Measurement of $b$-hadron pair production with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: A measurement of $b$-hadron pair production is presented, based on a data set corresponding to an integrated luminosity of 11.4 fb$^{-1}$ of proton-proton collisions recorded at $\sqrt{s} = 8$ TeV with the ATLAS detector at the LHC. Events are selected in which a $b$-hadron is reconstructed in a decay channel containing $J/\psi \rightarrow \mu\mu$, and a second $b$-hadron is reconstructed in a decay channel containing a muon. Results are presented in a fiducial volume defined by kinematic requirements on three muons based on those used in the analysis. The fiducial cross section is measured to be $17.7 \pm 0.1$(stat.) $\pm 2.0$(syst.) nb.

A number of normalised differential cross sections are also measured, and compared to predictions from the Pythia8, Herwig++, MadGraph5_aMC@NLO+Pythia8 and Sherpa event generators, providing new constraints on heavy flavour production.

KEYWORDS: B physics, Hadron-Hadron scattering (experiments)

ArXiv ePrint: 1705.03374
1 Introduction

The production of $b$- and $c$-quarks (heavy flavour) in proton-proton collisions provides a fruitful testing ground for the predictions of quantum chromodynamics (QCD). The mass of the $b$-quark introduces an additional scale, and the typical factorisation of QCD calculations into parton distribution functions, hard matrix element and softer parton shower components allows this mass to be introduced at different stages. Furthermore, there are
several possible schemes for the inclusion of the heavy-quark masses at these various stages and, given this freedom, the optimal settings must be determined by comparisons to data.

Several techniques have previously been used to measure heavy-flavour production. Some measurements rely on identifying hadronic jets containing a $b$-hadron\(^1\) based on the presence of a decay vertex or charged-particle tracks displaced from the primary interaction vertex due to the lifetime of the $b$-hadron. These measurements have been performed inclusively [1–4] and in the associated production of a vector boson ($V$) and jets [5–14]. Other classes of measurements identify semileptonic $b$-hadron decays through the presence of an electron or muon [15–20]. Finally, several measurements identify displaced decay vertices without any jet requirements [21, 22], or select specific decay channels such as $B^\pm \to J/\psi + X$ [23–29].

These measurements have highlighted disagreements both between different theoretical predictions and between those predictions and the data, motivating further study. Specifically, the region of small-angle $b\bar{b}$ pair production is particularly sensitive to the details of the various calculations, but so far remains only loosely constrained by the experimental data. Understanding this region, and heavy-flavour production more generally, is increasingly relevant at the LHC where, for example, searches for the Higgs boson produced in association with a vector boson ($VH$) and decaying to a $b\bar{b}$ pair rely extensively on the modelling of one of the main backgrounds: QCD production of $b\bar{b}$ pairs in association with vector bosons. These searches are most sensitive when the Higgs boson has significant momentum and the resulting $b$-quarks are produced at small opening angles, so constraining the theoretical predictions for the background in this region is an important input to these searches.

This paper describes a new measurement of the production of two $b$-hadrons, where one $b$-hadron decays to $J/\psi(\to \mu\mu) + X$ and the other to $\mu + Y$, resulting in three muons in the final state. The signal definition includes $J/\psi$ produced from feed-down from excited charmonium states, as well as muons produced in semileptonic cascade decays (the decay of a $c$-hadron produced in the decay of a $b$-hadron). To probe $b$-hadron production, several differential cross sections are measured, based on the kinematics of the $J/\psi$ (reconstructed from two muons) and the third muon. The variables considered are:

- the azimuthal separation between the $J/\psi$ and third muon, $\Delta\phi(J/\psi, \mu)$;\(^2\)
- the transverse momentum of the three-muon system, $p_T(J/\psi, \mu)$;
- the separation between the $J/\psi$ and the third muon in the azimuth-rapidity plane, $\Delta R(J/\psi, \mu)$,\(^3\) measured inclusively and split into $p_T(J/\psi, \mu) < 20\text{ GeV}$ (low $p_T$) and $p_T(J/\psi, \mu) \geq 20\text{ GeV}$ (high $p_T$) regions;

\(^1\)In this paper “$b$-hadron” and “$c$-hadron” are used as a shorthand for any weakly decaying hadron containing a valence $b$- or $\bar{b}$-quark, or a valence $c$- or $\bar{c}$-quark, respectively.

\(^2\)ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the $z$-axis along the beam pipe. The $x$-axis points from the IP to the centre of the LHC ring, and the $y$-axis points upward. Cylindrical coordinates ($r, \phi$) are used in the transverse plane, with $\phi$ being the azimuthal angle around the $z$-axis. Pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln\tan(\theta/2)$. Transverse momentum ($p_T$) is defined as the projection of the momentum in the transverse plane.

\(^3\)\(\Delta R(A, B)\) is defined as the separation between particles $A$ and $B$ in the azimuth-rapidity ($y$) plane, $\Delta R(A, B) = \sqrt{\Delta\phi(A, B)^2 + \Delta y(A, B)^2}$. 
• the separation in rapidity between the $J/\psi$ and the third muon, $\Delta y(J/\psi, \mu)$;
• the magnitude of the average rapidity of the $J/\psi$ and the third muon, $y_{\text{boost}}$;
• the mass of the three-muon system, $m(J/\psi, \mu)$;
• the ratio of the transverse momentum of the three-muon system to the invariant mass of the three-muon system, $p_T/m$, and its inverse, $m/p_T$.

Results are presented as the total cross section and normalised differential cross sections, all within a fiducial region defined in section 3.3.

2 ATLAS detector

The ATLAS detector [30] consists of an inner tracking system, referred to as the inner detector, surrounded by a superconducting solenoid, electromagnetic and hadronic calorimeters and an external muon spectrometer. Charged-particle tracks in the pseudorapidity range $|\eta| < 2.5$ are reconstructed with the inner detector, which is immersed in a 2 T axial field provided by the solenoid. The inner detector consists of pixel and microstrip semiconductor detectors, as well as a straw-tube transition radiation tracker. The solenoid is surrounded by sampling calorimeters, which span the pseudorapidity range up to $|\eta| = 4.9$. High-granularity liquid-argon (LAr) electromagnetic calorimeters are present up to $|\eta| = 3.2$. Hadronic calorimeters with scintillator tiles as active material cover $|\eta| < 1.74$, while LAr technology is used for hadronic calorimetry from $|\eta| = 1.5$ to $|\eta| = 4.9$. Outside the calorimeter system, air-core toroids provide a magnetic field for the muon spectrometer. Three stations of precision drift tubes and cathode-strip chambers provide measurements of muon tracks in the region $|\eta| < 2.7$. Resistive-plate and thin-gap chambers provide muon triggering capability up to $|\eta| = 2.4$.

3 Data set, event selection and simulation

3.1 Data set and event selection

This analysis uses proton-proton collision data recorded by the ATLAS experiment during 2012 at a centre-of-mass energy $\sqrt{s} = 8$ TeV. All events considered were recorded while the detector and trigger systems were functional and satisfied data quality requirements. Events are selected using a dimuon trigger where the muons are required to have opposite charge, be consistent with originating from the same production vertex, have $p_T(\mu) > 4$ GeV and $|\eta(\mu)| < 2.4$ and satisfy a loose dimuon mass selection, $2.5 < m(\mu^+, \mu^-) < 4.3$ GeV. The trigger was prescaled for various periods throughout 2012, and the integrated luminosity of the resulting data set is 11.4 fb$^{-1}$.

Muon candidates are reconstructed by combining inner detector tracks with tracks in the muon spectrometer. Candidates are required to have $|\eta| < 2.5$ and $p_T > 6$ GeV. Additional requirements are placed on the track quality, requiring a minimum number of hits in the different layers of each of the inner detector subcomponents. The $J/\psi$ candidates are
formed by selecting pairs of muons with opposite charge that have been identified to originate from a common vertex. These muons must also have $|\eta| < 2.3$ to lie within the trigger acceptance, and are required to match the direction of the corresponding trigger-level candidates. The invariant mass of the $J/\psi$ candidate is required to be in a window around the Particle Data Group (PDG) average of the $J/\psi$ mass of 3.097 GeV \cite{PDG}, $m(J/\psi_{\text{PDG}})$, specifically in the range 2.6–3.5 GeV. If more than one dimuon $J/\psi$ candidate satisfies the requirements above, the one with a reconstructed mass closest to the PDG mass is chosen.

A third-muon candidate is also required in the event. If there are more than three muons in an event the highest-$p_T$ muon that is not used in the $J/\psi$ candidate construction is defined to be the third muon.

3.2 Simulation

Inclusive $b$-hadron pair production is simulated with the Pythia8.186 \cite{PYTHIA} Monte Carlo event generator, based on a $2 \to 2$ matrix element calculation matched to a parton shower. The CTEQ6L1 \cite{CTEQ} parton distribution function (PDF) set is used, along with the AU2 \cite{AU2} tuned parameter settings. The $b$-quarks are treated as massless in the PDF set and the matrix element calculation, but the mass is reinstated during the parton shower. Events are filtered based on the presence of $J/\psi(\to \mu\mu)$ produced in the decay of a $b$-hadron, requiring two muons with $p_T > 6$ GeV.\footnote{All $J/\psi$ are decayed to $\mu\mu$. In events where two $J/\psi$ are present, the event is weighted by $\text{BR}(J/\psi \to \mu\mu)$ to correct for this.} The simulated collisions are overlaid with additional simulated minimum-bias collisions, to emulate the effect of multiple proton-proton interactions occurring during the same (in-time) or a nearby (out-of-time) bunch crossing, an effect called “pile-up”. These additional collisions are produced using Pythia8 with the A2 \cite{A2} tuned parameter settings and the MSTW2008 \cite{MSTW} PDF set. The simulated events are then passed through a GEANT4 \cite{GEANT4} simulation of the full ATLAS detector \cite{ATLAS}, and reconstructed using the same software as the real data.

In addition, an inclusive $pp \to b\bar{b}$ sample is simulated with the Herwig++ \cite{HERWIG} event generator using the CTEQ6L1 PDF set with UE-EE5 \cite{UE} parameter settings tuned for the underlying-event modelling. Again, this prediction includes a $2 \to 2$ matrix element calculation matched to a parton shower, but in this case, the $b$-quarks are considered massive in both the matrix element calculation and parton shower.

For both predictions, all $2 \to 2$ QCD processes are included and $b$-quarks can be produced either in the matrix element or in the subsequent parton shower phase.

3.3 Fiducial volume definition

Cross sections are defined in terms of observables at the particle level, defined in terms of particles with an average lifetime greater than 10 mm/$c$. The signal is defined by two final-state muons that originated from the decay of a $J/\psi$, which itself is a descendant of a $b$-hadron (including feed-down decays), and a third muon which originated from a different $b$-hadron (including cascade decays). Muons are “dressed” by adding the four-momentum of photons (excluding photons produced in the decays of hadrons) that are close to the
muon, defined by $\Delta R_\eta(\mu, \gamma) < 0.1$,\footnote{$\Delta R_\eta(A, B)$ is defined as the separation in the azimuth-pseudorapidity plane, $\Delta R_\eta(A, B) = \sqrt{\Delta \phi(A, B)^2 + \Delta \eta(A, B)^2}$.} to the muon four-momentum. The resulting dressed muons are required to have $p_T > 6$ GeV. The two muons from the $J/\psi$ are required to have $|\eta| < 2.3$ and the third muon is required to have $|\eta| < 2.5$, matching the trigger and reconstruction acceptances respectively.

4 Muon trigger and reconstruction efficiencies

The reconstructed $J/\psi$ and third muon candidates are taken from muons satisfying the selection criteria previously described. To extract the fiducial cross sections, the data must be corrected to account for inefficiencies in the selection of events containing muons. Corrections for the muon trigger and reconstruction efficiencies are obtained from data-driven techniques and applied to the observed events via an event weight.

The trigger efficiency correction \cite{41} is factored into several components:

- The single-muon trigger efficiency is calculated with a tag-and-probe method using $J/\psi$ candidate data. The efficiency is dependent on the kinematics of the muon and is parameterised as a function of $p_T$ and $q \times \eta$, where $q$ is the reconstructed muon charge. A correction derived from simulation is required to remove bias arising from the difference between the single-muon triggers used for the tag-and-probe method (which require $p_T(\mu) > 18$ GeV), and the dimuon trigger used in the analysis (which requires $p_T(\mu) > 4$ GeV for both muons).

- The dimuon trigger efficiency correction is calculated with data in two parts, a vertex-finding and opposite-charge correction, and a correction for the spatial overlap of muons in the trigger system, which typically results in a reduction in efficiency when $\Delta R_\eta(\mu, \mu) < 0.2$. The spatial overlap correction is calculated as a function of the angular separation between the two muons calculated in three separate dimuon rapidity intervals.

- A correction taken from simulation is applied to account for trigger inefficiencies in events with three muons where a pair of muons falls in the same trigger-level object, also typically for $\Delta R_\eta(\mu, \mu) < 0.2$. This correction is derived by fitting a linear function to the ratio of the number of $b$-hadron pair production events accepted by the trigger to all signal events, as a function of the separation of the third muon from the closest muon in the $J/\psi$ candidate.

The per-event trigger efficiency is given by the product of each single-muon efficiency, the additional dimuon trigger efficiency corrections and the three-muon efficiency correction.

The data are also corrected for the muon reconstruction efficiency \cite{42} which is applied per muon for each of the three muons in the event. The efficiency for a single muon is calculated in two parts: the efficiency for a muon track to be reconstructed in the inner detector is parameterised as a function of muon $p_T$ and $\eta$; then the efficiency for
reconstructing a muon given that an inner detector track has already been reconstructed is calculated using a tag-and-probe method using $Z \rightarrow \mu^+\mu^-$ and $J/\psi \rightarrow \mu^+\mu^-$ data. The correction is derived as a function of muon $p_T$ and $q \times \eta$.

5 Signal extraction

While the event selection described in section 3 provides a high-purity $J/\psi + \mu$ sample, the signal of interest in this analysis must still be extracted from the data. This is done in the following stages:

1. The yield of $J/\psi$ resulting from the decay of a $b$-hadron (the signal) is extracted using a simultaneous fit to the distributions of dimuon mass and pseudo-proper decay time, $\tau$, defined as

$$\tau = L_{xy} \times m(J/\psi_{PDG})/p_T(\mu^+\mu^-),$$

(5.1)

where $L_{xy}$ is the transverse distance between the primary vertex and the dimuon vertex positively (negatively) signed for a vertex with momentum vector pointing away from (towards) the primary vertex. The primary vertex is defined as the vertex formed from at least two tracks, each with $p_T > 400$ MeV, that has the largest summed track $p_T^2$ in the event. The quantity $p_T(\mu^+\mu^-)$ is the transverse momentum of the dimuon system. This fit is described in section 5.1, and the resulting determination of $J/\psi$ background contributions is used as an input to the next step.

2. Next, the yield of events containing an additional (i.e. third) muon resulting from the decay of a $b$-hadron is determined. First, the proportion of signal relative to background is enhanced by requiring $\tau > 0.25$ mm/c. The remaining contribution from certain backgrounds is determined from the $J/\psi$ fit, and other backgrounds are determined using a simultaneous fit to the transverse impact parameter significance, $S_{d_0}$, and the output of a boosted decision tree (BDT) trained to separate signal muons from misreconstructed muons, as described in section 5.2. The transverse impact parameter significance is defined as

$$S_{d_0} \equiv d_0/\sigma_{d_0},$$

(5.2)

where the transverse impact parameter, $d_0$, is the distance of closest approach of the track to the primary vertex point in the $r$-$\phi$ projection, with the $d_0$ sign given by the sign of the angular momentum of the track around the beam evaluated at the point of closest approach; and $\sigma_{d_0}$ is the (unsigned) uncertainty in the reconstructed $d_0$.

3. Some remaining irreducible sources of background are then subtracted from the fitted yields, as described in section 5.3.

4. Having determined the yield of $J/\psi$ and third muons, this is corrected for the effect of the $\tau$ requirement ($\tau > 0.25$ mm/c), as described in section 5.4.

5. Finally, the effects of detector resolution are corrected for, as described in section 5.5, determining the measured cross section for the signal.
This process is repeated for each kinematic bin of each differential cross section, resulting in the cross section in that bin. The width of these bins was chosen in order to retain the maximum information about the shape of each differential cross section, while maintaining a sufficient number of events in each bin to minimise the uncertainties. The data used in all fits are already corrected by the muon trigger and reconstruction efficiencies previously described.

5.1 Extraction of the $J/\psi$ composition

To identify the $J/\psi$ signal and background contributions to the data, a two-dimensional unbinned extended maximum-likelihood fit to the $J/\psi$ candidate mass and pseudo-proper decay time is performed. There are two components of the $J/\psi$ candidate dimuon mass spectrum: the real $J/\psi$ contribution, which is peaked at the $J/\psi$ mass; and the fake $J/\psi$ background contribution, which forms a continuum under the $J/\psi$ peak. The pseudo-proper decay time distribution for a real $J/\psi$ has two components. The first is for a $J/\psi$ from direct strong production in the hard scatter; this component is peaked at low $\tau$ and is referred to as prompt. The second corresponds to the $J/\psi$ signal from $b$-hadron decays that has, on average, larger values of $\tau$ due to the lifetime of the $b$-hadron, and is referred to as non-prompt. Both prompt and non-prompt $J/\psi$ include feed-down from excited charmonium states produced in either the hard scatter or the decay of a $b$-hadron. Similarly, there are three components of the fake $J/\psi$ background included in the fit, to account for different contributions to the pseudo-proper decay time (see below). The pseudo-proper decay time probability density function (p.d.f) components are convolved with a detector resolution function, modelled by a double Gaussian distribution centred at $\tau = 0 \text{ mm/c}$, with the widths of those Gaussian functions and their relative normalisation allowed to float freely in the fit.

In order to extract the number of non-prompt $J/\psi$, a fit model with five functional forms is used, based on the model used in ref. [26]. The forms are:

- **Non-prompt $J/\psi$**: the dimuon invariant mass is modelled using the sum of a Crystal Ball [43–45] and a Gaussian function. The $\tau$ distribution is modelled using a single-sided exponential decay function, convolved with the resolution function.

- **Prompt $J/\psi$**: the dimuon invariant mass is modelled using the same Crystal Ball and Gaussian function used for the non-prompt $J/\psi$. The $\tau$ distribution is modelled using a delta function at $\tau = 0 \text{ mm/c}$, convolved with the resolution function.

- **Prompt fake $J/\psi$ background**: the dimuon invariant mass distribution is modelled using a constant distribution (fits using a first order polynomial yield slopes consistent with zero for this component). The $\tau$ distribution is modelled using a delta function at $\tau = 0 \text{ mm/c}$, convolved with the resolution function.

- **Single-sided fake $J/\psi$ background**: the dimuon invariant mass distribution is modelled by a negative-slope exponential function. The $\tau$ distribution is modelled using a single-sided exponential decay function, convolved with the resolution function.
Figure 1. The two-dimensional dilepton mass-pseudo-proper decay time projection for a single differential observable bin: $0.4 < \Delta R(J/\psi, \mu) < 0.8$. The points with error bars are data, plotted at bin centres. The solid line is the projection of the unbinned maximum-likelihood fit to the data. The dilepton invariant mass projection (left) shows the total fake $J/\psi$ background component (dashed) in addition to the total p.d.f. The pseudo-proper decay time projection (right) also shows the prompt and non-prompt signal $J/\psi$ as dashed lines.

- Double-sided fake $J/\psi$ background: the dimuon invariant mass distribution is modelled by an exponential decay function. The $\tau$ distribution is modelled using a double-sided exponential decay function, convolved with the resolution function.

In the mass fit, the Gaussian and Crystal Ball functions share the same mean value. The Crystal Ball $n$ and $\alpha$ parameters are fixed to values derived from an inclusive $J/\psi$ fit, but the width is allowed to float. The mean values of the decay time model’s Gaussian function and double-sided exponential functions are fixed at zero. All other fit parameters are allowed to float.

The functional forms are combined into the fit model used in the unbinned extended maximum likelihood fit to the data. The fit for an example kinematic bin is shown in figure 1, where the prompt, single-sided and double-sided fake $J/\psi$ backgrounds are combined for clarity.

The stability and performance of the fit are checked by a closure test on simulated samples. Pseudo-data sets are produced by combining different numbers of prompt and non-prompt $J/\psi$ events. The two-dimensional $J/\psi$ model is then fitted to each pseudo-data set and the fractions of prompt and non-prompt $J/\psi$ events is extracted from the fit. The fractional difference between the input and fitted number of prompt and non-prompt $J/\psi$ events is compared, demonstrating that the fit is performing very well with deviations from the input composition consistently below 2%, which is comparable to the statistical uncertainty in the fit.

5.2 Extraction of the non-prompt muon signal

Having determined the $J/\psi$ contributions, the next step is to extract the signal yield: a non-prompt muon from the same hard scatter as a non-prompt $J/\psi$. To do this, a second two-
dimensional maximum-likelihood fit is performed, in this case using two observables that allow separation of the signal muons from background: the transverse impact parameter significance, and a BDT trained to separate signal muons from instrumental backgrounds.

The mechanisms of prompt and non-prompt $J/\psi$ production differ significantly, resulting in very different background contributions to the third-muon sample in each case. This leads to difficulties in fitting the third-muon distributions across the entire $J/\psi$ $\tau$ range. To increase the signal muon purity and improve the third-muon fit performance, the selected $J/\psi$ in each event is first required to have $\tau > 0.25$ mm/c, removing all of the prompt $J/\psi$ candidates. In addition, to reduce the $J/\psi$ candidates arising from the continuum background, a tighter invariant mass requirement of $2.95 < m(\mu^+,\mu^-) < 3.25$ GeV is applied. To account for the signal efficiency loss from these criteria, a correction is made once the signal yield is extracted from the fits.

There are several contributions to the third-muon background that are then considered: prompt muons produced at the primary vertex, muons produced in the decays of charged pions or kaons, third muons in events where the $J/\psi$ candidate is not a real $J/\psi$ but from the continuum background, and events where the $J/\psi$ and third muon are produced from different hard scatters in the same bunch crossing (referred to as the pile-up background).

There are two sources of fake muon background. Decay-in-flight (DIF) muons are the result of the decay of a charged pion or kaon. The small mass difference between the hadrons and resulting muons can result in only a small deflection of the track at the decay vertex. Therefore, the track of both the hadron and resulting muon can be reconstructed as coming from a single particle and a muon candidate may be identified. Muon candidates can also be reconstructed when charged hadrons leave tracks in the inner detector and charged particles from the shower in the hadronic calorimeter leave tracks in the muon spectrometer, referred to as hadronic shower leakage muons. Both the DIF and hadronic shower leakage sources of fake muons contribute significantly at low angular separations between the $J/\psi$ and third muon, as they can arise through decays such as $B^\pm \rightarrow J/\psi + K^\pm$.

To discriminate between the signal and fake muons, a BDT is constructed from kinematic variables that are sensitive to the production mechanism of the muons. These variables are:

- **Track deflection significance**: this parameter is the maximum value of the significance of the difference in track curvature calculated upstream and downstream of a point somewhere along the track reconstructed in the inner detector. DIF muons originating from a point inside the inner detector typically have higher values of track deflection significance than the signal muons.

- **Track deflection neighbour significance**: computed considering track segments, between adjacent hits, along the inner detector track. The largest value over the whole track of the significance of the angular difference between adjacent track segments is taken. This variable quantifies the significance of a deflection along a muon track; DIF muons originating from a point inside the inner detector populate larger values.
- **Momentum balance significance**: the significance of the difference between the track momenta reconstructed in the inner detector and in the muon spectrometer. If a pion or kaon decays outside the inner detector, the inner detector track may be matched to a lower-momentum muon spectrometer track produced by the resulting muon. The imbalance between the two track momenta is higher for these DIF muons than for inner detector and muon spectrometer tracks produced by a single muon. This variable also offers discrimination between signal and hadronic shower leakage muons.

- **Absolute pseudorapidity, $|\eta|$**: muon candidates produced in the background processes are more likely to be produced at high absolute pseudorapidities.

The BDT is developed in the TMVA framework [46]. The Pythia8 simulation is separated into two independent samples; one is used for training the BDT and the other is used to validate the performance. Figure 2 (left) shows a clear discrimination between the distribution of the BDT output for the signal and the fake muon background.

The BDT is trained using signal muons and fake muons, both taken from the simulated signal sample using the list of particles produced in each simulated collision, the Monte Carlo (MC) event record. Signal muons are defined as reconstructed muons, $\mu_{\text{reco}}$, that can be associated to a muon from the MC event record, $\mu_{\text{MC}}$, which was produced in the decay of a $b$-hadron. The association between reconstructed and simulated muons is performed by requiring $\Delta R_{\eta}(\mu_{\text{MC}}, \mu_{\text{reco}}) < 0.02$. Muons from DIF are identified as reconstructed muons which fail the signal definition, but are associated to a muon in the MC event record which has a charged pion or kaon as a parent. In this case, a looser association is required due to the change in the momentum at the point the pion/kaon decays: $\Delta R_{\eta}(\mu_{\text{MC}}, \mu_{\text{reco}}) < 0.15$. Finally, the remaining reconstructed muons may be identified as hadronic shower leakage if they do not match any muon from the MC event record, but do match a charged pion or kaon. Again, a looser association of $\Delta R_{\eta}(\pi/K_{\text{MC}}, \mu_{\text{reco}}) < 0.15$ is used to correctly identify all of this background.

There is also a background of third muons in events where the $J/\psi$ candidate comes from the dimuon continuum background (the “fake $J/\psi$” contribution introduced in section 5.1). Events lying in the regions outside the dimuon signal mass region ($2.95 < m(\mu^+\mu^-) < 3.25$ GeV) but within the range $2.60 < m(\mu^+\mu^-) < 3.50$ GeV consist almost entirely of this fake $J/\psi$ background. These events are therefore used to form the $S_d$ and BDT templates used for the fake $J/\psi$ background. These templates are normalised using the results from the two-dimensional $J/\psi$ fit (described in section 5.1) to determine the number of fake $J/\psi$ events in the signal mass region.

The background from pile-up is studied using $\Delta z_0$, defined as the difference between the reconstructed $z$-position (at their respective points of closest approach to the beam axis) of the third-muon track and the $J/\psi$ candidate muon which maximises the value of $\Delta z_0$. Figure 2 (right) shows the $\Delta z_0$ distribution for data after all event selection criteria are applied. The distribution consists of two components: a peaked structure centred on zero, representing events where the $J/\psi$ candidate and third muon are produced in the same proton-proton interaction, and a Gaussian-distributed component from pile-up spanning a wide $\Delta z_0$ range. To suppress the pile-up background, events are required to have
Figure 2. Left: distribution of the BDT output for signal muons (dark blue) and background fake muons (light blue) taken from simulation. Right: data $\Delta z_0$ distribution for the inclusive data set including the Gaussian pile-up background fit. The Gaussian function is fitted using the data outside the region marked by the dashed arrows.

$|\Delta z_0| < 40$ mm, and the pile-up background within this signal region is estimated by fitting a Gaussian model to the broad $\Delta z_0$ distribution, excluding the signal region from the fit. The integral of the Gaussian function within $|\Delta z_0| < 40$ mm gives the number of residual pile-up events in the signal region, and the shape of third muon BDT and $S_{d_0}$ distributions for pile-up events is taken from the pure pile-up region outside this $|\Delta z_0|$ range.

In summary, the two-dimensional fit of $S_{d_0}$ vs BDT for the third muon is constructed from the following components:

- **Signal $\mu$:** both the BDT and $S_{d_0}$ fit templates are taken from the Pythia8 simulated sample, using reconstructed muons matched to a muon in the simulated event record which derives from a $b$-hadron. This component is expected to populate the high values of the BDT output, signifying real muons, and have a wide $S_{d_0}$ distribution indicating production away from the interaction point. The shapes of the templates are fixed but the normalisation floats in the fit.

- **Prompt $\mu$:** both the BDT and $S_{d_0}$ templates are taken from $J/\psi$ muons in the inclusive $pp \rightarrow J/\psi$ Pythia8 simulation, where $J/\psi$ production is dominated by prompt production. These muons are real and thus should occupy the high values in the BDT output distribution and will have a narrow $S_{d_0}$ distribution as they are produced at the interaction point. The shapes of the templates are fixed but the normalisation floats in the fit.

- **Prompt and non-prompt fake $\mu$:** both the BDT and $S_{d_0}$ fake muon templates are taken from the Pythia8 simulated sample. The same BDT template is used for both the prompt and non-prompt components. The BDT shape has a large contribution at low values. The $S_{d_0}$ template is derived separately for the prompt and non-prompt components as fake muons can have both prompt as well as non-prompt sources. The shape of the templates are fixed but the normalisation floats in the fit.
Figure 3. The two-dimensional $S_{d0}$(left)–BDT(right) fit for a single differential observable bin: $0.4 < \Delta R(J/\psi, \mu) < 0.8$. The points with error bars are data, plotted at bin centres. The solid line is the projection of the unbinned maximum-likelihood fit to the data. The signal $\mu$, prompt $\mu$, non-prompt fake $\mu$, prompt fake $\mu$, fake $J/\psi$ and pile-up contributions are represented by dashed lines.

- **Fake $J/\psi$:** the BDT and $S_{d0}$ templates are derived from data and fixed in the fit.
- **Pile-up:** the BDT and $S_{d0}$ templates are derived from data and fixed in the fit.

The three $S_{d0}$ templates derived from simulation have a small shift in the mean of the distribution relative to data, due to the modelling of the beam-spot position. To correct for this, a small shift in the mean is derived independently for each of these three templates, by rerunning the fit once and allowing the shift to float; the shifted mean is then fixed at the fitted value for all other fits. The shifts for the three samples are consistent within uncertainties, and approximately $-0.05$.

Having derived templates for each of the expected third muon components in the data, an extended maximum-likelihood fit is carried out. The data are fitted in each bin of the various observables with the signal muon, prompt muon, fake $J/\psi$ and pile-up templates derived in the same kinematic bin. Figure 3 shows the result of the third-muon fits to the data for an example bin. While it can be seen that some of the templates suffer from statistical fluctuations, these have a very small effect on the fit result, and a systematic uncertainty is derived to cover these effects, as described in section 6.4.

The stability of the fit is verified with a closure test using simulated samples. Pseudo-data sets are created with varying fractions of prompt and non-prompt third-muon events and the fits repeated. The numbers of fitted prompt and non-prompt third muon events are compared to the values used to construct the pseudo-data sets. The differences between input and fitted values are typically below the 1% level, demonstrating that the fits perform well. Two additional qualitative cross-checks on the modelling of the fake muon component are performed by considering data control regions with requirements orthogonal to those of the signal region such that they are expected to contain more fake muons. The first control region is defined by reversing the pile-up rejection requirement so that the probability of a charged pion or kaon faking a third muon increases. The second control region looking at
prompt dimuon events is defined by reversing the pseudo-proper decay time requirement
where the fraction of $J/\psi$ production via weak decays is reduced, and the fraction of events
containing a charged pion or kaon faking a third muon is increased. The fits in these control
regions are performed inclusively as there is insufficient data to split into differential bins.
The third-muon fit procedure is the same as described above except that no pile-up template
is used in the pile-up control region. In both control regions, the fit behaves as expected,
giving a good description of the data and returning a higher fraction of fake muons.

5.3 Irreducible backgrounds

There are three additional sources of background that could not be constrained in the fits,
either because their contribution is too small to be reliably determined, or because their
characteristics are very similar to the signal. As no reliable data-driven determination is
possible, these irreducible backgrounds are instead subtracted from the post-fit signal yield
based on estimates derived from simulation.

The first source of irreducible background is $B_c \to J/\psi + \mu + X$ production. As both
the $J/\psi$ and third muon originate in the decay of the same hadron, this is considered to
be a source of background, and is concentrated at low values of $\Delta R(J/\psi, \mu)$, the region
of particular interest in this analysis. The production fraction of $B_c$ [47] and branching
fraction [31] of $B_c \to J/\psi \mu + X$ mean that this background is expected to be very small
and, in the absence of an identifiable signal in the data, the estimate is taken directly from
simulation. A prediction of the $B_c$ contribution passing the event selection is calculated
in each differential observable bin from both the PYTHIA8 and HERWIG++ simulated
samples. The average of the two predictions is then subtracted from the fitted signal yield
to remove the background from $B_c$ decays.

Another source of signal-like muons in this analysis is semileptonic decays of $c$-hadrons.
In the case where the $c$-hadron is produced in the decay of a $b$-hadron, these muons are
counted as part of the signal, but all other $c$-hadrons are considered as a source of back-
ground muons. This population of events is again expected to be small, as it requires a
displaced $J/\psi$ produced in a $b$-hadron decay, as well as a separate $c$-hadron; production
modes include separate $g \to b\bar{b}$ and $g \to c\bar{c}$ splittings in the same hard scatter, or double
parton scattering producing $b\bar{b} + c\bar{c} + X$ in a single proton-proton collision. To determine
the rate of such events in data, it is possible to use the third-muon $S_{d_0}$ fits: $c$-hadrons have
shorter lifetimes than $b$-hadrons, producing a narrower $S_{d_0}$ distribution. However, the
contribution is found to be so small that it is not possible to reliably extract the $c$-hadron
background with the number of events available in each fit. Instead the rate is derived from
simulation, where it is found that approximately 5% of all third muons passing the selection
originate from this source of background. To derive a correction for each bin of the mea-
sured kinematic distributions, the rate of $b+c$-hadron production is derived separately from
PYTHIA8 and HERWIG++ simulation, two event generators which have different models
of the parton shower, underlying event and double parton scattering, and the average of
these is used to remove the expected contribution from the fitted number of signal events.

The final background considered is from events where a charged pion or kaon traverses
the detector to the muon spectrometer without interacting with the detector material or
decaying, referred to as \textit{sail-through}. This background has a signature very similar to the signal third muons due to the presence of a well-defined track in the inner detector and a well-matched track in the muon spectrometer. The estimate of this background is taken from simulation, where a reconstructed muon not associated with a muon in the MC event record is matched to a charged kaon or pion which has no decay vertex inside the detector. The third-muon yield in each differential bin is corrected, after fitting, by removing the expected number of sail-through events. This number is estimated from the \textsc{Pythia8} simulation, by calculating the ratio of the numbers of sail-through and fake muons and using this to scale the background estimate from the number of fake muons fitted in data.

A summary of the relative sizes and distributions of all the backgrounds as a function of the \( \Delta R(J/\psi, \mu) \) observable is given in figure 4.

\textbf{5.4 Extrapolation to the full range of pseudo-proper decay time}

Once the signal yield has been determined, a correction must be applied to extrapolate the results obtained in the third muon fit (for \( \tau > 0.25 \text{ mm/c} \)) to the full range of \( J/\psi \) pseudo-proper decay time. The rate of non-prompt third muons is expected to be constant as a function of \( \tau \) for non-prompt \( J/\psi \) mesons, and this is confirmed in two tests. First in data, by repeating the third muon fit in different bins of \( \tau \) and observing that the ratio of non-prompt \( J/\psi \) events to each of the third muon fit components remains constant. Second with simulation, where the rate of non-prompt muons is indeed found to be constant as a function of \( \tau \) in events containing a \( J/\psi \) originating from a \( b \)-hadron decay.

The extrapolation to the full \( \tau \)-spectrum is then performed by simply correcting the third muon yield found in the \( \tau > 0.25 \text{ mm/c} \) region by an extrapolation factor taken as the ratio of all non-prompt \( J/\psi \) events over the full \( \tau \) range to the number of \( J/\psi \) events found above the \( \tau > 0.25 \text{ mm/c} \) requirement, as determined from the \( J/\psi \) fit. The correction is derived individually for each differential observable bin, based on the fit in that bin, and is typically a factor of around 1.9 with no significant kinematic dependence.
5.5 Resolution corrections

The final step in converting the measurement to a full particle-level cross section is to correct for the effects of detector resolution on the momentum and $\eta$ of the muons. Detector resolution can have two main effects, causing events to migrate between bins, or in and out of the fiducial volume.

Migration between bins can occur when events passing both the particle-level and detector-level selections are reconstructed in different bins of the differential cross sections. In this analysis, the analysis bins are all significantly wider than the detector resolution, so migrations between bins are a very small effect.

Migrations in and out of the fiducial volume must also be considered. Detector resolution effects can move an event into or out of the fiducial region (for example, by migrating individual muons above or below the muon $p_T > 6\text{ GeV}$ requirement). These effects are again found to be very small.

The final resolution correction factor combines both the bin-to-bin and fiducial migration effects, and as both these effects are small, a simple correction to each bin is sufficient. These corrections are derived by simply taking the ratio of the particle-level distributions to the detector-level distributions in the PYTHIA8 sample, where each sample distribution is derived independently and events are not required to pass both selections simultaneously. This ratio is then applied to the data to correct from detector-level quantities to particle-level quantities. The size of the correction is typically less than 2%, although for certain kinematic bins it can be as large as 5%.

6 Systematic uncertainties

Various systematic uncertainties are accounted for in this measurement. They broadly fit into three categories: uncertainties associated with the muon efficiency corrections to data, $J/\psi$ fit model systematic uncertainties and uncertainties in the background components in the fits. Each source of systematic uncertainty is considered individually by repeating the full differential analysis from the beginning with the systematic change implemented; the deviation from the nominal result is then taken as the uncertainty. All of the systematic uncertainties, apart from those concerning $J/\psi$ modelling, are allowed to vary independently upwards and downwards; from these two changes, the largest deviation from the nominal result is symmetrised and assigned as the uncertainty in the measurement from that source. The total upward or downward systematic uncertainty in the cross section is then calculated as the sum in quadrature of all contributions in the upward or downward directions respectively. When calculating the uncertainty in the normalised differential cross sections, shape-only systematic uncertainties are derived by varying each source of uncertainty up and down, while preserving the overall normalisation of the distribution. The total upwards or downwards systematic uncertainty for each bin is then the quadrature sum of these shape-only systematic uncertainties in the upwards or downwards directions respectively.

The main sources of systematic uncertainty in the cross section measurement are shown in figure 5 for the $\Delta R(J/\psi, \mu)$ observable. The statistical uncertainty includes the statistical uncertainty of the data and the statistical uncertainty of the third-muon templates.
Figure 5. The contribution from each systematic uncertainty source as a function of $\Delta R(J/\psi, \mu)$. Details of the different contributions are given in the text.

taken from simulation. The trigger uncertainty includes the uncertainties in the single-muon trigger efficiency, the dimuon efficiency correction, the nearby-muon correction and the simulation-based correction to the trigger efficiency. The muon reconstruction uncertainty includes the uncertainties in the muon reconstruction efficiency and the inner detector track reconstruction efficiency. The fit model uncertainty includes several variations of the functional forms used in the fit model and the fitting procedure. The BDT uncertainty includes several uncertainties on the simulation-derived templates and a data-driven uncertainty. The Fake $J/\psi$, $B_c$ and $c$-hadron uncertainties are included as described in section 6.4. The uncertainties in the resolution correction, pile-up and pile-up double counting, also described in section 6.4, are omitted from the figure for clarity but are included in all following calculations. The luminosity uncertainty is constant and therefore also not included in the figure.

All these uncertainties are described in more detail in the remainder of this section.

6.1 Luminosity uncertainty

A 1.9% uncertainty is assigned to the delivered integrated luminosity. The methodology used to determine this uncertainty is described in ref. [48].

6.2 Muon trigger and reconstruction efficiency uncertainties

As described in section 4, the trigger efficiency has several components: the single-muon efficiency; the efficiency to identify pairs of muons with opposite-charge and vertex requirements; and the efficiency to resolve pairs of nearby muons including a third-muon correction.

The single-muon efficiency is parameterised as a function of $p_T$ and $q \cdot \eta$. Each bin in the trigger efficiency map has an associated uncertainty resulting primarily from the limited data available to derive the efficiency. A Gaussian p.d.f. is formed for each map bin, with
the mean being given by the central value and the width by the uncertainty in that bin. Modified efficiency maps are formed by sampling randomly from the p.d.f. in each map bin; multiple maps can be created by repeating this procedure. The data are corrected for trigger efficiency using each modified map in turn to determine the number of events in the reweighted data set after applying the trigger efficiency correction. The distribution of the corrected data yields using all the maps is then fitted with a Gaussian function; the mean of this Gaussian function gives the nominal event yield after corrections, and the width gives the systematic uncertainty in the trigger efficiency weighting procedure.

The trigger efficiency map is also corrected by a factor derived from simulation to remove the bias in the maps due to the tag-and-probe threshold being different to that of the analysis selection, and a systematic uncertainty in this correction is defined by using the data-driven maps without the simulation-based correction applied and repeating the analysis. The change in the extracted differential cross sections from using the uncorrected maps is taken as a systematic uncertainty.

The efficiency to identify pairs of muons with opposite-charge and vertex requirements also has an uncertainty, and the impact of this is assessed separately by varying the nominal correction for this efficiency by one standard deviation when reweighting the data set. In addition, an uncertainty assigned to the efficiency correction for cases where the third muon is close to a trigger muon is assigned by varying the correction parameters within their uncertainties.

The uncertainty in the muon reconstruction efficiency is factorised into two components. A constant 0.5% uncertainty is included for the efficiency of reconstructing a muon track in the inner detector [42]. This is added coherently for each of the three muons in an event, resulting in a 1.5% systematic uncertainty. As in the case of the trigger map efficiency, the uncertainty in the muon reconstruction maps is defined by the spread of the data set yields when using a set of modified maps created by sampling from p.d.f.s defined in each bin of the efficiency map by the nominal efficiency and the uncertainty in that value.

The combined systematic uncertainty in the reconstruction and trigger efficiencies on the cross section is 10%. The relative normalised uncertainty can be as large as 10% in certain kinematic regions. This is dominated by the uncertainty in the dimuon correction to the single trigger efficiency except in the small $\Delta\phi(J/\psi, \mu)$ and $\Delta R(J/\psi, \mu)$ regions where the systematic uncertainty due to a third muon being close to a triggered muon is largest.

6.3 $J/\psi$ mass-lifetime model uncertainty

To assess potential bias in the fitted number of non-prompt $J/\psi$ candidates extracted from the two-dimensional dimuon fit due to the models used, various changes are made to the functions describing the fit components:

- The $J/\psi$ mass model is changed to a combination of two Gaussian functions.
- The non-prompt $J/\psi$ pseudo-proper decay time model is changed to a double exponential function convolved with the same resolution function.
- The resolution model is changed to a single Gaussian function.
• The fixed parameters in the Crystal Ball function are varied by ±10%.

• The dimuon mass model of the prompt and double-sided fake $J/\psi$ backgrounds is changed to be a first-order polynomial function.

• The single-sided fake $J/\psi$ background dimuon mass model is changed to a first order polynomial function.

• The single-sided fake $J/\psi$ pseudo-proper decay time model is changed to a double exponential function.

The analysis is repeated for each of the varied $J/\psi$ models, with only one change at a time. When performing fits to lower statistics regions, the fits are naturally less constrained, leading to a potential double counting of statistical uncertainties when performing the varied fits. To avoid this, the envelope of the largest deviation from the nominal event yield when considering all the individual model changes is taken as the total systematic uncertainty for the $J/\psi$ model uncertainty. This envelope is calculated separately in each differential bin.

The fit model variations give a 5% uncertainty in the total cross section and at most a 15% relative uncertainty in the normalised differential cross section. This is dominated by the change in decay time parameterisation of the non-prompt $J/\psi$ mesons and the mass parameterisation of the single-sided fake $J/\psi$ background.

6.4 Third-muon uncertainties

There are several contributions to the systematic uncertainty in the third-muon fit from the derivation of the templates and the estimation of the various background components.

There is a statistical uncertainty in the templates used for the third-muon fits, and the effect of this uncertainty is assessed through an ensemble test. Pseudo-templates are created by randomly sampling from the default templates, and the fit repeated using each of the modified pseudo-templates. This procedure provides a distribution of results for the extracted signal muon yield, and this distribution is fitted with a Gaussian function, the width of which is taken as a systematic uncertainty from the statistical fluctuations of the templates. The uncertainty is typically at the level of 1–2%. The shapes of the templates derived from simulation are found to have minimal dependence on the physics modelling in the simulation, due to the similar lifetimes of all $b$-hadrons and the very weak kinematic dependence of $S_{d_0}$, so no additional systematic uncertainty was required.

The contribution from events with a fake $J/\psi$ candidate which is combined with a third muon from another proton-proton interaction is double-counted by the fitting process. It would be possible to remove this double-counting by creating a four-dimensional fit but the complexity and lack of stability in such a fit make this very impractical and, since the level of double counting is very small, this is not necessary. Instead, each kinematic bin can be corrected by determining the number of fake $J/\psi$ events that are due to pile-up and removing that from the number of fake $J/\psi$ events extracted from the fit. The number of fake $J/\psi$ events that are due to pile-up is determined by fitting the $\Delta z_0$ distribution of events.
outside the signal mass window. Approximately 2% of fake $J/\psi$ events are found to be from pile-up events and the fake $J/\psi$ contribution is of order 10% of the selected data, therefore the signal yield is typically altered by 0.2%, but varies depending on the kinematic region.

The fake muon template contains two types of background with similar behaviour: DIF and hadronic leakage, as explained in section 5.2. Both backgrounds are due to the decays or interactions of charged pions and kaons. To assess the robustness of the simulation of the fake muon background, the templates used in the third muon fits are systematically altered. The BDT response is subtly different for pions and kaons but, due to the limited number of fake muon candidates in simulation, the two sources of fake muons are combined. The ratio of pions to kaons populating the fake muon templates is changed by $\pm 50\%$ to cover any effect of the combination. In addition, the fraction of pions and kaons decaying inside and outside the inner detector is varied by $\pm 50\%$. Finally the ratio of DIF muons and hadronic leakage muons in the fake muon template is changed by $\pm 50\%$. The BDT response is different for the two types of fake muons but the available number of simulated events do not allow separation of the two contributions in the template which is composed of approximately 75% DIF muons. The fractional composition of the fake muon template is changed to cover any mismodelling in simulation of the composition of the two sources of fake muons. The effects these systematic shifts have on the BDT template result in an uncertainty of less than 1% in the total cross section and up to 2% relative uncertainty in the differential normalised cross section in certain kinematic regions.

Furthermore, a data-driven uncertainty in the shape of the BDT distribution for fake muons is derived using two control regions. A pile-up region is defined by reversing the $|\Delta z_0|$ requirements, and a fake $J/\psi$ region is formed by selecting for events outside the dimuon mass window. The third-muon BDT distribution in these events is fitted with the usual simulation-derived BDT templates for real and fake muons. The fitted real component is then subtracted from data, leaving an estimate of the BDT template for fake muons. Due to statistical limitations, this process can only be performed on the inclusive data set and not differentially. This data-driven template is then used instead of the simulation-derived fake muon templates for each differential fit. The difference with respect to the usual result is then used to form the systematic uncertainty. The envelope from the largest deviation in each bin from either the pile-up or fake $J/\psi$ derived template is assigned as a systematic uncertainty in the modelling of fake muons in simulation. The corresponding uncertainty in the total cross section is 1% and can be as large as a 10% relative uncertainty in the normalised differential cross sections.

The uncertainty in the fake $J/\psi$ background estimate is assessed by changing the normalisation of the fake $J/\psi$ templates in the third-muon fits. The number of fake $J/\psi$ events is derived from the two-dimensional dimuon fits, given by normalisation of the three fake $J/\psi$ components within the dimuon signal mass window. Due to the dimuon pseudo-proper decay time requirement in the third-muon fit region, the single-sided fake $J/\psi$ component is the only background to contribute non-negligibly to the high pseudo-proper decay time region. The differential fits are repeated with fake $J/\psi$ template normalisation altered by the uncertainty ($\pm 1\sigma$) in the normalisation of the single-sided background. The systematic uncertainty assigned is less than 1% on the total cross section but as much as 3% relative uncertainty in the normalised differential cross sections.
For the uncertainty in the pile-up background, a similar procedure is used. The templates used in the third-muon fits are changed by altering the normalisation within the uncertainty. The uncertainty is derived from the Gaussian fit to $\Delta z_0$ and is applied as a $\pm 1\sigma$ variation to the nominal pile-up templates. The corresponding systematic uncertainties are very small, 0.4% in the total cross section and up to 2% in the normalised differential cross sections.

The $B_c$ background prediction is taken from the average of Pythia8 and Herwig++ simulation predictions. The difference between the two predictions is assigned as an uncertainty in the number of $B_c$-hadrons in the data set. Similarly the number of events estimated to be from $b+c$-hadrons is taken from the average of Pythia8 and Herwig++ simulation predictions. A systematic uncertainty is assigned to this prediction, using the difference between Pythia8 and Herwig++ for the rate of $b+c$-events. The systematic uncertainty assigned to the $B_c$-hadron prediction gives a small uncertainty of below 0.1% on the total cross section. Typically the relative uncertainty in the normalised differential cross section is also very small, except at low $\Delta R(J/\psi, \mu)$ where it can be almost 2%. For the $b+c$-hadron estimation, the systematic uncertainty in the total cross section is 2% and the relative uncertainty in the normalised differential cross section can be as large as 9%.

The sail-through background prediction is difficult to constrain with data, so the simulation-based estimate is varied by $\pm 50\%$ to assign an uncertainty due to mismodelling. The corresponding uncertainty in the total cross section is 0.1% and the relative uncertainty in the normalised differential cross section is less than 1% everywhere.

6.5 Resolution correction uncertainty

The uncertainty in the factors used to correct for events migrating in and out of the acceptance is estimated based on the statistical uncertainty in the simulated sample used to derive the correction. The correction, described in section 5.5, is derived from the ratio of events passing the particle-level selection to events passing the detector-level selection in simulation. The uncertainty in this ratio is calculated assuming these samples are uncorrelated. That is not entirely the case as they are derived from the same simulated sample, so this represents a conservative estimate of an uncertainty in this correction. The relative fractional uncertainty due to the resolution corrections for $\Delta \phi(J/\psi, \mu)$ is typically at the 1% level.

7 Results and interpretation

The total measured cross section in the fiducial region, defined in section 3.3, is

$$\sigma(B(\rightarrow J/\psi[\rightarrow \mu^+\mu^-] + X)B(\rightarrow \mu + X)) = 17.7 \pm 0.1(\text{stat}) \pm 2.0(\text{syst}) \text{ nb.}$$

While leading order calculations are not expected to accurately reproduce this total cross section, the normalised differential cross sections are used to test the accuracy of a number of predictions. First, comparisons are made using Pythia8, exploring several different options for the $g \rightarrow b\bar{b}$ splitting kernel, as this process dominates the region of particular interest: small-angle $b$-hadron production. The details of these settings are given in ref. [49]
Option label | Descriptions
--- | ---
Opt. 1 | The same splitting kernel, \((1/2)(z^2 + (1-z)^2)\), for massive as massless quarks, only with an extra \(\beta\) phase-space factor. This was the default setting in PYTHIA8.1, and currently must also be used with the MC@NLO \[50\] method.

Opt. 4 | A splitting kernel \(z^2 + (1-z)^2 + 8r_qz(1-z)\), normalised so that the \(z\)-integrated rate is \((\beta/3)(1 + r/2)\), and with an additional suppression factor \((1 - m_{qq}^2/m^{2}_{dipole})^3\), which reduces the rate of high-mass \(q\bar{q}\) pairs. This is the default setting in PYTHIA8.2.

Opt. 5 | Same as Option 1, but reweighted to an \(\alpha_s(km_{qq}^2)\) rather than the normal \(\alpha_s(p_T^2)\), with \(k = 1\).

Opt. 5b | Same as Option 5, but setting \(k = 0.25\).

Opt. 8 | Same as Option 4, but reweighted to an \(\alpha_s(km_{qq}^2)\) rather than the normal \(\alpha_s(p_T^2)\), with \(k = 1\).

Opt. 8b | Same as Option 8, but setting \(k = 0.25\).

**Table 1.** Description of PYTHIA8 options. Options 2, 3, 6 and 7 are less well physically motivated and not considered here. The notation used is as follows: \(r_q = m_q^2/m_{qq}^2\), \(\beta = \sqrt{1-4r_q}\), with \(m_q\) the quark mass and \(m_{qq}\) the \(q\bar{q}\) pair invariant mass.

and summarised in table 1. The settings explore one of the main theoretical degrees of freedom when evaluating gluon-splitting to heavy quarks: whether to use the relative \(p_T\) (Opt. 1 and 4) or mass (Opt. 5, 8, 5b and 8b) of the splitting to set the scale when determining the value of \(\alpha_s\) to be used in that splitting.

The distributions of \(\Delta R(J/\psi, \mu)\), \(m(J/\psi, \mu)\), \(\Delta \phi(J/\psi, \mu)\) and \(p_T/m\) are shown in figure 6. In general, PYTHIA8 does not reproduce the shape of the angular distributions in data within uncertainties. The \(p_T\)-based scale splitting kernels (Opt. 1 and 4) generally give a better description of the low \(\Delta R(J/\psi, \mu)\) region, with the kernel of Opt. 4 performing the best. This region is more suppressed in the mass-based scale kernels, although this suppression is overcome when lowering the scale by a factor of four (Opt. 5b and 8b), with Opt. 8b in particular performing comparably to Opt. 4. This pattern is repeated across the other differential cross sections considered.

To extend the comparisons, the HERWIG++ sample described in section 3 is included. Two samples are also simulated using MADGRAPH5_AMC@NLOv2.2.2 \[51\] at leading order in QCD interfaced to the PYTHIA8.186 parton shower model. In both samples, the CKKW-L \[52, 53\] merging procedure is applied, with a merging scale of 15 GeV. The A14 \[54\] tuned parameter settings are used together with the NNPDF2.3LO PDF set \[55\] for PYTHIA8. The EVTGEN1.2.0 program \[56\] is used for properties of the \(b\)-hadron and \(c\)-hadron decays. Both samples are generated with up to one additional parton in the matrix
Figure 6. Measured normalised differential cross sections as a function of $\Delta R(J/\psi, \mu)$, $m(J/\psi, \mu)$, $\Delta \phi(J/\psi, \mu)$ and $p_T/m$ compared to Pythia8 predictions with the different gluon-splitting parameter settings described in table 1. The bottom pane shows the ratio to data of the options that use a mass-based splitting scale. The middle pane shows the $p_T$-based scale options, along with the mass-based option that agrees best with data (Opt. 8b) for comparison.

element calculation. One is generated in the $5$-flavour scheme where massless $b$-quarks are included in PDF (the NNPDF3.0NLO [57] PDF set is used) and as possible initial state partons in the matrix element calculation; this is referred to as $5fl$ in the figures. The other is generated in the $4$-flavour scheme where $b$-quark mass is included in the calculation, but $b$-quarks are excluded from the PDF (the NNPDF3.0NLO 4fl PDF set is used), but are generated in the matrix element; this is referred to as $4fl$ in the figures.

Finally, a sample is simulated using the Sherpa2.1.1 [58] event generator. Matrix elements are calculated with two or three outgoing partons at leading order, and merged
with the SHERPA parton shower [59] using the ME+PS@LO prescription [60]. The CT10nlo [61] PDF set is used in conjunction with dedicated parton shower tuning developed by the SHERPA authors. Only the 5-flavour scheme sample is considered for the SHERPA event generator.

Due to the computational demands in producing sufficient events in the three-muon fiducial volume, both the MadGraph5_AMC@NLO+PYTHIA8 and SHERPA samples were produced only to the level of two $b$-hadrons. These predictions are corrected to a three-muon prediction using the *transfer functions* described in appendix A; the theoretical uncertainty in these transfer functions completely dominates the statistical uncertainty of the samples, and the quadrature sum of these two uncertainties is shown as an uncertainty bar for the MadGraph5_AMC@NLO+PYTHIA8 and SHERPA samples in the figures.

Figures 7, 8 and 9 compare the various predictions to the measured normalised differential cross sections. The best performing PYTHIA8 prediction (“Opt. 4”) is included again for reference along with the HERWIG++, MadGraph5_AMC@NLO+PYTHIA8 and SHERPA predictions. Considering first the $\Delta R(J/\psi, \mu)$ distribution, agreement with data is slightly better for HERWIG++ than PYTHIA8. The 4- and 5-flavour MadGraph5_AMC@NLO+PYTHIA8 predictions tend to sit on either side of the data, with the 4-flavour prediction generally being closer in shape. The 5-flavour scheme SHERPA prediction is similar in shape to the 5-flavour scheme MadGraph5_AMC@NLO+PYTHIA8 prediction, although the modelling is worse overall. The differences between 4- and 5-flavour predictions are enhanced in the high-\(p_T\) $\Delta R(J/\psi, \mu)$ distribution, with the 4-flavour remaining close to the data while both 5-flavour predictions move further away.

The general trends seen in $\Delta R(J/\psi, \mu)$ are also visible in $\Delta \phi(J/\psi, \mu)$, while in $\Delta y(J/\psi, \mu)$ the MadGraph5_AMC@NLO+PYTHIA8 and SHERPA predictions all provide a good description of the data, while PYTHIA8 and HERWIG++ tend to fall away at high $\Delta y(J/\psi, \mu)$. In $y_{\text{boost}}$, a distribution expected to be more sensitive to the PDFs than to the details of event generators, a comparable picture is seen across all predictions.

Moving to the other kinematic variables, the low $m(J/\psi, \mu)$ region again provides discrimination between the 4- and 5-flavour predictions, and is one case where the 5-flavour MadGraph5_AMC@NLO+PYTHIA8 prediction lies closer to the data than the 4-flavour prediction. However, at high values of the ratio $p_T/m$, the 4-flavour prediction clearly provides a much better description of the data than either of the 5-flavour predictions. Indeed, considering all distributions, the 4-flavour prediction from MadGraph5_AMC@NLO+PYTHIA8 provides the best description of the data overall. The predictions of PYTHIA8 and HERWIG++ are generally comparable, with indications that some further tuning could yield an improved description of the data. It should be noted, however, that the theoretical uncertainties in these predictions were not evaluated for this study.
Figure 7. Measured normalised differential cross sections as a function of $\Delta R(J/\psi, \mu)$, $\Delta \phi(J/\psi, \mu)$, low- and high-$p_T$ $\Delta R(J/\psi, \mu)$. Comparisons are made with predictions of Pythia8 and Herwig++. MadGraph5$aMC@NLO+Pythia8$ and Sherpa predictions are also compared having been corrected from the two-$b$-hadron production to the three-muon final state via transfer functions (indicated with *). There is no entry for these predictions in the lowest low-$p_T$ $\Delta R(J/\psi, \mu)$ bin as the transfer function is not defined in this bin. The Pythia8 “Opt. 4” gluon-splitting parameter settings are described in table 1. The ratio to data of the MadGraph5$aMC@NLO+Pythia8$ and Sherpa (middle pane), and Pythia8 and Herwig++ (bottom pane) are also shown.
Figure 8. Measured normalised differential cross sections as a function of $\Delta y(J/\psi, \mu)$, $y_{\text{boost}}$, $p_T(J/\psi, \mu)$ and $m(J/\psi, \mu)$. Comparisons are made with predictions of PYTHIA8 and HERWIG++. MADGRAPH5_AMC@NLO+PYTHIA8 and SHERPA predictions are also compared having been corrected from the two-$b$-hadron production to the three-muon final state via transfer functions (indicated with *). The PYTHIA8 “Opt. 4” gluon splitting parameter settings are described in table 1. The ratio to data of the MADGRAPH5_AMC@NLO+PYTHIA8 and SHERPA (middle pane), and PYTHIA8 and HERWIG++ (bottom pane) are also shown.
Figure 9. Measured normalised differential cross sections as a function of $m/p_T$ and $p_T/m$. Comparisons are made with predictions of PYTHIA8 and HERWIG++. MADGRAPH5_AMC@NLO+PYTHIA8 and SHERPA predictions are also compared having been corrected from the two-$b$-hadron production to the three-muon final state via transfer functions (indicated with *). The PYTHIA8 “Opt. 4” gluon splitting parameter settings are described in table 1. The ratio to data of the MADGRAPH5_AMC@NLO+PYTHIA8 and SHERPA (middle pane), and PYTHIA8 and HERWIG++ (bottom pane) are also shown.

8 Conclusion

A measurement of the production of $b$-hadron pairs in the $B(\rightarrow J/\psi[\rightarrow \mu^+\mu^-]+X)B(\rightarrow \mu+X)$ decay mode has been presented, using an integrated luminosity of 11.4 fb$^{-1}$ of proton-proton collisions at $\sqrt{s} = 8$ TeV recorded by the ATLAS detector at the LHC. A fiducial volume is defined by requiring two muons from the decay of a $J/\psi$, which itself originates from the decay of a $b$-hadron (including feed-down), and a third muon from the decay of a different $b$-hadron (including cascade decays). All muons are required to have $p_T > 6$ GeV, the two muons from the $J/\psi$ must have $|\eta| < 2.3$ and the third muon must have $|\eta| < 2.5$. The total cross section in this fiducial volume is measured to be $17.7 \pm 0.1(\text{stat}) \pm 2.0(\text{syst})$ nb.

Normalised differential cross sections were measured for ten kinematic observables designed to probe the underlying mechanisms of $b$-hadron production. These include a determination of nearby $b$-hadron pair production down to zero opening angle. This region is particularly sensitive to the production of $b$-quarks via gluon-splitting, which suffers from large theoretical uncertainties. Constraining this region is vital for many LHC measurements, including $H \rightarrow bb$ in the $VH$ production mode.

Predictions for the three-muon cross section are compared to the data. Several choices for the $g \rightarrow bb$ splitting kernel in PYTHIA8 are considered, along with nominal predictions from HERWIG++, MADGRAPH5_AMC@NLO+PYTHIA8 and SHERPA. These cover a range of different matrix element calculations and parton shower models, as well as 4- and 5-flavour treatments of the calculation. Of the PYTHIA8 options considered,
the $p_T$-based splitting kernel gives the best agreement with data, performing comparably to *Herwig++*. The best overall agreement with data comes from the 4-flavour *MadGraph5_aMC@NLO+Pythia8* prediction, which outperformed *Pythia8* and *Herwig++*, and the 5-flavour predictions from *MadGraph5_aMC@NLO+Pythia8* and *Sherpa*, though it should be noted that the associated theoretical uncertainties must be evaluated before the level of agreement with data can be fully quantified.

In conclusion, the measurements presented here provide a new test of QCD calculations, motivate the choices of predictions used to model $b$-hadron production, and further tuning of parameters in the calculations to improve the description of the data.

**Acknowledgments**

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

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; SPT, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DSM/IRFU, France; SRNSF, 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; NWO, Netherlands; RCN, Norway; MNiSW 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, ERDF, FP7, Horizon 2020 and Marie Sklodowska-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; CERCA Programme Generalitat de Catalunya, Generalitat Valenciana, Spain; the Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (U.K.) and BNL (U.S.A.), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. [62].
A Transfer functions

The cross sections measured in this analysis are computationally costly to simulate using available event generators, due to the low production rate of $b$-hadrons from inclusive simulation, and the low branching ratio for the three-muon final state. This is potentially a limiting factor in the use of these results for testing and tuning theory predictions. To get around this problem, a set of transfer functions are derived, which can be used to translate the predictions for a less computationally intensive inclusive two-$b$-hadron final state into the desired three-muon final state. These transfer functions therefore capture the average kinematics of $b$-hadron decays.

The transfer functions are calculated using simulation, by taking the ratio of a three-muon prediction to a two-$b$-hadron prediction. The three-muon prediction is calculated for the fiducial volume defined in section 3.3, and an equivalent two-$b$-hadron fiducial volume is chosen such that the shape of the $\Delta R(b$-hadron,$b$-hadron) distribution is as similar as possible to the shape of the $\Delta R(J/\psi ,\mu )$ distribution for the three-muon prediction. The dependence of the $\Delta R(b$-hadron,$b$-hadron) distribution on the definition of the two-$b$-hadron fiducial volume is studied using the Pythia8 and Herwig++ event generators described in section 7, and based on requiring two $b$-hadrons with a minimum $p_T$ requirement and within a defined rapidity region. The dependence on the rapidity requirement was found to be small, so $|y| < 2.4$ was chosen for both $b$-hadrons as it lies between the $\eta$ requirements on muons in the three-muon selection. The dependence on $p_T$ is more significant, and the optimal value of $p_T > 15.5$ GeV was chosen based on a fine scan of $p_T$ selections.

While these requirements give reasonable agreement between three-muon and two-$b$-hadron selections for angular variables, the $p_T$ and mass of the two-$b$-hadron system requires further treatment in order to reproduce more closely the equivalent three-muon distribution. Due to the additional particles produced in the decay, the muons carry only a fraction of the momentum of the $b$-hadrons, so the $b$-hadron momenta are scaled down in order to match the equivalent three-muon quantity more closely. While no choice of value for a simple scaling parameter gives good agreement across all $p_T$ and mass values, a scaling factor of 1.75 provides a reasonable compromise. This scaling is only used for mass and $p_T$. The two-$b$-hadron fiducial volume is therefore defined by requiring two weakly decaying $b$-hadrons with $p_T > 15.5$ GeV and $|y| < 2.4$. Both the mass and $p_T$ of the combined two-$b$-hadron system are scaled down by a factor of 1.75.

Transfer functions are then derived for each of the measured differential cross sections using the six different Pythia8 configurations described in section 7, and Herwig++. For each event generator, predictions for the two-$b$-hadron fiducial definition and the three-muon fiducial definition are derived independently, both are normalised to unit area, and the ratio of three-muon to two-$b$-hadron cross section is taken. The six Pythia8 and one Herwig++ transfer functions for each distribution are all slightly different, and used to define an envelope. The geometric centre of the envelope is taken as the default value for the transfer function in each bin, and the spread defines an uncertainty around that central value.
Figure 10. The transfer functions and uncertainty envelopes for $\Delta R(b\text{-hadron}, b\text{-hadron})$ (left) and $m(b\text{-hadron}, b\text{-hadron})$, with the 1.75 scale factor applied (right).

The resulting transfer functions for $m(b\text{-hadron}, b\text{-hadron})$ and $\Delta R(b\text{-hadron}, b\text{-hadron})$ are shown in figure 10. It can be seen that the transfer function is relatively constant in $\Delta R(b\text{-hadron}, b\text{-hadron})$, but shows more structure in mass around the nearby (low mass) and back-to-back (higher mass) kinematic edges. The kinematic requirements on the $b$-hadrons result in no event populating the lowest $\Delta R(b\text{-hadron}, b\text{-hadron})$ bin for the $p_T < 20 \text{ GeV}$ selection, and so the transfer function is not defined in this one bin.

The transfer functions are applied as a bin-by-bin multiplicative factor to two-$b$-hadron predictions from MadGraph5_aMC@NLO+Pythia8 and SHERPA, before these predictions are normalised to unit area. The uncertainty in the transfer functions is added in quadrature to any statistical uncertainty in the predictions from these event generators, and completely dominates those statistical uncertainties.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References


[10] D0 collaboration, V.M. Abazov et al., Measurement of the ratio of differential cross sections $\sigma(pp \rightarrow Z + b\text{jet})/\sigma(pp \rightarrow Z + \text{jet})$ in pp collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. D 87 (2013) 092010 [arXiv:1301.2233] [nSPIRE].

[11] D0 collaboration, V.M. Abazov et al., Measurement of the ratio of inclusive cross sections $\sigma(pp \rightarrow Z + 2\text{jets})/\sigma(pp \rightarrow Z + 2\text{jets})$ in pp collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. D 91 (2015) 052010 [arXiv:1501.05325] [nSPIRE].


* Also at Department of Physics, The University of Texas at Austin, Austin TX, United States of America
* Also at Institute of Theoretical Physics, Iria State University, Tbilisi, Georgia
* Also at CERN, Geneva, Switzerland
* Also at Georgian Technical University (GTU), Tbilisi, Georgia
* Also at Ochanomizu University, Tokyo, Japan
* Also at Manhattan College, New York NY, United States of America
* Also at Departamento de Física, Pontificia Universidad Católica de Chile, Santiago, Chile
* Also at Department of Physics, The University of Michigan, Ann Arbor MI, United States of America
* Also at The City College of New York, New York NY, United States of America
* Also at School of Physics, Shandong University, Shandong, China
* Also at Departamento de Física Teorica y del Cosmos and CAFPE, Universidad de Granada, Granada, Portugal
* Also at Department of Physics, California State University, Sacramento CA, United States of America
* Also at Moscow Institute of Physics and Technology State University, Dolgoprudny, Russia
* Also at Departement de Physique Nucleaire et Corpusculaire, Université de Genève, Geneva, Switzerland
* Also at Institut de Física d’Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Barcelona, Spain
* Also at School of Physics, Sun Yat-sen University, Guangzhou, China
* Also at Institute for Nuclear Research and Nuclear Energy (INRNE) of the Bulgarian Academy of Sciences, Sofia, Bulgaria
* Also at Faculty of Physics, M.V.Lomonosov Moscow State University, Moscow, Russia
* Also at National Research Nuclear University MEPhI, Moscow, Russia
* Also at Department of Physics, Stanford University, Stanford CA, United States of America
* Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary
* Also at Giresun University, Faculty of Engineering, Turkey
* Also at CPPM, Aix-Marseille Université and CNRS/IN2P3, Marseille, France
* Also at Department of Physics, Nanjing University, Jiangsu, China
* Also at University of Malaya, Department of Physics, Kuala Lumpur, Malaysia
* Also at Institute of Physics, Academia Sinica, Taipei, Taiwan
* Also at LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
* Deceased