prescaling, where tighter selections were applied: $p_T > 6$ GeV or $|\eta| < 1.05$ for one of the muons. A full track reconstruction of the muon candidates was performed at the software trigger levels, where an additional loose selection was applied to the dimuon invariant mass $m_{\mu\mu}$ and the events were assigned to the $J/\psi$ stream ($2.5 < m_{\mu\mu} < 4.3$ GeV) or to the $B$ stream ($4.0 < m_{\mu\mu} < 8.5$ GeV).

Events from the 2012 dataset are divided into three mutually exclusive trigger categories and they are assigned to the tightest trigger they passed, in case they were selected in more than one trigger stream.

We define the following 2012 trigger categories:

- $T_1$: “higher threshold” trigger with $p_T > 6$ GeV and $> 4$ GeV for the two muons;
- $T_2$: “barrel” trigger with $p_T > 4$ GeV for both muon candidates and at least one of them with $|\eta| < 1.05$ (and $T_1$ requirement not satisfied);
$T_3$: basic dimuon trigger with $p_T > 4$ GeV for both muon candidates (and $T_1$, $T_2$ requirements not satisfied).

Events belonging to a given category are all associated with the same pattern of Level-1 prescaling. The event sample in the $T_2$ ($T_3$) category has an equivalent integrated luminosity equal to 97.7\% (81.3\%) of the luminosity of the $T_1$ category. The impact of the trigger Level-1 prescale on the total sample of collected events is minor, since the majority of the events belong to the $T_1$ category.

The events in the reference channels $B^+ \rightarrow J/\psi K^+$ and $B^0_s \rightarrow J/\psi \phi$ collected in 2012 and used in this analysis belong to a prescaled sample of events, which was processed together with the signal events. The effective prescaling factor is equal to 7.3, and does not affect the sensitivity of this analysis, given the large number of available events in the normalisation channel. This factor is included in the $\alpha_k$ parameters in Eq. (1).

A fourth category is defined for events from the 2011 dataset. They were collected with a trigger requirement $p_T > 4$ GeV for both muon candidates, and prescaling was not applied to this sample.

After off-line reconstruction, a preliminary selection is performed on candidates for $B^0 \rightarrow \mu^+\mu^-$, $B^+ \rightarrow J/\psi K^+ \rightarrow \mu^+\mu^- K^+$ and $B^0_s \rightarrow J/\psi \phi \rightarrow \mu^+\mu^- K^+K^-$. In the ID system, muons are required to have at least one hit in the pixel detector, five hits in the semiconductor tracker (two hits per each double-sided layer), and six hits in the transition-radiation tracker, if $0.1 < |\eta| < 1.9$. They are also required to be reconstructed in the MS, and to have $|\eta| < 2.5$ and $p_T > 4$ GeV. Kaon candidates have to satisfy similar requirements in the ID, except that at least nine instead of six hits are required in the transition-radiation tracker and a looser requirement of $p_T > 1$ GeV is imposed.

$B$ meson properties are computed based on a decay vertex fitted to two, three or four tracks, depending on the decay process to be reconstructed. The $\chi^2$ per degree of freedom in the vertex fit is required to be less than six for the $B$ vertex, and less than ten for the $J/\psi \rightarrow \mu\mu$ vertex. The conditions $2915 < m(\mu\mu) < 3275$ MeV and $1005 < m(KK) < 1035$ MeV are required on ID track combinations for the $J/\psi \rightarrow \mu\mu$ and the $\phi \rightarrow KK$ vertices, respectively. In the $B^+ \rightarrow J/\psi K^+$ and $B^0_s \rightarrow J/\psi \phi$ fits the reconstructed $J/\psi$ mass is constrained to the world average value [23].

Reconstructed $B$ candidates are required to satisfy $p_T^B > 8.0$ GeV and $|p_T^B| < 2.5$. The dimuon invariant mass for $B^0 \rightarrow \mu^+\mu^-$ candidates is calculated using the combined ID and MS information, in order to improve the mass resolution in the end-caps with respect to using ID information only [24].

The dimuon mass range considered for the $B^0 \rightarrow \mu^+\mu^-$ decay is 4766–5966 MeV in which the 5166–5526 MeV range is defined as the signal region while the low-mass and high-mass regions (4766–5166 and 5526–5966 MeV) are the signal mass sidebands. For the reference channels, the mass range considered is 4930–5630 (5050–5650) MeV for $B^+ \rightarrow J/\psi K^+$ ($B^0_s \rightarrow J/\psi \phi$) in which the 5180–5380 (5297–5437) MeV range is the peak region and the two low and high mass ranges are the mass sidebands used for background subtraction.

The coordinates of the PVs are obtained from charged tracks not used in the decay vertices, and are transversely constrained to the luminous region of the colliding beams. The matching of a $B$ candidate to a PV is made by projecting the direction of the candidate to the point of closest approach to the collision axis, and choosing the PV with the smallest separation along $z$. Simulation shows that this method achieves a correct matching probability of better than 99\%.

To reduce of the large background in the $B^0 \rightarrow \mu^+\mu^-$ channel before the final selection based on multivariate classifiers, a loose collinearity requirement is applied between the momentum of the $B$ candidate
\((p^B)\) and the spatial separation between the PV and the decay vertex (\(\Delta x\)). The absolute value of the difference in azimuthal angle \(\Delta \phi_{2D}\) is required to be smaller than 1.0 rad. Using the difference in rapidity \(\Delta \eta\), the combination \(\Delta R = \sqrt{\Delta \phi_{2D}^2 + \Delta \eta^2}\) is required to be smaller than 1.5. These requirements reduce the background by a factor of 0.4, with a signal efficiency of 95%.

After the preliminary selection, approximately \(2.6 \times 10^6 B^0_{(s)} \rightarrow \mu^+\mu^-\) and \(2.3 \times 10^6 B^+ \rightarrow J/\psi K^+\) candidates are found in the signal regions.

5 Background composition

The background to the \(B^0_{(s)} \rightarrow \mu^+\mu^-\) signal can be separated into three categories:

- **Combinatorial** background, the dominant component due to pairs of uncorrelated muons and characterised by a small dependence on the dimuon invariant mass;
- **Partially reconstructed** \(B \rightarrow \mu\mu X\) decays, characterised by non-reconstructed final-state particles (X) and thus accumulating in the low-mass sideband;
- **Peaking** background, due to \(B^0_{(s)} \rightarrow hh'\) decays, with both hadrons misidentified as muons.

The combinatorial background consists mainly of muons independently produced in the fragmentation and decay trees of a \(b\) and a \(\bar{b}\) quark (opposite-side muons). It is studied in the signal mass sidebands, and it is found to be correctly described by the inclusive MC sample of combinatorial background from semileptonic decays of \(b\) and \(c\) hadrons. Section 8 contains data–MC comparisons for the combinatorial background. As discussed in Section 7, a multivariate classifier trained on MC samples is used to reduce this component.

The partially reconstructed decays consist of several topologies: (a) same-side (SS) combinatorial background from decay cascades \((b \rightarrow c\mu\nu \rightarrow s(d)\mu\mu\nu\nu\); (b) same-vertex (SV) background from \(B\) decays containing a muon pair (e.g. \(B^0 \rightarrow K^{*0}\mu\mu, B \rightarrow J/\psi X \rightarrow \mu\mu X'\)); (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.

Inclusive MC samples of SS events, SV events, and \(B_c \rightarrow J/\psi\mu\nu\) decays were generated together with the large sample of combinatorial background. All subsamples have a dimuon mass distribution accumulating below the mass range considered in this analysis. The high-mass tail extends to the signal region and becomes a significant fraction of the background only after applying a selection against the combinatorial background.

The semileptonic decays with final-state hadrons misidentified as muons consist mainly of three-body charmless decays \(B^0 \rightarrow \pi\mu\nu, B^0_c \rightarrow K\mu\nu\) and \(\Lambda_c \rightarrow p\mu\nu\) in which the tail of the mass distribution extends to the signal region. Due to branching fractions of the order of \(10^{-6}\), this background is not large, and is further reduced by the dedicated muon identification requirements, discussed in Section 6. The MC mass distributions of these partially reconstructed decay topologies are shown in Figure 1(a) after applying the preliminary selection criteria described in Section 4.

Finally, the peaking background is due to \(B^0_{(s)}\) decays containing two hadrons misidentified as muons, which populate the signal region and cannot be observed in the sidebands. This is shown in Figure 1(b).
6 Multivariate analysis for reducing hadron misidentification

In the preliminary selection, muon candidates are formed from the combination of tracks reconstructed independently in the ID and MS [25]. The performance of the muon reconstruction in ATLAS is presented in Ref. [24]. Additional studies were performed for this analysis to minimise and evaluate the amount of background related to hadrons erroneously identified as muons.

Detailed simulation studies were performed for the channels $B^0 \rightarrow hh'$ and $\Lambda_b \rightarrow ph$, with $h^{(*)} = \pi^\pm, K^\pm$. A full Geant4-based simulation [19] in all systems of the ATLAS detector is used for this purpose. 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 meson. Hence, despite the notation of fake muons, this background is generally related to true muons measured in the MS, but not produced promptly in the decay of a $B$ meson. The contribution from hadronic punch-through into the MS is expected from simulation to amount only to 3% (8%) of the total number of fake candidates from kaons (pions).

The simulation shows that after the preliminary selection the probability for a kaon (pion) to be misidentified as a muon is 0.4% (0.2%). This fraction is found to be largely independent of the transverse momentum and rapidity of the track, as well as other variables related to the underlying event or pile-up. The misidentification rate for protons is found to be negligible (< 0.01%).

The muon candidate is further required to match the trigger requirements, resulting in a reduction in the number of retained tracks by a factor 0.58, and to pass an additional multivariate selection, implemented as a boosted decision tree (BDT) [26]. This selection, referred to as fake-BDT, is based on variables
Table 1: Description of the eight variables used in the discrimination between signal muons and those from hadron decays in flight and punch-throughs.

1. Absolute value of the track rapidity measured in the ID.
2. Ratio $q/p$ (charge over momentum) measured in the MS.
3. Scattering curvature significance: maximum variation of the track curvature between adjacent layers of the ID.
4. $\chi^2$ of the track reconstruction in the MS.
5. Number of hits used to reconstruct the track in the MS.
6. Ratio of the values of $q/p$ measured in the ID and in the MS, corrected for the average energy loss in the calorimeter.
7. $\chi^2$ of the match between the tracks reconstructed in the ID and MS.
8. Energy deposited in the calorimeters along the muon trajectory obtained by combining ID and MS tracks.

The background due to $B^0_s \rightarrow hh'$, with double misidentification $hh' \rightarrow \mu\mu$, has a distribution in the reconstructed mass peaking at 5250 MeV, close to the $B^0_s$ mass and is effectively indistinguishable from the $B^0$ signal (see Figure 1(b)). Beyond the muon and fake-BDT selection, these events have the same acceptance and selection efficiency as the $B^0 \rightarrow \mu^+\mu^-$ signal. Therefore, the expected number of peaking-background events can be estimated from the number of observed $B^+ \rightarrow J/\psi K^+$ events, in a way analogous to what is done for the signal, using Eq. (1). World average [23] 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. The resulting total expected number of peaking-background events, after the final selection (including a multivariate cut against $\mu^+\mu^-$ continuum background, the continuum-BDT discussed in Section 7), is equal to 0.7, with a 10% uncertainty from the normalisation procedure.

The simulation of hadron misidentification was validated and calibrated with studies performed on data. The fractions of fake muons after the preliminary selection were evaluated on samples of $\phi \rightarrow K^+K^-$ and $B^+ \rightarrow J/\psi K^+$ events, and found to be consistent with the simulation within a factor $1.2 \pm 0.2$. This factor and its square $1.4 \pm 0.5$ are used as scale correction and systematic uncertainty in the single and double misidentification probability, respectively. Hence, the expected number of peaking background events is equal to $1.0 \pm 0.4$.

A further test of the peaking background was performed on the final sample of $B^0_{(s)} \rightarrow \mu^+\mu^-$ candidates. Inverting the selection applied with the fake-BDT, the number of events containing real muons is largely reduced, while the number of peaking-background events is approximately three times larger than in the sample obtained with the nominal selection. The result of a fit to the background-enhanced sample is $0.5 \pm 3.0$ events of peaking background, in good agreement with the expectation.
The efficiency of the fake-BDT selection when applied to muons from $B^0 \rightarrow J/\psi K^+$ candidates selected in data was found to be accurate to better than 1%.

Besides the peaking background, the selection with the fake-BDT also reduces the semileptonic contributions with a single misidentified hadron. The expected number of events from $B^0 \rightarrow \pi \mu \nu$ and $B^0_s \rightarrow K \mu \nu$ in the final sample is $107 \pm 27$. The $\Lambda_b \rightarrow p \mu \nu$ contribution is negligible due to the smaller production cross section and the fake rejection for protons at the level of $10^{-5}$.

7 Multivariate analysis for continuum discrimination

A multivariate analysis, implemented as a BDT, is employed to enhance the signal relative to the combinatorial background. This classifier, referred to as the continuum-BDT, is based on the 15 variables described in Table 2. 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 separation between production and decay vertices $\Delta x$; (b) variables describing the muons forming the $B$ meson candidate; and (c) variables related to the rest of the event. The selection of the variables aims to optimise the discrimination power of the classifier, while minimising the dependence on the invariant mass of the muon pair.

Most of the discriminating variables are part of the set used in the previous analysis based on data collected in 2011 [14], while others were modified or added, exploiting the statistical power of the large samples of MC events used for training and validating the classifier. To minimise the dependence of the classifier on the effects of the pile-up, requirements of compatibility with the same vertex matched to the dimuon candidate are placed on the additional tracks considered for the variables $I_0, DOCA_{strik}$ and $N_{close}$.

The correlation between the discriminating variables was studied in the MC samples for signal and combinatorial background discussed in Section 3, and on data from the sidebands of the $\mu^+\mu^-$ mass distribution. Different degrees of correlation are present, with significant linear correlation among the variables $\chi^2_{PV,DV,ag}, L_{xy,1}^{\text{max},-\text{sig}.}, |d_0|^{\text{max},-\text{sig}.}$, and $\chi^2_{\mu,xyP\nu}$. Conversely, the variables $IP_{BD}^3$, DOCA$_{\mu\mu}$ and $I_{0,7}$ have negligible correlation with any of the others used in the classifier.

The MC sample for signal and the large MC sample of semileptonic decays of hadrons containing $b$ or $c$ quarks are used for training and testing the classifier. As discussed in Section 3, signal and background samples are reweighted according to the distributions of $p_T$ and $|\eta|$ of the dimuon and of the number of reconstructed PVs observed in data. To correctly reproduce the data collected in 2012, the MC events belonging to different trigger streams are reweighted according to the relative equivalent luminosity and to two different versions of the Level-2 muon reconstruction algorithm used during the data taking. The BDT training is done using TMVA [26].

Figure 2 shows the distribution of the BDT output variable for signal and background, separately for combinatorial background and partially reconstructed events. Also shown is the BDT distribution for dimuon candidates from data, from the sidebands of the mass distribution. With the normalisation observed in the data before the application of the BDT selection, the combinatorial background dominates over the other sources by a factor of about 2000 in the number of events contained in the full mass window and full range of the continuum-BDT output. In both the signal and background MC samples, the absolute value of the linear correlation coefficient between the BDT output and the dimuon invariant mass is smaller than 1%.
When the BDT classifier is applied to $B^+ \rightarrow J/\psi K^+$ and $B_0^\text{s} \rightarrow J/\psi \phi$ candidates, the variables related to the decay products of the $B$ mesons refer only to the muons from the decay of the $J/\psi$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_\text{T}^B$</td>
<td>Magnitude of the $B$ candidate transverse momentum $p_\text{T}^B$.</td>
</tr>
<tr>
<td>$\chi^2_{\text{PV},\text{DV}}$</td>
<td>Significance of the separation $\Delta x$ between production (i.e. associated PV) and decay (DV) vertices in the transverse projection: $\Delta x = \sum_{\Delta}^{-1} \Delta x$, where $\Delta x$ is the covariance matrix.</td>
</tr>
<tr>
<td>$\Delta R$</td>
<td>Three-dimensional opening between $p_\text{T}^B$ and $\Delta x$: $\sqrt{\Delta R^2 + \Delta y^2}$.</td>
</tr>
<tr>
<td>$</td>
<td>\alpha_{2D}</td>
</tr>
<tr>
<td>$L_{\mu\mu}$</td>
<td>Projection of $\Delta x$ along the direction of $p_\text{T}^B$: $(\Delta x \cdot p_\text{T}^B)/</td>
</tr>
<tr>
<td>IP$^{3D}_B$</td>
<td>Three-dimensional impact parameter of the $B$ candidate to the associated PV.</td>
</tr>
<tr>
<td>DOCA$_{\mu\mu}$</td>
<td>Distance of closest approach (DOCA) of the two tracks forming the $B$ candidate (three-dimensional).</td>
</tr>
<tr>
<td>$\Delta \phi_{\mu\mu}$</td>
<td>Difference in azimuthal angle between the momenta of the two tracks forming the $B$ candidate.</td>
</tr>
<tr>
<td>$</td>
<td>d_0</td>
</tr>
<tr>
<td>$</td>
<td>d_0</td>
</tr>
<tr>
<td>$P_{\text{min}}^L$</td>
<td>Value of the smaller projection of the momenta of the muon candidates along $p_\text{T}^B$.</td>
</tr>
<tr>
<td>$I_{0.7}$</td>
<td>Isolation variable defined as ratio of $</td>
</tr>
<tr>
<td>DOCA$_{\text{xtrk}}$</td>
<td>DOCA of the closest additional track to the decay vertex of the $B$ candidate. Tracks matched to a PV different from the $B$ candidate are excluded.</td>
</tr>
<tr>
<td>$N_{\text{close}}^{\text{xtrk}}$</td>
<td>Number of additional tracks compatible with the decay vertex (DV) of the $B$ candidate with $\ln(\chi^2_{\text{xtrk,DV}}) &lt; 1$. The tracks matched to a PV different from the $B$ candidate are excluded.</td>
</tr>
<tr>
<td>$\chi^2_{\mu,\text{PV}}$</td>
<td>Minimum $\chi^2$ for the compatibility of a muon in the $B$ candidate with a PV different from the one associated with the $B$ candidate.</td>
</tr>
</tbody>
</table>

The final selection requires a continuum-BDT output value larger than 0.24, corresponding to a signal relative efficiency of 54% (see Section 11), and to a reduction of the combinatorial background by a factor of about $10^{-3}$. 

### 8 Data–MC comparisons

The distributions of the discriminating variables are used to compare the MC sample of semileptonic decays with data in the dimuon sidebands. Figure 3 shows the distributions for two discriminating variables. Agreement with the sideband data is fair and the discrepancies observed do not compromise the
Figure 2: Continuum-BDT distribution for the signal and background events: signal \( B^0 \), partially reconstructed \( B \) events (SS+SV), \( B_s \) decays and continuum. The solid histograms are obtained from MC simulation, while the points represent data collected in the sidebands. All distributions are normalised to unity. The distributions are shown after the preliminary selection, and before applying any reweighting to the variables used in the classifier.

use of this MC background sample for the purpose of training the continuum-BDT. The continuum MC simulation is not used for computation of efficiencies or normalisation purposes.

Figure 3: Data and continuum MC distributions of \( |\alpha_{2D}| \) (a) and \( \chi^2_{\mu,xPV} \) (b) variables (see Table 2). The dots correspond to the 2012 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. Discrepancies between MC events and sideband data like the one observed for \( \chi^2_{\mu,xPV} \) do not compromise significantly the optimisation of the continuum-BDT classifier.

The distributions of the discriminating variables are also used for the comparison of \( B^+ \rightarrow J/\psi K^+ \) and \( B_s^0 \rightarrow J/\psi \phi \) events between MC simulation and data. To perform such comparison, for each variable the contribution of the background is subtracted from the signal. For this purpose, a maximum-likelihood fit is
Figure 4: Data and MC distributions in $B^+ \rightarrow J/\psi K^+$ events for the discriminating variables: $|\alpha_{2D}|$ (a), $\chi^2_{PV, DV}$ (b) and $I_{0,7}$ (c). The variable $I_{0,7}$ is also shown for $B^0_s \rightarrow J/\psi \phi$ events (d). The black dots correspond to the sideband-subtracted data, while the red histogram corresponds to the MC distribution, normalised to the number of data events. Differences in shape between MC events and data are accounted for as systematic effects. The discrepancy shown for $I_{0,7}$ in the $B^+ \rightarrow J/\psi K^+$ channel is the most significant among all variables and both reference channels.

performed to the mass distribution, separately in the four trigger and data categories. For $B^+$, the signal is described by two overlying Gaussian distributions, an error function for the partially reconstructed decays and an exponential function for the continuum background. The fit model is simpler than the one used for the extraction of the $B^+$ signal used for normalisation after the final selection, described in Section 9, but it is sufficient for the purpose discussed here. For $B^0_s \rightarrow J/\psi \phi$, a Gaussian distribution is used for the signal and a third-order Chebychev polynomial for the background. For each discriminating variable, the background distribution observed in the sidebands is interpolated to the signal region, normalised according to the result of the likelihood fit, and subtracted from the distribution observed in the signal region.

Figure 4 shows examples of the distributions of the discriminating variables obtained from data and MC
simulation. In general, the overall shapes of distributions are in good agreement between data and MC events. Observed differences are accounted for as systematic effects with the procedure described in Section 10. The discrepancy shown for the isolation variable $I_{0.7}$ in the $B^+ \rightarrow J/\psi K^+$ channel is the most significant one among all variables and both reference channels.

9 Yield extraction for the normalisation channel $B^+ \rightarrow J/\psi K^+$

The $B^+$ yield for the normalisation channel is extracted with an unbinned extended maximum-likelihood fit to the $J/\psi K^+$ 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.

Figure 5: $J/\psi K^+$ invariant mass distribution for all $B^+$ candidates in the $T_1$ trigger category in 2012 data in linear (a) and logarithmic (b) scale. The result of the fit is overlaid. The various components of the spectrum are described in the text. The insets at the bottom of the plots show the bin-by-bin pulls for the fits, 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: $B^+ \rightarrow J/\psi K^+$ events, Cabibbo-suppressed $B^+ \rightarrow J/\psi \pi^+$ events on 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 combinatorial background composed mostly of $b\bar{b} \rightarrow J/\psi X$ events. The shape of the $B^+ \rightarrow J/\psi K^+$ distribution is parameterised using a Johnson $S_U$ function $^{[27, 28]}$ and a Gaussian function for the $T_1$, $T_2$ and 2011 categories, while a single Johnson $S_U$ function is used for the $T_3$ category. The final $B^+ \rightarrow J/\psi K^+$ yield includes the contribution from radiative decays. The $B^+ \rightarrow J/\psi \pi^+$ events are modelled by the sum of a Johnson $S_U$ and a Gaussian function, where all parameters are determined from the MC simulated data. The PRD are described with combinations of Fermi–Dirac and exponential functions, slightly different between the different categories in the low-mass region. Their shape parameters are determined from simulation. Finally, the combinatorial background is modelled with an exponential function with the shape parameter extracted from the fit. As an example,
the fit for the $T_1$ category is shown in Figure 5. The results of the fits in all data categories are shown in Table 3.

Table 3: Results of the fits to the events reconstructed as $B^+ \to J/\psi K^+$ in each trigger and data category. Uncertainties are statistical and systematic, respectively.

<table>
<thead>
<tr>
<th>Category</th>
<th>$N_{J/\psi K^+}$</th>
<th>$N_{J/\psi \pi^+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>$46,860 \pm 290 \pm 280$</td>
<td>$1,420 \pm 230 \pm 440$</td>
</tr>
<tr>
<td>$T_2$</td>
<td>$5,200 \pm 84 \pm 100$</td>
<td>$180 \pm 51 \pm 89$</td>
</tr>
<tr>
<td>$T_3$</td>
<td>$2,512 \pm 91 \pm 42$</td>
<td>$85 \pm 77 \pm 30$</td>
</tr>
<tr>
<td>2011</td>
<td>$95,900 \pm 420 \pm 1,100$</td>
<td>$3,000 \pm 340 \pm 1,140$</td>
</tr>
</tbody>
</table>

Some of the systematic effects are included automatically in the fit: the effect of limited MC sample size, for example, is included in the uncertainties through 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 MC 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, uncertainties in the relative fractions and shapes of PRD, and in the shape of the combinatorial background. In each case, the difference with respect 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 shape of the combinatorial background, the relative fractions of PRD and the signal charge asymmetry. The total statistical and systematic uncertainty in the $B^+$ normalisation yield amounts to 0.8%.

9.1 $J/\psi \pi^+/J/\psi K^+$ ratio measurement

For further validation of the fit to the $B^+ \to J/\psi K^+$ yield, the fit described in Section 9 is used to extract the yields for $B^+ \to J/\psi K^+$ and $B^+ \to J/\psi \pi^+$ decays and obtain the ratio $\rho_{\pi/K}$ of the corresponding branching fractions. The measurement is performed separately in the four categories, and combined into an error-weighted mean $\rho_{\pi/K}$. Table 3 shows the fitted yields.

Most systematic effects cancel in the measurement of this ratio. Residual systematic uncertainties in the ratio of the branching fractions come from the uncertainties in the $K^-/K^+$, $\pi^-/\pi^+$ and $K^+/\pi^+$ relative efficiencies. For each systematic effect the ratio is re-evaluated, therefore accounting for correlated effects. The largest systematic uncertainty in the measured ratio comes from the combinatorial background model parameterisation (23%), followed by the effect of the uncertainties in the PRD fraction estimates (15%). All other systematic sources have uncertainties at the level of 10% or less. The final result for the ratio of branching fractions is:

$$\rho_{\pi/K} = \frac{\mathcal{B}(B^+ \to J/\psi \pi^+)}{\mathcal{B}(B^+ \to J/\psi K^+)} = 0.035 \pm 0.003 \pm 0.012,$$

where the first error is statistical and the second is systematic. The result is in agreement with the most accurate available results from LHCb ($0.0383 \pm 0.0011 \pm 0.0007$ [29]) and BABAR ($0.0537 \pm 0.0045 \pm 0.0011$ [30]).
10 Evaluation of the $B^0(\pi) \to \mu^+\mu^-$ to $B^+ \to J/\psi K^+$ efficiency ratio

The ratio of efficiencies for $B^+ \to J/\psi K^+$ and $B^0(\pi) \to \mu^+\mu^-$ enters the $D_{\text{norm}}$ term defined in Eq. (2). Both channels are measured in the fiducial volume of the $B$ meson defined as $p_T^B > 8.0$ GeV and $|\eta_B| < 2.5$.

The efficiencies within the fiducial volume include both acceptance and selection efficiency. The acceptance is defined by the selection placed on the particles in the final state: $p_T^\mu > 4.0$ GeV and $|\eta_\mu| < 2.5$ for muons, $p_T^K > 1.0$ GeV and $|\eta_K| < 2.5$ for kaons. In addition to the reweighting of the distributions of $p_T^B$, $|p^B|$ and the number of reconstructed PVs observed in data, the MC samples are reweighted according to the equivalent integrated luminosity associated with each trigger category and the Level-2 muon trigger algorithms used in 2012.

The selection efficiency includes trigger, reconstruction, and signal selection. The trigger efficiencies are taken from a data-driven study based on the comparison of single-muon and dimuon triggers for events containing muon pairs from the decays of $J/\psi$ and $\Upsilon$ resonances [31]. The signal selection requires the output of the continuum-BDT to be larger than 0.24.

All efficiency terms are computed separately for the three trigger selections used in 2012 and for the 2011 sample. Table 4 provides the values of the efficiency ratios $R^*_k$, for each of the categories ($k = 1 - 4$), together with the statistical and systematic uncertainties described below.

The efficiency ratios shown in Table 4 are computed using the mean lifetime of $B^0$ [23, 32] in the MC generator. The same efficiency ratios apply to the $B^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays, within the MC statistical uncertainty of ±0.5%.

Table 5 provides a summary of the systematic uncertainties in $D_{\text{norm}}$. Each contribution is discussed below.

The statistical uncertainties in the efficiency ratios come from the finite number of events available for the simulated samples. The systematic uncertainty affecting $R^*_k$ comes from four sources. A first contribution is due to the uncertainties in the DDW and in the trigger efficiencies. This term is assessed from pseudo-MC studies, performed by varying the corrections within their statistical uncertainties. The RMS value of the distribution of $R^*_k$ obtained from pseudo-MC samples is taken as the systematic uncertainty. The uncertainties range from ±1% to ±5% depending on the category considered, with a combined effect of ±1.5% for the data collected in 2012, and ±2.2% for 2011.

Table 4: Values of the efficiency ratios $R^*_k$ for the 2012 trigger categories and the 2011 sample, and their relative contributions to $D_{\text{norm}}$ (Eq. (2)). The first uncertainty is statistical and the second systematic. The systematic component includes the uncertainties from the MC reweighting and from data–MC discrepancies, as described in the text. The correction due to the $B^0$ effective lifetime value discussed in the text is not applied to the numbers shown.

<table>
<thead>
<tr>
<th>Data category ($k$)</th>
<th>$R^*<em>k = (e</em>{J/\psi K^+}/e_{\mu^+\mu^-})_k$</th>
<th>Relative contribution to $D_{\text{norm}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>$0.180 \pm 0.001 \pm 0.009$</td>
<td>68.3%</td>
</tr>
<tr>
<td>$T_2$</td>
<td>$0.226 \pm 0.004 \pm 0.014$</td>
<td>6.0%</td>
</tr>
<tr>
<td>$T_3$</td>
<td>$0.189 \pm 0.005 \pm 0.022$</td>
<td>3.5%</td>
</tr>
<tr>
<td>2011</td>
<td>$0.156 \pm 0.002 \pm 0.009$</td>
<td>22.2%</td>
</tr>
</tbody>
</table>
The difference between the efficiency ratio value obtained after each reweighting and the value without reweighting is taken as the systematic uncertainty due to mis-modelling of that variable. A comparison of data and MC simulation based on the $B^+ \rightarrow J/\psi K^+$ channel is preferred to the one with $B^0_s \rightarrow J/\psi \phi$, since the latter is generally found to be statistically consistent but less precise. The isolation variable $I_{0.7}$ is computed on charged-particle tracks only, and differences between $B^+$ and $B^0_s$ are expected and were observed in previous studies [14]. For this variable, the reweighting procedure is based on $B^+$ data for the corresponding MC sample, and on $B^0_s \rightarrow J/\psi \phi$ data for the $B^0_s \rightarrow \mu^+ \mu^-$ sample. The main discrepancy for the 2012 MC samples is observed in the isolation variable $I_{0.7}$, both in the normalisation channel ($-8.9\%$ in $\varepsilon_{J/\psi K^+}$ with the weights extracted from the $B^+ \rightarrow J/\psi K^+$ channel) and in the signal ($-3.8\%$ in $\varepsilon_{\mu^+ \mu^-}$ reweighting with $B^0_s \rightarrow J/\psi \phi$), with an effect of $-5.3\%$ on the efficiency ratio. The absolute values of the shifts in the ratio observed for other variables are smaller than $2\%$, with absolute shifts in the efficiencies of either channel smaller than $5\%$. For the 2011 sample a similar pattern is observed, but with smaller amplitudes: $3.2\%$ shift for $I_{0.7}$ and less than $1\%$ in any other variable. Because of the relatively large mis-modelling of $I_{0.7}$, the 2012 MC samples obtained after reweighting are taken as a reference, thus correcting the central value of the efficiency ratio. A systematic error of $\pm 3.2\%$, derived from the uncertainty in the reweighting procedure, is taken as the contribution of $I_{0.7}$ to the uncertainty in $R_{B^+}$. The correlation of the variable $I_{0.7}$ with the other discriminating variables is found to be negligible in the MC samples and in $B^+$ data. For all other variables, the shifts observed with the reweighting procedure are taken as contributions to the systematic uncertainty. In each category, the total uncertainty is the sum in quadrature of the uncertainties over all discriminating variables, and is the dominant contribution to the systematic uncertainties shown in Table 4.

A third source of systematic uncertainty arises from differences between the $B^0_s \rightarrow \mu^+ \mu^-$ and the $B^+ \rightarrow J/\psi K^+$ channels related to the reconstruction efficiency of the kaon track and of the $B^+$ decay vertex. The corresponding systematic uncertainty is $\pm 3.6\%$.

Finally, a fourth type of systematic uncertainty is related to residual uncertainties in the trigger efficiencies, which are evaluated through data-driven studies performed using $J/\psi K^+$ and $\mu^+ \mu^-$ candidates. The uncertainties cancel to a large extent in the comparison of the signal and normalisation channels, and the residual systematic uncertainty, averaged across the categories, is $\pm 1.5\%$.

The efficiency ratios enter in Eq. (1) with the $D_{\text{norm}}$ term defined in Eq. (2). For each category $k$, the
efficiency ratio is multiplied by the number of observed \( B^+ \) candidates and the trigger prescaling factor. The relative contributions of the \( T_1, T_2, T_3 \) and 2011 categories are shown in Table 4. The uncertainties in \( R^0_c \) are weighted accordingly and combined. For the trigger categories of the 2012 data sample, the correlations among the uncertainties due to DDW, trigger efficiency and mis-modelling of the discriminating variables are taken into account. No significant correlation is found between the 2011 and 2012 samples in the mis-modelling of the discriminating variables. The total uncertainty is equal to \( \pm 5.9\% \), including the relatively small uncertainty in the yield of \( B^+ \rightarrow J/\psi K^+ \). For comparison, the combined effect on \( D_{\text{norm}} \) of all corrections applied to the simulated data is equal to \( \pm 3.2\% \), with the contributions from DDW, data-driven trigger efficiencies and \( l_{0.7} \) reweighting equal to \( \pm 3.6\% \), \( \pm 3.9\% \) and \( \pm 4.1\% \), respectively.

A correction to the efficiency ratio for \( B^0_s \rightarrow \mu^+\mu^- \) is expected because of the width difference \( \Delta \Gamma_s \) between the \( B^0_s \) eigenstates. According to the SM, the decay \( B^0_s \rightarrow \mu^+\mu^- \) proceeds predominantly through the heavy state \( B_{s,H} \) [33], which has width \( \Gamma_{s,H} = \Gamma_s - \Delta \Gamma_s / 2 \), i.e. \( (6.2 \pm 0.5)\% \) smaller than the average \( \Gamma_s \) [32]. The variation in the value of the \( B^0_s \rightarrow \mu^+\mu^- \) mean lifetime was tested with MC simulation, and found to change the \( B^0_s \) efficiency by \( \pm 4\% \), and consequently the \( B^0_s \rightarrow \mu^+\mu^- \) efficiency ratio. This correction is applied to the central value of \( D_{\text{norm}} \) used in Section 12 for the determination of \( 2(B(B^0_s \rightarrow \mu^+\mu^-)) \). Due to the small value of \( \Delta \Gamma_s \), no correction needs to be applied to the \( B^0 \rightarrow \mu^+\mu^- \) decay.

### 10.1 Comparison of normalisation yields with other measurements

The systematic errors in acceptance and efficiency are minimised by using \( B^+ \rightarrow J/\psi K^+ \) as the normalisation channel and evaluating only ratios of efficiencies. However, event counts and absolute efficiency values for the reference channels can be used to extract the production cross sections for the purposes of comparisons with other measurements. The relevance of such comparisons is mainly in the validation of the selection efficiency in the analysis presented here, in particular for the final selection performed with the continuum-BDT classifier.

The yield of \( B^+ \) can be compared to the measurement performed by ATLAS with 2.7 fb\(^{-1}\) of data collected at \( \sqrt{s} = 7 \) TeV [34], and based on the same decay channel. In the comparison, the data collected at \( \sqrt{s} = 8 \) TeV for the present analysis were restricted to the phase space \( p_T > 9.0 \) GeV and \( |\eta| < 2.25 \) used for the previous result. Trigger and preliminary selections are very similar, but the selections against continuum background and fake muons are used only in the present analysis. The difference in the collision energy is taken into account by comparing the measured production cross section to the prediction based on the fixed-order next-to-leading-log (FONLL) approximation [35]. The theoretical uncertainty in the extrapolation from 7 to 8 TeV is expected to be small compared to experimental uncertainties. The ratio of the observed to the predicted cross section was measured in Ref. [34] as \( 1.24 \pm 0.04 \pm 0.09 \), where only measurement uncertainties, respectively statistical and systematic, are included in the error. The corresponding value from the present analysis is \( 1.17 \pm 0.02 \pm 0.14 \), with the uncertainty dominated by the systematic error in the efficiency of the continuum-BDT selection. The result is in agreement with the previous measurement. Correlated systematic errors between the two analyses amount to \( \pm 0.05 \).

The measurements of \( B^0_s \rightarrow J/\psi \phi \) and \( B^+ \rightarrow J/\psi K^+ \), together with the corresponding acceptance and efficiency values, can be used to extract the production ratio \( B^0_s/B^+ \), for \( 10 \leq p_T^B \leq 20 \) GeV and \( |\eta| < 2.5 \), in \( pp \) collisions at \( \sqrt{s} = 8 \) TeV. Using world averages values [23] for the branching fractions to the final states, the resulting mean ratio of the hadronisation fractions \( f_s/f_u \) is equal to \( 0.236 \pm 0.014 \pm 0.018 \pm 0.021 \), where the first error is statistical, the second is the systematic uncertainty in the efficiency ratio.
and the third is the uncertainty in the branching fractions. The ratio is uniform across the kinematic range observed, and it varies by only $-2\%$ if the $B^0_s$ and $B^+$ signals are extracted without applying the continuum-BDT selection. The normalisation procedure might not be free of bias, since the value of $\mathcal{B}(B^0_s \rightarrow J/\psi \phi)$ includes assumptions about $f_s$, and updating the assumptions may change it by about 5%. The result nevertheless provides a satisfactory consistency check with the available measurements [36, 37]. The most direct comparison is with the recent value $f_s/f_d = 0.240 \pm 0.020$ [37], obtained by ATLAS from the analysis of 2.7 fb$^{-1}$ of data collected at $\sqrt{s} = 7$ TeV, and performed over the same $p_T$ and $\eta^B$ ranges used in this analysis. The uncertainty in that measurement is dominated by the prediction of the ratio of branching fractions $\mathcal{B}(B^0_s \rightarrow J/\psi \phi) / \mathcal{B}(B^0 \rightarrow J/\psi K^{*0})$. The ratio of the efficiency-corrected event yields observed at $\sqrt{s} = 8$ TeV in the present analysis can be compared to the corresponding value from Ref. [37] after rescaling by the ratio of branching fractions $\mathcal{B}(B^* \rightarrow J/\psi K^*) / [\mathcal{B}(B^0 \rightarrow J/\psi K^{*0}) \times \mathcal{B}(K^{*0} \rightarrow K^+\pi^-)]$, which is known with better accuracy than $\mathcal{B}(B^0_s \rightarrow J/\psi \phi)$. In this way, some systematic uncertainties are removed, and the ratio of the two results is 0.96 $\pm$ 0.12. The largest contribution to the systematic uncertainty is from the efficiency of the continuum-BDT selection used in the present analysis.

In conclusion, the observed event rates for the normalisation channels are in agreement with previous measurements within uncertainties of about 12%.

11 Extraction of the signal yield

Dimuon candidates passing the preliminary selection and the multivariate selections against hadron misidentification and continuum background are classified according to three intervals in the continuum-BDT output: 0.240–0.346, 0.346–0.446 and 0.446–1. Each interval corresponds to an equal efficiency of 18% for signal events, and they are ordered according to increasing signal-to-background ratio. In each continuum-BDT interval, events from the four trigger and data categories are merged.

An unbinned extended maximum-likelihood fit is performed on the dimuon invariant mass distribution simultaneously across the three continuum-BDT intervals. The result of the fit is the total yield of $B^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ events in the three BDT intervals. The parameters describing the background are allowed to vary freely and are determined by the fit. The fit model for signal and background is described in Section 11.1. The systematic uncertainties related to the BDT intervals, to the signal and to the background model are discussed in Sections 11.1 and 11.2, and are included in the likelihood with Gaussian multiplicative factors with width equal to the systematic uncertainty.

11.1 Signal and background model

The model for describing signal and background is based on simulation and on data collected in the mass sidebands of the search region.

The mass distribution of the $B^0 \rightarrow \mu^+\mu^-$ signal is described by a superposition of two Gaussian distributions, both centred at the $B^0$ or $B^0_s$ mass. The parameters are extracted from MC simulation, and they are taken to be uncorrelated with the BDT output. Systematic uncertainties in the mass scale and resolutions are considered separately. Figure 6 shows the mass distributions for $B^0$ and $B^0_s$, obtained from MC events and normalised to the SM expectations.
Figure 6: Dimuon invariant mass distribution for the $B^0_s$ and $B^0$ signals from MC simulation. The double Gaussian fits are overlaid. The two distributions are normalised to the SM prediction for the expected yield with an integrated luminosity of 25 fb$^{-1}$.

The efficiency of the three intervals in the continuum-BDT output for $B^0_{(s)} \to \mu^+\mu^-$ events is calibrated with studies performed on the reference channels. The distribution of the BDT output is compared between MC and background-subtracted data. The differences observed in the ratio of data over simulation are described with a linear dependence on the BDT output. The slopes are equal within ±12% between $B^+ \to \phi K^+$ and $B^0 \to \phi$ and the mean value is used to reweight the BDT-output distribution in the $B^0_{(s)} \to \mu^+\mu^-$ MC sample. The corresponding variation in the efficiencies are equal to +0.018 and −0.018 respectively in the first and third BDT intervals. The values of the lower edge of the second and third BDT intervals are corrected in MC simulation to obtain equal efficiencies of 0.180 in each interval.

The systematic uncertainties in the efficiency of the BDT intervals are obtained with a procedure similar to the one used for the event selection (Section 10). 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 total systematic uncertainty due to mis-modelling of that variable. In each BDT interval, the sum in quadrature of the variations of all discriminating variables is found to be similar in the $B^+ \to \phi K^+$ and $B^0 \to \phi$ channels, and the average of the two is taken as the total systematic uncertainty in the efficiency. Values of ±0.026, ±0.010 and ±0.023 are found respectively in the first, second and third interval. Gaussian terms are included in the likelihood in order to describe these uncertainties, taking care of constraining the sum of the efficiencies of the three intervals, since that uncertainty is already included in the selection efficiency.

Figure 7 shows the distribution of the continuum-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 the correction based on the BDT output and the systematic uncertainty discussed above. The reweighting on the $I_{0,7}$ variables, discussed in Section 10 for the evaluation of the efficiency of the final event selection (BDT output > 0.24), is not applied to the events shown in Figure 7, and in the evaluation of the relative efficiency of the intervals used for the extraction of the $B^0_{(s)} \to \mu^+\mu^-$ signal. Reweighting
the BDT output is preferred over reweighting $I_{0.7}$, because of correlations present between the discriminating variables after the final selection is applied.

Finally, for the $B_{s}^{0} \rightarrow \mu^{+} \mu^{-}$ signal, the lifetime difference between $B_{s,1}$ and $B_{s}^{0}$ requires further corrections to the efficiency of the BDT intervals of +0.003 and +0.018 respectively in the second and third interval.

Figure 7: Data and MC distributions of $B^{+} \rightarrow J/\psi K^{+}$ (a) and $B_{s}^{0} \rightarrow J/\psi \phi$ (b) for the continuum-BDT output and the MC distributions for $B_{s}^{0} \rightarrow \mu^{+} \mu^{-}$ (c). The MC samples are normalised to the number of data events. A linear correction has been applied to the MC distributions, equal for all channels, and a systematic uncertainty is assigned to the distribution of the $B_{s}^{0} \rightarrow \mu^{+} \mu^{-}$ MC sample, as discussed in the text and illustrated by the dashed line and the envelope shown in (c). The vertical dashed lines in (c) correspond to the boundaries of the continuum-BDT intervals used for the signal extraction.

The background is composed of the types of events described in Section 5: (a) the combinatorial background, which is strongly reduced by the selection progressively applied in the three intervals of continuum-BDT; (b) the background from partially reconstructed SS and SV events, which is present mainly in the low-mass sideband; (c) the peaking background.
The dependence of the combinatorial background on the dimuon mass is described with a first-order polynomial. In the simulation, the slope of the distribution is similar in the three continuum-BDT intervals. The correlation between continuum-BDT and dimuon mass is small, and similar between simulation and sideband data within large statistical uncertainties. Hence the slope of the mass dependence is described by independent parameters in the three intervals, subject to loose Gaussian constraints of uniformity within ±40% between the first and second interval, and ±80% between the first and the third. Such variations of slope are larger than those observed in simulation, and consistent with those determined from data. Deviations from these assumptions are discussed below in Section 11.2. The normalisation of the combinatorial background is also extracted independently in each BDT interval.

The SS+SV background has a mass distribution peaking below the low-mass sideband region. The mass dependence is derived from data in the low-mass sideband region, and described with an exponential function with equal shape in the three continuum-BDT intervals. The value of the shape parameter is extracted from the fit to data. The normalisation values are extracted independently in each interval.

The mass distribution of the peaking background is very similar to the $B^0$ signal, as shown in Figure 1(b). In the fit, this contribution is included with fixed mass shape and with a normalisation of 1.0 ± 0.4 events, as discussed in Section 6. This contribution is equally distributed among the three intervals of continuum-BDT.

The fitting procedure is tested with pseudo-MC experiments, as discussed below. The use of three intervals in the continuum-BDT output is found to optimise the performance of the likelihood fit, with all BDT intervals contributing to the determination of the background, while the second and in particular the third interval provide sensitivity to the signal yield.

### 11.2 Systematic uncertainties in the fit

Studies based on pseudo-MC 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 the pseudo-MC samples. The corresponding deviations in the average numbers $N_s$, $N_d$ of $B^0_s$ and $B^0$ events returned by the fit, run in the nominal configuration, are taken as systematic uncertainties. The amplitude of the variations in the generation of the pseudo-MC samples 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 three-body $B^0_s$ decays and to $B^0_s \rightarrow J/\psi \mu^+ \mu^-$), and in others from uncertainties determined from data in the sidebands and from simulation (shapes of the background components and their variation across the continuum-BDT intervals).

The pseudo-MC experiments were generated with the abundance of the background components extracted from data in the mass sidebands of the signal region. The signal was generated with different configurations, corresponding to the SM prediction, to smaller values of $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ and to smaller/larger values of $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$.

For all variations in the assumptions and all configurations of the signal amplitudes, the distributions of the differences between results and generated values, divided by the fit errors (pull distributions), are found to be correctly described by Gaussian functions with widths approximately equal to one and values of the mean smaller than 0.2 for $B^0_s$ and smaller than 0.4 for $B^0$. The distributions obtained from pseudo-MC samples generated according to the nominal fit model are used to evaluate fit biases. For $B^0_s$, the fit
bias is negligible. For $B^0$ the absolute value is smaller than 25% of the fit error, and it is included as an additional systematic uncertainty.

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. For $B^0_s$, the total systematic uncertainty is found to increase with the assumed size of the signal, with a dependence $\sigma_{\text{syst}}(N_s) = \sqrt{2^2 + (0.06 \times N_s)^2}$. The total systematic uncertainty for $B^0$ is approximately $\sigma_{\text{syst}}(N_d) = 3$. Most of the shifts observed have opposite sign for $N_s$ and $N_d$, resulting in a combined correlation coefficient in the systematic errors of $\rho_{\text{syst}} = -0.7$.

The fit to 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 combined Gaussian distribution parameterised by the values of $\sigma_{\text{syst}}(N_s)$, $\sigma_{\text{syst}}(N_d)$ and $\rho_{\text{syst}}$.

### 11.3 Results of the signal yield extraction

Including both the 2012 and 2011 data-taking periods, 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. The values $509 \pm 28$, $32 \pm 6$ and $4.8 \pm 1.9$ events are obtained respectively in the three intervals of continuum-BDT. For comparison, the total expected number of signal events according to the SM prediction is 41 and 5 respectively for $N_s$ and $N_d$, equally distributed among the three intervals.\(^2\)

Once the signal region is made accessible, a total of 1951 events in the full mass range of 4766–5966 MeV are used for the likelihood fit to signal and background. Without applying any boundary on the values of the fitted parameters, the values determined by the fit are $N_s = 16 \pm 12$ and $N_d = -11 \pm 9$, where the uncertainties correspond to likelihood variations of $-2 \Delta \ln(L) = 1$. The likelihood includes the systematic uncertainties discussed above, but statistical uncertainties largely dominate. The primary result of this analysis is obtained by applying the natural boundary of non-negative yields, for which the fit returns the values $N_s = 11$ and $N_d = 0$. The uncertainties in the result of the fit are discussed in Section 12, where the measured values of the branching fractions are presented.

Figure 8 shows the dimuon mass distributions in the three intervals of continuum-BDT, together with the projections of the likelihood fit.

For comparison, the value $N_d$ can be constrained according to the SM expectation for the ratio $\mathcal{B}(B^0 \to \mu^+\mu^-)/\mathcal{B}(B^0_s \to \mu^+\mu^-)$ \(^1\) multiplied by the ratio of the hadronisation probabilities $f_d/f_s$ \(^3\), rather than being extracted independently from the fit. In this case the value of $N_s$ changes by $-0.8$, while $N_d = N_s/8.3 \approx 1.2$.

### 12 Branching ratio extraction

The branching fractions for the decays $B^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are extracted from data using a profile-likelihood fit. The likelihood is obtained from the one used for $N_s$ and $N_d$ replacing the fit parameters with the corresponding branching fractions divided by normalisation terms in Eq. (1), and including Gaussian multiplicative factors for the normalisation uncertainties.

\(^2\) The values of the single-event sensitivity are discussed in Section 12.
Figure 8: Dimuon invariant mass distributions in the unblinded data, in the three intervals of continuum-BDT output. Superimposed is the result of the maximum-likelihood fit, obtained imposing the boundary of non-negative signal contributions. The total fit is shown as a black continuous line, the filled area corresponds to the observed signal component, the blue dashed line to the SS+SV background, and the green dashed line to the combinatorial background.

The normalisation terms include external inputs for the $B^+$ branching fraction and the relative hadronisation probability. The first is obtained from world averages [23] as the product of $\mathcal{B}(B^+ \to J/\psi K^+) = (1.027 \pm 0.031) \times 10^{-3}$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033)\%$. The second is equal to one for $B^0$, while for $B^0_s$ it is taken from the ATLAS measurement $f_s/f_d = 0.240 \pm 0.020$ [37], assuming $f_u/f_d = 1$ [32].

The efficiency- and luminosity-weighted number of events for the normalisation channel enters in Eq. (1) with the denominator $D_{\text{norm}}$ (Eq. (2)). The values $D_{\text{norm}} = (2.88 \pm 0.17) \times 10^6$ for $B^0_s$ and $(2.77 \pm 0.16) \times 10^6$ for $B^0$ are obtained using Tables 3 and 4 for each category, together with the combined uncertainty from Table 5, and including the +4% correction to the $B^0_s \to \mu^+\mu^-$ efficiency due to the lifetime difference between $B_{sH}$ and $B^0_s$. 
The combination of $B^+$ branching fraction, hadronisation probabilities and $\mathcal{D}_{\text{norm}}$, i.e. the single-event sensitivity, is equal to $(8.9 \pm 1.0) \times 10^{-11}$ for $B^+_s \rightarrow \mu^+\mu^-$ and $(2.21 \pm 0.15) \times 10^{-11}$ for $B^0 \rightarrow \mu^+\mu^-$. The values of the branching fractions that optimise the profile-likelihood within the constraint of non-negative values are $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) = 0.9 \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = 0$. That constraint is applied for all results discussed in this section if not otherwise stated.

A Neyman construction [38] is used to determine the 68.3% confidence interval for $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ with pseudo-MC experiments, obtaining:

$$\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) = (0.9^{+1.1}_{-0.8}) \times 10^{-9}.$$ 

The uncertainties include both the statistical and systematic contributions. The two components are separated by repeating the likelihood fit after setting all systematic uncertainties to zero. The statistical uncertainty is dominant, with the systematic uncertainty equal to $\pm 0.3 \times 10^{-9}$.

The observed significance of the $B^0_s \rightarrow \mu^+\mu^-$ signal is determined from pseudo-MC experiments, with a hypothesis test based on the likelihood ratio $-\ln[L(\text{no--signal})/L(\text{max})]$ [39], and is equal to 1.4 standard deviations. For this test, $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ is left free to be determined in the fit. The corresponding expected significance for the SM prediction is 3.1 standard deviations.

Pseudo-MC experiments are also used to evaluate the compatibility of the observation with the SM prediction. A hypothesis test based on $-\ln[L(\text{SM})/L(\text{max})]$ is performed for the simultaneous fit to $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$. The result is $p = 0.048 \pm 0.002$, corresponding to 2.0 standard deviations.

Figure 9 shows the contours in the plane of $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ drawn for values of $-2 \Delta \ln(L)$ equal to 2.3, 6.2 and 11.8, relative to the maximum of the likelihood, allowing negative values of the branching fractions. The maximum within the physical boundary is shown with error bars indicating the 68.3% interval for the value of $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$. Also shown are the corresponding contours obtained in the combination of the results of the CMS and LHCb experiments, and the prediction based on the SM.

Using the CL$_s$ method [40] implemented with pseudo-MC experiments, an upper limit is placed on the $B^0_s \rightarrow \mu^+\mu^-$ branching fraction at the 95% confidence level:

$$\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) < 3.0 \times 10^{-9} \ (95\% \ CL).$$

The limit is obtained under the hypothesis of background only, with $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ left free to be determined in the fit. The expected limit is $1.8^{+0.7}_{-0.4} \times 10^{-9}$.

An upper limit based on the CL$_s$ method is also set on $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$. The expected limit obtained from pseudo-MC samples generated according to the observed amplitudes of backgrounds and $B^0_s$ signal is $(5.7^{+2.1}_{-1.8}) \times 10^{-10}$ at a confidence level of 95%. The observed limit is:

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 4.2 \times 10^{-10} \ (95\% \ CL).$$

The observed upper limit is above the SM prediction and also covers the central value of the combination of the measurements by CMS and LHCb [13]. The expected significance for $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ according to the SM prediction is equal to 0.2 standard deviations.
Figure 9: Contours in the plane $\mathcal{B}(B^0_s \to \mu^+\mu^-), \mathcal{B}(B^0 \to \mu^+\mu^-)$ for intervals of $-2 \Delta \ln(L)$ equal to 2.3, 6.2 and 11.8 relative to the absolute maximum of the likelihood, without imposing the constraint of non-negative branching fractions. Also shown are the corresponding contours for the combined result of the CMS and LHCb experiments, the SM prediction, and the maximum of the likelihood within the boundary of non-negative branching fractions, with the error bars covering the 68.3% confidence range for $\mathcal{B}(B^0_s \to \mu^+\mu^-)$.

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 25 fb$^{-1}$ of 7 TeV and 8 TeV proton–proton collision data collected by the ATLAS experiment in Run 1 of LHC.

For $B^0$ an upper limit $\mathcal{B}(B^0 \to \mu^+\mu^-) < 4.2 \times 10^{-10}$ is placed at the 95% confidence level, based on the CL$_s$ method. The limit is compatible with the predictions based on the SM and with the combined result of the CMS and LHCb experiments.

For $B^0_s$ the result is $\mathcal{B}(B^0_s \to \mu^+\mu^-) = (0.9^{+1.1}_{-0.8}) \times 10^{-9}$, where the errors include both the statistical and systematic uncertainties. An upper limit $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 3.0 \times 10^{-9}$ at 95% CL is placed, lower than the SM prediction, and in better agreement with the measurement of CMS and LHCb.

A $p$-value of 4.8% is found for the compatibility of the results with the SM prediction.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.
We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; STFC, Belarus; CNPq and FAPESC, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEADSM/IRFU, France; GNSF, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; RCN, Norway; MNI SW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, 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, the Canada Council, CANARIE, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BSF, GIF and Minerva, Israel; BRF, Norway; the Royal Society and Leverhulme Trust, United Kingdom.

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


The ATLAS Collaboration

Division of Physics, TOBB University of Economics and Technology, Ankara, Turkey
5 LAPP, CNRS/IN2P3 and Université Savoie Mont Blanc, Annecy-le-Vieux, France
6 High Energy Physics Division, Argonne National Laboratory, Argonne IL, United States of America
7 Department of Physics, University of Arizona, Tucson AZ, United States of America
8 Department of Physics, The University of Texas at Arlington, Arlington TX, United States of America
9 Physics Department, University of Athens, Athens, Greece
10 Physics Department, National Technical University of Athens, Zografou, Greece
11 Department of Physics, The University of Texas at Austin, Austin TX, United States of America
12 Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
13 Institut de Física d’Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Barcelona, Spain, Spain
14 Institute of Physics, University of Belgrade, Belgrade, Serbia
15 Department for Physics and Technology, University of Bergen, Bergen, Norway
16 Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley CA, United States of America
17 Department of Physics, Humboldt University, Berlin, Germany
18 Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern, Switzerland
19 School of Physics and Astronomy, University of Birmingham, Birmingham, United Kingdom
20 (a) Department of Physics, Bogazici University, Istanbul; (b) Department of Physics Engineering, Gaziantep University, Gaziantep; (d) Istanbul Bilgi University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey; (e) Bahcesehir University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey, Turkey
21 Centro de Investigaciones, Universidad Antonio Narino, Bogota, Colombia
22 (a) INFN Sezione di Bologna; (b) Dipartimento di Fisica e Astronomia, Università di Bologna, Bologna, Italy
23 Physikalisches Institut, University of Bonn, Bonn, Germany
24 Department of Physics, Boston University, Boston MA, United States of America
25 Department of Physics, Brandeis University, Waltham MA, United States of America
26 (a) Universidade Federal do Rio De Janeiro COPPE/EE/IF, Rio de Janeiro; (b) Electrical Circuits Department, Federal University of Juiz de Fora (UFJJF), Juiz de Fora; (c) Federal University of Sao Joao del Rei (UFSSJ), Sao Joao del Rei; (d) Instituto de Fisica, Universidade de Sao Paulo, Sao Paulo, Brazil
27 Physics Department, Brookhaven National Laboratory, Upton NY, United States of America
28 (a) Transilvania University of Brasov, Brasov, Romania; (b) National Institute of Physics and Nuclear Engineering, Bucharest; (c) National Institute for Research and Development of Isotopic and Molecular Technologies, Physics Department, Cluj Napoca; (d) University Politehnica Bucharest, Bucharest; (e) West University in Timisoara, Timisoara, Romania
29 Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
30 Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
31 Department of Physics, Carleton University, Ottawa ON, Canada
32 CERN, Geneva, Switzerland
33 Enrico Fermi Institute, University of Chicago, Chicago IL, United States of America
34 (a) Departamento de Física, Pontificia Universidad Católica de Chile, Santiago; (b) Departamento de Física, Universidad Técnica Federico Santa María, Valparaíso, Chile
35 (a) Institute of High Energy Physics, Chinese Academy of Sciences, Beijing; (b) Department of Modern Physics, University of Science and Technology of China, Anhui; (c) Department of Physics, Nanjing University, Jiangsu; (d) School of Physics, Shandong University, Shandong; (e) Department of
Physics and Astronomy, Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai Jiao Tong University, Shanghai; (also affiliated with PKU-CHEP); (f) Physics Department, Tsinghua University, Beijing 100084, China
36 Laboratoire de Physique Corpusculaire, Clermont Université and Université Blaise Pascal and CNRS/IN2P3, Clermont-Ferrand, France
37 Nevis Laboratory, Columbia University, Irvington NY, United States of America
38 Niels Bohr Institute, University of Copenhagen, Kobenhavn, Denmark
39 (a) INFN Gruppo Collegato di Cosenza, Laboratori Nazionali di Frascati; (b) Dipartimento di Fisica, Università della Calabria, Rende, Italy
40 (a) AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Krakow; (b) Marian Smoluchowski Institute of Physics, Jagiellonian University, Krakow, Poland
41 Institute of Nuclear Physics Polish Academy of Sciences, Krakow, Poland
42 Physics Department, Southern Methodist University, Dallas TX, United States of America
43 Physics Department, University of Texas at Dallas, Richardson TX, United States of America
44 DESY, Hamburg and Zeuthen, Germany
45 Institut für Experimentelle Physik IV, Technische Universität Dortmund, Dortmund, Germany
46 Institut für Kern- und Teilchenphysik, Technische Universität Dresden, Dresden, Germany
47 Department of Physics, Duke University, Durham NC, United States of America
48 SUPA - School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
49 INFN Laboratori Nazionali di Frascati, Frascati, Italy
50 Fakultät für Mathematik und Physik, Albert-Ludwigs-Universität, Freiburg, Germany
51 Section de Physique, Université de Genève, Geneva, Switzerland
52 (a) INFN Sezione di Genova; (b) Dipartimento di Fisica, Università di Genova, Genova, Italy
53 (a) E. Andronikashvili Institute of Physics, Iv. Javakhishvili Tbilisi State University, Tbilisi; (b) High Energy Physics Institute, Tbilisi, Georgia
54 II Physikalisches Institut, Justus-Liebig-Universität Giessen, Giessen, Germany
55 SUPA - School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
56 II Physikalisches Institut, Georg-August-Universität, Göttingen, Germany
57 Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, Grenoble, France
58 Department of Physics, Hampton University, Hampton VA, United States of America
59 Laboratory for Particle Physics and Cosmology, Harvard University, Cambridge MA, United States of America
60 (a) Kirchhoff-Institut für Physik, Ruprecht-Karls-Universität Heidelberg, Heidelberg; (b) Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg; (c) ZITI Institut für technische Informatik, Ruprecht-Karls-Universität Heidelberg, Mannheim, Germany
61 Faculty of Applied Information Science, Hiroshima Institute of Technology, Hiroshima, Japan
62 (a) Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong; (b) Department of Physics, The University of Hong Kong, Hong Kong; (c) Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
63 Department of Physics, Indiana University, Bloomington IN, United States of America
64 Institut für Astro- und Teilchenphysik, Leopold-Franzens-Universität, Innsbruck, Austria
65 University of Iowa, Iowa City IA, United States of America
66 Department of Physics and Astronomy, Iowa State University, Ames IA, United States of America
67 Joint Institute for Nuclear Research, JINR Dubna, Dubna, Russia
68 KEK, High Energy Accelerator Research Organization, Tsukuba, Japan
69 Graduate School of Science, Kobe University, Kobe, Japan

41
Netherlands

109 Department of Physics, Northern Illinois University, DeKalb IL, United States of America
110 Budker Institute of Nuclear Physics, SB RAS, Novosibirsk, Russia
111 Department of Physics, New York University, New York NY, United States of America
112 Ohio State University, Columbus OH, United States of America
113 Faculty of Science, Okayama University, Okayama, Japan
114 Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman OK, United States of America
115 Department of Physics, Oklahoma State University, Stillwater OK, United States of America
116 Palacký University, RCPTM, Olomouc, Czech Republic
117 Center for High Energy Physics, University of Oregon, Eugene OR, United States of America
118 LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
119 Graduate School of Science, Osaka University, Osaka, Japan
120 Department of Physics, University of Oslo, Oslo, Norway
121 Department of Physics, Oxford University, Oxford, United Kingdom

122 (a) INFN Sezione di Pavia; (b) Dipartimento di Fisica, Università di Pavia, Pavia, Italy
123 Department of Physics, University of Pennsylvania, Philadelphia PA, United States of America
124 National Research Centre "Kurchatov Institute" B.P.Konstantinov Petersburg Nuclear Physics Institute, St. Petersburg, Russia
125 (a) INFN Sezione di Pisa; (b) Dipartimento di Fisica E. Fermi, Università di Pisa, Pisa, Italy
126 Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh PA, United States of America

127 (a) Laboratório de Instrumentação e Física Experimental de Partículas - LIP, Lisboa; (b) Faculdade de Ciências, Universidade de Lisboa, Lisboa; (c) Department of Physics, University of Coimbra, Coimbra; (d) Centro de Física Nuclear da Universidade de Lisboa, Lisboa; (e) Departamento de Física, Universidade do Minho, Braga; (f) Departamento de Física Teorica y del Cosmos and CAFPE, Universidad de Granada, Granada (Spain); (g) Dep Fisica and CEFITEC of Faculdade de Ciencias e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
128 Institute of Physics, Academy of Sciences of the Czech Republic, Praha, Czech Republic
129 Czech Technical University in Prague, Praha, Czech Republic
130 Faculty of Mathematics and Physics, Charles University in Prague, Praha, Czech Republic
131 State Research Center Institute for High Energy Physics (Protvino), NRC KI, Russia
132 Particle Physics Department, Rutherford Appleton Laboratory, Didcot, United Kingdom
133 (a) INFN Sezione di Roma; (b) Dipartimento di Fisica, Sapienza Università di Roma, Roma, Italy
134 (a) INFN Sezione di Roma Tor Vergata; (b) Dipartimento di Fisica, Università di Roma Tor Vergata, Roma, Italy
135 (a) INFN Sezione di Roma Tre; (b) Dipartimento di Matematica e Fisica, Università Roma Tre, Roma, Italy
136 (a) Faculté des Sciences Ain Chock, Réseau Universitaire de Physique des Hautes Energies - Université Hassan II, Casablanca; (b) Centre National de l’Energie des Sciences Techniques Nucleaires, Rabat; (c) Faculté des Sciences Semlalia, Université Cadi Ayyad, LPHEA-Marrakech; (d) Faculté des Sciences, Université Mohamed Premier and LPTPM, Oujda; (e) Faculté des sciences, Université Mohammed V, Rabat, Morocco
137 DSM/IRFU (Institut de Recherches sur les Lois Fondamentales de l’Univers), CEA Saclay (Commissariat à l’Energie Atomique et aux Energies Alternatives), Gif-sur-Yvette, France
138 Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz CA, United States of America
173 Department of Physics, University of Wisconsin, Madison WI, United States of America
174 Fakultät für Physik und Astronomie, Julius-Maximilians-Universität, Würzburg, Germany
175 Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal, Germany
176 Department of Physics, Yale University, New Haven CT, United States of America
177 Yerevan Physics Institute, Yerevan, Armenia
178 Centre de Calcul de l’Institut National de Physique Nucléaire et de Physique des Particules (IN2P3), Villeurbanne, France
a Also at Department of Physics, King’s College London, London, United Kingdom
b Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
c Also at Novosibirsk State University, Novosibirsk, Russia
d Also at TRIUMF, Vancouver BC, Canada
e Also at Department of Physics & Astronomy, University of Louisville, Louisville, KY, United States of America
f Also at Department of Physics, California State University, Fresno CA, United States of America
g Also at Department of Physics, University of Fribourg, Fribourg, Switzerland
h Also at Departament de Física de la Universitat Autònoma de Barcelona, Barcelona, Spain
i Also at Departamento de Física e Astronomia, Faculdade de Ciencias, Universidade do Porto, Portugal
j Also at Tomsk State University, Tomsk, Russia
k Also at Universita di Napoli Parthenope, Napoli, Italy
l Also at Institute of Particle Physics (IPP), Canada
m Also at National Institute of Physics and Nuclear Engineering, Bucharest, Romania
n Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg, Russia
o Also at Department of Physics, The University of Michigan, Ann Arbor MI, United States of America
p Also at Centre for High Performance Computing, CSIR Campus, Rosebank, Cape Town, South Africa
q Also at Louisiana Tech University, Ruston LA, United States of America
r Also at Institucio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona, Spain
s Also at Graduate School of Science, Osaka University, Osaka, Japan
t Also at Department of Physics, National Tsing Hua University, Taiwan
u Also at Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen, Netherlands
v Also at Department of Physics, The University of Texas at Austin, Austin TX, United States of America
w Also at Institute of Theoretical Physics, Ilia State University, Tbilisi, Georgia
x Also at CERN, Geneva, Switzerland
y Also at Georgian Technical University (GTU), Tbilisi, Georgia
z Also at Ochadai Academic Production, Ochanomizu University, Tokyo, Japan
aa Also at Manhattan College, New York NY, United States of America
ab Also at Hellenic Open University, Patras, Greece
ac Also at Academia Sinica Grid Computing, Institute of Physics, Academia Sinica, Taipei, Taiwan
ad Also at School of Physics, Shandong University, Shandong, China
ae Also at Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia
af Also at Section de Physique, Université de Genève, Geneva, Switzerland
ag Also at Eotvos Lorand University, Budapest, Hungary
ah Also at International School for Advanced Studies (SISSA), Trieste, Italy
ai Also at Department of Physics and Astronomy, University of South Carolina, Columbia SC, United States of America
aj Also at School of Physics and Engineering, Sun Yat-sen University, Guangzhou, China
\textsuperscript{ak} Also at Institute for Nuclear Research and Nuclear Energy (INRNE) of the Bulgarian Academy of Sciences, Sofia, Bulgaria
\textsuperscript{ad} Also at Faculty of Physics, M.V.Lomonosov Moscow State University, Moscow, Russia
\textsuperscript{am} Also at Institute of Physics, Academia Sinica, Taipei, Taiwan
\textsuperscript{an} Also at National Research Nuclear University MEPhI, Moscow, Russia
\textsuperscript{ao} Also at Department of Physics, Stanford University, Stanford CA, United States of America
\textsuperscript{ap} Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary
\textsuperscript{aq} Also at Flensburg University of Applied Sciences, Flensburg, Germany
\textsuperscript{ar} Also at University of Malaya, Department of Physics, Kuala Lumpur, Malaysia
\textsuperscript{as} Also at CPPM, Aix-Marseille Université and CNRS/IN2P3, Marseille, France
\textsuperscript{*} Deceased