Study of W boson production in association with beauty and charm

LHCb Collaboration; Anderson, J; Bernet, R; Bowen, E; Bursche, A; Chiapolini, N; Chrzaszcz, M; Elsasser, C; Graverini, E; Lionetto, F; Mauri, A; Müller, K; Steinkamp, O; Storaci, B; Tresch, M; et al

Abstract: The associated production of a W boson with a jet originating from either a light parton or heavy-flavor quark is studied in the forward region using proton-proton collisions. The analysis uses data corresponding to integrated luminosities of 1.0 and 2.0 fb$^{-1}$ collected with the LHCb detector at center-of-mass energies of 7 and 8 TeV, respectively. The W bosons are reconstructed using the $W \rightarrow \mu\nu$ decay and muons with a transverse momentum, $p_T$, larger than 20 GeV in the pseudorapidity range $2.0 < \eta < 4.5$. The partons are reconstructed as jets with $p_T > 20$ GeV and $2.2 < \eta < 4.2$. The sum of the muon and jet momenta must satisfy $p_T > 20$ GeV. The fraction of $W$+jet events that originate from beauty and charm quarks is measured, along with the charge asymmetries of the $W$+$b$ and $W$+$c$ production cross-sections. The ratio of the $W$+jet to $Z$+jet production cross-sections is also measured using the $Z \rightarrow \mu\mu$ decay. All results are in agreement with Standard Model predictions.

DOI: [https://doi.org/10.1103/PhysRevD.92.052001](https://doi.org/10.1103/PhysRevD.92.052001)

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: [https://doi.org/10.5167/uzh-122831](https://doi.org/10.5167/uzh-122831)
Published Version

Originally published at:
LHCb Collaboration; Anderson, J; Bernet, R; Bowen, E; Bursche, A; Chiapolini, N; Chrzaszcz, M; Elsasser, C; Graverini, E; Lionetto, F; Mauri, A; Müller, K; Steinkamp, O; Storaci, B; Tresch, M; et al (2015). Study of W boson production in association with beauty and charm. Physical Review D (Particles, Fields, Gravitation and Cosmology), 92(5):052001.
DOI: [https://doi.org/10.1103/PhysRevD.92.052001](https://doi.org/10.1103/PhysRevD.92.052001)
Study of $W$ boson production in association with beauty and charm

The LHCb collaboration†

Abstract

The associated production of a $W$ boson with a jet originating from either a light parton or heavy-flavor quark is studied in the forward region using proton-proton collisions. The analysis uses data corresponding to integrated luminosities of 1.0 and 2.0 fb$^{-1}$ collected with the LHCb detector at center-of-mass energies of 7 and 8 TeV, respectively. The $W$ bosons are reconstructed using the $W \rightarrow \mu\nu$ decay and muons with a transverse momentum, $p_T$, larger than 20 GeV in the pseudorapidity range $2.0 < \eta < 4.5$. The partons are reconstructed as jets with $p_T > 20$ GeV and $2.2 < \eta < 4.2$. The sum of the muon and jet momenta must satisfy $p_T > 20$ GeV. The fraction of $W$+jet events that originate from beauty and charm quarks is measured, along with the charge asymmetries of the $W+b$ and $W+c$ production cross sections. The ratio of the $W$+jet to $Z$+jet production cross sections is also measured using the $Z \rightarrow \mu\mu$ decay. All results are in agreement with Standard Model predictions.


© CERN on behalf of the LHCb collaboration, license CC-BY-4.0

†Authors are listed at the end of this article.
1 Introduction

Measurements of $W$+jet production in hadron collisions provide important tests of the Standard Model (SM), especially of perturbative quantum chromodynamics (QCD) in the presence of heavy-flavor quarks. Such measurements are also sensitive probes of the parton distribution functions (PDFs) of the proton. The ratio of the $W$+jet to $Z$+jet production cross sections is a test of perturbative QCD methods and constrains the light-parton PDFs of the proton.

The jet produced in association with the $W$ boson may originate either from a $b$ quark ($W+b$), $c$ quark ($W+c$) or light parton. Several processes contribute to the $W+b$ and $W+c$ final states at next-to-leading order (NLO) in perturbative QCD. The dominant mechanism for $W+c$ production is $gs \rightarrow Wc$, but there are also important contributions from $gs \rightarrow Wcg$, $gg \rightarrow Wc\bar{s}$, and $q\bar{q} \rightarrow Wc\bar{c}$ [1]. Therefore, measuring the ratio of the $W+c$ to $W$+jet production cross sections in the forward region at the LHC provides important constraints on the $s$ quark PDF [2,3] at momentum transfers of $Q \approx 100$ GeV ($c = 1$ throughout this article) and momentum fractions down to $x \approx 10^{-5}$. Previous measurements of the proton $s$ quark PDF were primarily based on deep inelastic scattering experiments with $Q \approx 1$ GeV and $x$ values $\mathcal{O}(0.1)$ [4–6]. The $W+c$ cross section has been measured at the Tevatron [7,8] and at the LHC [9,10] in the central region.

In the so-called four-flavor scheme, theoretical calculations are performed considering only the four lightest quarks in the proton [11]. Production of $W+b$ proceeds via $q\bar{q} \rightarrow Wg$ with $g \rightarrow b\bar{b}$ at leading order. If the $b$ quark content of the proton is considered, i.e. the five-flavor scheme, then single-$b$ production via $gb \rightarrow Wbq$ also contributes [12]. The ratio of the $W+b$ to $W$+jet cross sections thus places constraints both on the intrinsic $b$ quark content of the proton and the probability of gluon splitting into $b\bar{b}$ pairs. The $W+b$ cross section has been measured in the central region at the Tevatron [13,14] and at the LHC [15].

LHCb has measured the cross sections for inclusive $W$ and $Z$ production in proton-proton (pp) collisions at center-of-mass energy $\sqrt{s} = 7$ TeV [16–19], providing precision tests of the SM in the forward region. Additionally, measurements of the $Z$+jet and $Z+b$ cross sections have been made [20,21]. In this article, the associated production of a $W$ boson with a jet originating from either a light parton or a heavy-flavor quark is studied using pp collisions at center-of-mass energies of 7 and 8 TeV. The production of the $W+b$ final state via top quark decay is not included in the signal definition in this analysis, but is reported separately in Ref. [22].

A comprehensive approach is taken, where the inclusive $W$+jet, $W+b$ and $W+c$ contributions are measured simultaneously, rather than split across multiple measurements as in Refs. [9,10,15,23,26]. The identification of $c$ jets, in conjunction with $b$ jets, is performed using the tagging algorithm described in Ref. [27], which improves upon previous $c$-tagging methods where muons or exclusive decays were required to identify the jet [9,10]. For each center-of-mass energy, the following production cross section ratios are measured: $\sigma(Wb)/\sigma(Wj)$, $\sigma(Wc)/\sigma(Wj)$, $\sigma(W^{+}j)/\sigma(Zj)$, $\sigma(W^{-}j)/\sigma(Zj)$, $A(Wb)$, and
\[ A(Wc), \text{ where} \]
\[ A(Wq) \equiv \frac{\sigma(W^+q) - \sigma(W^-q)}{\sigma(W^+q) + \sigma(W^-q)}. \]  

The analysis is performed using the \( W \rightarrow \mu \nu \) decay and jets clustered with the anti-\( k_T \) algorithm \[28\] using a distance parameter \( R = 0.5 \). The following fiducial requirements are applied: both the muon and the jet must have momentum transverse to the beam, \( p_T \), greater than 20 GeV; the pseudorapidity of the muon must fall within \( 2.0 < \eta(\mu) < 4.5 \); the jet pseudorapidity must satisfy \( 2.2 < \eta(j) < 4.2 \); the muon and jet must be separated by \( \Delta R(\mu, j) > 0.5 \), where \( \Delta R \equiv \sqrt{\Delta \eta^2 + \Delta \phi^2} \) and \( \Delta \eta(\Delta \phi) \) is the difference in pseudorapidity (azimuthal angle) between the muon and jet momenta; and the transverse component of the sum of the muon and jet momenta must satisfy \( p_T(\mu + j) \equiv (\vec{p}(\mu) + \vec{p}(j))_T > 20 \text{ GeV} \).

All results reported in this article are for within this fiducial region, \textit{i.e.} no extrapolation outside of this region is performed.

The article is organized as follows: the detector, data sample and simulation are described in Sect. 2; the event selection is given in Sect. 3; the signal yields are determined in Sect. 4; the systematic uncertainties are outlined in Sect. 5; and the results are presented in Sect. 6.

### 2 The LHCb detector and data set

The LHCb detector \[29,30\] is a single-arm forward spectrometer covering the pseudorapidity range 2 < \( \eta < 5 \), designed for the study of particles containing b or c quarks. The detector includes a high-precision tracking system consisting of a silicon-strip vertex detector surrounding the \( pp \) interaction region \[31\], a large-area silicon-strip detector located upstream of a dipole magnet with a bending power of about 4 Tm, and three stations of silicon-strip detectors and straw drift tubes \[32\] placed downstream of the magnet. The tracking system provides a measurement of momentum, \( p \), of charged particles with a relative uncertainty that varies from 0.5\% at low momentum to 1.0\% at 200 GeV. The minimum distance of a track to a primary vertex, the impact parameter, is measured with a resolution of \((15 + 29/p_T) \mu \text{m}\), with \( p_T \) in GeV. Different types of charged hadrons are distinguished using information from two ring-imaging Cherenkov detectors. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic calorimeter and a hadronic calorimeter. The electromagnetic and hadronic calorimeters have energy resolutions of \( \sigma(E)/E = 10\%/\sqrt{E} \pm 1\% \) and \( \sigma(E)/E = 69\%/\sqrt{E} \pm 9\% \) (with \( E \) in GeV), respectively. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers \[33\].

The trigger \[34\] consists of a hardware stage, based on information from the calorimeter and muon systems, followed by a software stage, which applies a full event reconstruction. This analysis requires at least one muon candidate that satisfies the trigger requirement of \( p_T > 10 \text{ GeV} \). Global event cuts (GECs), which prevent high-occupancy events from dominating the processing time of the software trigger, are also applied and have an
efficiency of about 90% for $W+jet$ and $Z+jet$ events.

Two sets of $pp$ collision data collected with the LHCb detector are used: data collected during 2011 at $\sqrt{s} = 7$ TeV, corresponding to an integrated luminosity of 1.0 fb$^{-1}$, and data collected during 2012 at $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 2.0 fb$^{-1}$. Simulated $pp$ collisions, used to study the detector response, to define the event selection and to validate data-driven techniques, are generated using PYTHIA [35] with an LHCb configuration [36]. Decays of hadronic particles are described by EVTGEN [37] in which final-state radiation (FSR) is generated using PHOTOS [38]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit [39] as described in Ref. [40].

Results are compared with theoretical calculations at NLO using MCFM [41] and the CT10 PDF set [42]. The theoretical uncertainty is a combination of PDF, scale, and strong-coupling ($\alpha_s$) uncertainties. The PDF and scale uncertainties are evaluated following Refs. [42] and [43], respectively. The $\alpha_s$ uncertainty is evaluated as the envelope obtained using $\alpha_s(M_Z) \in [0.117, 0.118, 0.119]$ in the theory calculations.

3 Event selection

The signature for $W+jet$ events is an isolated high-$p_T$ muon and a well-separated jet, both produced in the same $pp$ interaction. Muon candidates are identified with tracks that have associated hits in the muon system. The muon candidate must have $p_T(\mu) > 20$ GeV and pseudorapidity within $2.0 < \eta(\mu) < 4.5$. Background muons from $W \to \tau \to \mu$ decays or semileptonic decays of heavy-flavor hadrons are suppressed by requiring the muon impact parameter to be less than 0.04 mm [16]. Background from high-momentum kaons and pions that enter the muon system and are misidentified as muons, is reduced by requiring that the sum of the energy of the associated electromagnetic and hadronic calorimeter deposits does not exceed 4% of the momentum of the muon candidate.

Jets are clustered using the anti-$k_T$ algorithm with a distance parameter $R = 0.5$, as implemented in FASTJET [44]. Information from all the detector subsystems is used to create charged and neutral particle inputs to the jet-clustering algorithm using a particle flow approach [20]. During 2011 and 2012, LHCb collected data with a mean number of $pp$ collisions per beam crossing of about 1.7. To reduce contamination from multiple $pp$ interactions, charged particles reconstructed within the vertex detector may only be clustered into a jet if they are associated with the same $pp$ collision.

Signal events are selected by requiring a muon candidate and at least one jet with $\Delta R(\mu, j) > 0.5$. For each event the highest-$p_T$ muon candidate that satisfies the trigger requirements is selected, along with the highest-$p_T$ jet from the same $pp$ collision. The high-$p_T$ muon candidate is not removed from the anti-$k_T$ inputs and so is clustered into a jet. This jet, referred to as the muon jet and denoted as $j_{\mu}$, is used to discriminate between $W+jet$ and dijet events. The requirement $p_T(j_{\mu} + j) > 20$ GeV is made to suppress dijet backgrounds, which are well balanced in $p_T$, unlike $W+jet$ events where there is undetected energy from the neutrino. Furthermore, the distribution of the fractional muon
candidate $p_T$ within the muon jet, $p_T(\mu)/p_T(j_\mu)$, is used to separate vector bosons from jets. For vector-boson production, this ratio deviates from unity only due to muon FSR, activity from the underlying event, or from neutral-particle production in a separate $pp$ collision, whereas for jet production this ratio is driven to smaller values by the presence of additional radiation produced in association with the muon candidate.

Events with a second, oppositely charged, muon candidate from the same $pp$ collision are vetoed. However, when the dimuon invariant mass is in the range $60 < M(\mu^+\mu^-) < 120$ GeV, such events are selected as $Z$+jet candidates and the $p_T(j_\mu+j)$ requirement is not applied. Two $Z$+jet data samples are selected at each center-of-mass energy: a data sample where only the $\mu^+$ is required to satisfy the trigger requirements and one where only the $\mu^-$ is required to satisfy them. The first sample is used to measure $\sigma(W^+j)/\sigma(Zj)$, while the second is used for $\sigma(W^-j)/\sigma(Zj)$. This strategy leads to approximate cancellation of the uncertainty in the trigger efficiency in the measurement of these ratios.

The reconstructed jets must have $p_T(j) > 20$ GeV and $2.2 < \eta(j) < 4.2$. The reduced $\eta(j)$ acceptance ensures nearly uniform jet reconstruction and heavy-flavor tagging efficiencies. The momentum of a reconstructed jet is scaled to obtain an unbiased estimate of the true jet momentum. The scaling factor, typically between 0.9 and 1.1, is determined from simulation and depends on the jet $p_T$ and $\eta$, the fraction of the jet transverse momentum measured with the tracking systems, and the number of $pp$ interactions in the event. No scaling is applied to the momentum of the muon jet. Migration of events in and out of the jet $p_T$ fiducial region due to the detector response is corrected for by an unfolding technique. Data-driven methods are used to obtain the unfolding matrix, with the resulting corrections to the measurements presented in this article being at the percent level.

The jets are identified, or tagged, as originating from the hadronization of a heavy-flavor quark by the presence of a secondary vertex (SV) with $\Delta R < 0.5$ between the jet axis and the SV direction of flight, defined by the vector from the $pp$ interaction point to the SV position. Two boosted decision trees (BDTs) [45,46], BDT($bc$|$udsg$) and BDT($b$|c), trained on the characteristics of the SV and the jet, are used to separate heavy-flavor jets from light-parton jets, and to separate $b$ jets from $c$ jets. The two-dimensional distribution of the BDT response observed in data is fitted to obtain the SV-tagged $b$, $c$ and light-parton jet yields. The SV-tagger algorithm is detailed in Ref. [27], where the heavy-flavor tagging efficiencies and light-parton mistag probabilities are measured in data.

4 Background determination

Contributions from six processes are considered in the $W$+jet data sample: $W$+jet signal events; $Z$+jet events where one muon is not reconstructed; top quark events producing a $W$+jet final state; $Z \rightarrow \tau\tau$ events where one $\tau$ lepton decays to a muon and the other decays hadronically; QCD dijet events; and vector boson pair production. Simulations based on NLO predictions show that the last contribution is negligible.

The signal yields are obtained for each muon charge and center-of-mass energy independently. The $p_T(\mu)/p_T(j_\mu)$ distribution is fitted to determine the $W$+jet yield of
each data sample. To determine the \( W+b \) and \( W+c \) yields, the subset of candidates with an SV-tagged jet is binned according to \( p_T(\mu)/p_T(j_{\mu}) \). In each \( p_T(\mu)/p_T(j_{\mu}) \) bin, the two-dimensional SV-tagger BDT-response distributions are fitted to determine the yields of \( b \)-tagged and \( c \)-tagged jets, which are used to form the \( p_T(\mu)/p_T(j_{\mu}) \) distributions for candidates with \( b \)-tagged and \( c \)-tagged jets. These \( p_T(\mu)/p_T(j_{\mu}) \) distributions are fitted to determine the SV-tagged \( W+b \) and \( W+c \) yields. Finally, to obtain \( \sigma(Wb)/\sigma(Wj) \) and \( \sigma(Wc)/\sigma(Wj) \), the jet-tagging efficiencies of \( \epsilon_{\text{tag}}(b) \approx 65\% \) and \( \epsilon_{\text{tag}}(c) \approx 25\% \) are accounted for. In all fits performed in this analysis, the templates are histograms with fixed shapes.

The \( p_T(\mu)/p_T(j_{\mu}) \) distributions are shown in Fig. [1] (in this and subsequent figures the pull represents the difference between the data and the fit, in units of standard deviations). The \( W \) boson yields are determined by performing binned extended-maximum-likelihood fits to these distributions with the following components:

- The \( W \) boson template is obtained by correcting the \( p_T(\mu)/p_T(j_{\mu}) \) distribution observed in \( Z+j \) events for small differences between \( W \) and \( Z \) decays derived from simulation.

- The template for \( Z \) boson events where one muon is not reconstructed is obtained by correcting, using simulation, the \( p_T(\mu)/p_T(j_{\mu}) \) distribution observed in fully reconstructed \( Z+j \) events for small differences expected in partially reconstructed \( Z+j \) events. The yield is fixed from the fully reconstructed \( Z+j \) data sample, where simulation is used to obtain the probability that the muon is missed, either because it is out of acceptance or it is not reconstructed.

- The templates for \( b \), \( c \) and light-parton jets are obtained using dijet-enriched data samples. These samples require \( p_T(j_{\mu}+j) < 10 \) GeV and, for the heavy-flavor samples, either a stringent \( b \)-tag or \( c \)-tag requirement on the associated jet. The templates are corrected for differences in the \( p_T(j_{\mu}) \) spectra between the dijet-enriched and signal regions. The contributions of \( b \), \( c \) and light-parton jets are each free to vary in the \( p_T(\mu)/p_T(j_{\mu}) \) fits.

The \( p_T(\mu)/p_T(j_{\mu}) \) fits determine the \( W+j \) yields, which include contributions from top quark and \( Z \rightarrow \tau \tau \) production. The top quark and \( Z \rightarrow \tau \tau \) contributions cannot be separated from \( W+j \) since their \( p_T(\mu)/p_T(j_{\mu}) \) distributions are nearly identical to that of \( W+j \) events. The subtraction of these backgrounds is described below.

The yields of events with \( W \) bosons associated with \( b \)-tagged and \( c \)-tagged jets are obtained by fitting the two-dimensional SV-tagger BDT-response distributions for \( \sqrt{s} = 7 \) and 8 TeV and for each muon charge separately in bins of \( p_T(\mu)/p_T(j_{\mu}) \). The SV-tagger BDT templates used in this analysis are obtained from the data samples enriched in \( b \) and \( c \) jets used in Ref. [27]. As a consistency check, the two-dimensional BDT distributions are fitted using templates from simulation; the yields shift only by a few percent. Figure [2] shows the BDT distributions combining all data in the most sensitive region, \( W+j \) events with \( p_T(\mu)/p_T(j_{\mu}) > 0.9 \). This is the region where the muon carries a large fraction of the muon-jet momentum and is, therefore, highly isolated. Figure [3] shows the distributions in
a dijet dominated region \((0.5 < p_T(\mu)/p_T(j_\mu) < 0.6)\). In the dijet region the majority of SV-tagged jets associated with the high-\(p_T\) muon candidate are found to be \(b\) jets. This is due to the large semileptonic branching fraction of \(b\) hadrons. In the \(W+\)jet signal region there are significant contributions from both \(b\) and \(c\) jets.  

As a consistency check, the \(b\), \(c\), and light-parton yields are obtained in the \(p_T(\mu)/p_T(j_\mu) > 0.9\) signal region from a fit using only two of the BDT inputs, both of which rely only on basic SV properties, the track multiplicity and the corrected mass, which is defined as

\[
M_{\text{cor}} = \sqrt{M^2 + |\vec{p}|^2 \sin^2 \theta + |\vec{p}| \sin \theta}, \tag{2}
\]

where \(M\) and \(|\vec{p}|\) are the invariant mass and momentum of the particles that form the SV, and \(\theta\) is the angle between \(|\vec{p}|\) and the flight direction. The corrected mass, which is the minimum mass for a long-lived hadron whose trajectory is consistent with the flight direction, peaks near the \(D\) meson mass for \(c\) jets and consequently provides excellent discrimination.
against other jet types. The SV track multiplicity identifies $b$ jets well, since $b$-hadron decays typically produce many displaced tracks. In Fig. 4, the distributions of $M_{\text{cor}}$ and SV track multiplicity for a subsample of SV-tagged events with $\text{BDT}(bc|udsg) > 0.2$ (see Fig. 2) are fitted simultaneously. The templates used in these fits are obtained from data.
in the same manner as the SV-tagger BDT templates. After correcting for the efficiency of requiring $\text{BDT}(bc|udsg) > 0.2$, the $b$ and $c$ yields determined from the fits to $M_{\text{cor}}$ and SV track multiplicity and from the two-dimensional BDT fits are consistent. The mistag probability for $W$+light-parton events in this sample is found to be approximately 0.3%,
which agrees with the value obtained from simulation.

From the SV-tagger BDT fits, the $b$ and $c$ yields are obtained in bins of $\sqrt{s}$, muon charge, and $p_T(\mu)/p_T(j_\mu)$. The $p_T(\mu)/p_T(j_\mu)$ distributions for muons associated with $b$-tagged and $c$-tagged jets are shown in Figs. 5 and 6. These distributions are fitted to determine the $W+b$ and $W+c$ final-state yields as in the inclusive $W+$jet sample. The $Z+b$ and $Z+c$ yields are obtained by fitting the SV-tagger BDT distributions in the fully reconstructed $Z+$jet data samples and then correcting for the missed-muon probability. The fits are shown in Figs. 5 and 6 for each muon charge and center-of-mass energy. The yields obtained still include contributions from top quark production and $Z\rightarrow\tau\tau$.

The $Z\rightarrow\tau\tau$ background, where one $\tau$ lepton decays into a muon and the other into a hadronic jet, contaminates the $W+c$ sample due to the similarity of the $c$-hadron and $\tau$ lepton masses. The $p_T(SV)/p_T(j)$ distribution, where $p_T(SV)$ is the transverse momentum of the particles that form the SV, is used to discriminate between $c$ and $\tau$ jets, since SVs produced from $\tau$ decays usually carry a larger fraction of the jet energy than SVs from $c$-hadron decays. Figure 7 shows fits to the $p_T(SV)/p_T(j)$ distributions observed in data where the $b$ and light-parton yields are fixed using the results of BDT fits performed on the data samples. A requirement of $\text{BDT}(bc|udsg) > 0.2$ is applied to this sample to remove the majority of SV-tagged light-parton jets while retaining 90% of $b$, $c$ and $\tau$ jets. The only free parameter in these fits is the fraction of jets identified as charm in the SV-tagger BDT fits that originate from $\tau$ leptons. The $p_T(SV)/p_T(j)$ templates are obtained from simulation. The $Z\rightarrow\tau\tau$ yields are consistent with SM expectations and are about 25 times smaller than the $W+c$ yields. These results are extrapolated to the inclusive sample using simulation.

The top quark background is determined in the dedicated analysis of Ref. [22], where a reduced fiducial region is used to enrich the relative top quark content. The yields
Figure 5: Fits to $p_T(\mu)/p_T(j_\mu)$ distributions for $b$-tagged data samples for $\sqrt{s} = 7$ and 8 TeV.

and charge asymmetries of the $W + b$ final state as functions of $p_T(\mu + b)$ are used to discriminate between $W + b$ and top quark production. The results obtained in Ref. [22] are consistent with SM expectations and are extrapolated to the fiducial region of this analysis using simulation based on NLO calculations. The extrapolated top quark yields are subtracted from the observed number of $W + b$ candidates to obtain the signal yields. Top quark production is found to be responsible for about 1/3 of events that contain a $W$ boson and $b$ jet. A summary of all signal yields is given in Table 1.

5 Systematic uncertainties

A summary of the relative systematic uncertainties separated by source for each measurement is provided in Table 2. A detailed description of each contribution is given below.

The $p_T$ distributions of muons from $W$ and $Z$ bosons produced in association with
Figure 6: Fits to $p_T(\mu)/p_T(j_\mu)$ distributions for $c$-tagged data samples for $\sqrt{s} = 7$ and 8 TeV.

$b$, $c$ and light-parton jets are nearly identical. This results in a negligible uncertainty from muon trigger and reconstruction efficiency on cross section ratios involving only $W$ bosons. In the ratios $\sigma(W^+j)/\sigma(Zj)$ and $\sigma(W^-j)/\sigma(Zj)$, the muon from the $Z$ boson decay with the same charge as that from the $W$ decay is required to satisfy the same trigger and selection requirements as the $W$ boson muon, giving negligible uncertainty from the trigger and selection efficiency. The efficiency for reconstructing and selecting the additional muon from the $Z$ boson decay is obtained from the data-driven studies of Ref. [17]. A further data-driven correction is applied to account for the higher occupancy in events with jets [20]; a 2% systematic uncertainty is assigned to this correction.

The GEC efficiency is obtained following Ref. [20]: an alternative dimuon trigger requirement with a looser GEC is used to determine the fraction of events that are rejected. The GEC efficiencies for all final states are found to be consistent within a statistical precision of 1%, which is assigned as a systematic uncertainty. As a further check, the number of jets per event reconstructed in association with $W$ or $Z$ bosons is compared and found to be consistent.
Figure 7: Fits to the $p_T(SV)/p_T(jet)$ distributions in (left) 7 TeV and (right) 8 TeV data for candidates with $p_T(\mu)/p_T(j_\mu) > 0.9$ and BDT($bc|udsg$) > 0.2.

Table 1: Summary of signal yields. The two $Zj$ yields denote the charge of the muon on which the trigger requirement is made. The $Zj$ yields given are the numbers of candidates observed, while the $W$ boson yields are obtained from fits. The yield due to top quark production is subtracted in these results.

<table>
<thead>
<tr>
<th>Mode</th>
<th>$\mu^+$</th>
<th>$\mu^-$</th>
<th>$\mu^+$</th>
<th>$\mu^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Zj$</td>
<td>2364</td>
<td>2357</td>
<td>6680</td>
<td>6633</td>
</tr>
<tr>
<td>$Wj$</td>
<td>27 400 ± 500</td>
<td>17 500 ± 400</td>
<td>70 700 ± 1100</td>
<td>44 800 ± 800</td>
</tr>
<tr>
<td>$Wb$-tag</td>
<td>160 ± 31</td>
<td>51 ± 27</td>
<td>400 ± 43</td>
<td>236 ± 45</td>
</tr>
<tr>
<td>$Wc$-tag</td>
<td>295 ± 36</td>
<td>338 ± 31</td>
<td>795 ± 56</td>
<td>802 ± 55</td>
</tr>
</tbody>
</table>

The jet reconstruction efficiencies for heavy-flavor and light-parton jets in simulation are found to be consistent within 2%, which is assigned as a systematic uncertainty for flavor-dependencies in the jet-reconstruction efficiency. The jet $p_T$ detector response is studied with a data sample enriched in $b$ jets using SV tagging. The $p_T(SV)/p_T(jet)$ distribution observed in data is compared to templates obtained from simulation in bins of jet $p_T$. The resolution and scale in simulation for each jet $p_T$ bin are varied to find the best description of the data and to construct a data-driven unfolding matrix. The results obtained using this unfolding matrix are consistent with those obtained using a matrix determined by studies of $p_T$ balance in $Z+jet$ events [20], where no heavy-flavor tagging is applied. The unfolding corrections are at the percent level and their statistical precision is
assigned as the uncertainty.

The heavy-flavor tagging efficiencies are measured from data in Ref. [27], where a 10% uncertainty is assigned for $b$ and $c$ jets. The cross-check fits of Sect. 4 using the corrected mass and track multiplicity, remove information associated with jet quantities, such as $p_T$, from the yield determination and produce yields consistent at the 5% level. This is assigned as the uncertainty for the SV-tagged yield determination.

The $W$ boson template for the $p_T(\mu)/p_T(j_\mu)$ distribution is derived from data, as described in Sect. 4. The fit is repeated using variations of this template, e.g. using a template taken directly from simulation and using separate templates for $W^+$ and $W^-$, to assess a systematic uncertainty. The dijet templates are obtained from data in a dijet-enriched region. The residual, small $W$ boson contamination is subtracted using two methods: the $W$ boson yield expected in the dijet-enriched region is taken from simulation; and the $p_T(\mu)/p_T(j_\mu)$ distribution in the dijet-enriched region is fitted to a parametric function to estimate the $W$ boson yield. The difference in the $W$ boson yields obtained using these two sets of dijet templates is at most 2%. The uncertainty on $W/Z$ ratios due to the $W$ boson and dijet templates is 4%. The uncertainty due to the $W$ boson template cancels to good approximation in the measurements of $\sigma(Wb)/\sigma(Wj)$ and $\sigma(Wc)/\sigma(Wj)$; however, the uncertainty due to the dijet templates is larger due to the enhanced dijet background levels. Variations of the dijet templates are considered, with 10% and 5% uncertainties assigned on $\sigma(Wb)/\sigma(Wj)$ and $\sigma(Wc)/\sigma(Wj)$.

The systematic uncertainty from top quark production is taken from Ref. [22], while the systematic uncertainty from $Z \to \tau\tau$ is evaluated by fitting the data using variations of the $p_T(SV)/p_T(j_\mu)$ templates. All other electroweak backgrounds are found to be negligible from NLO predictions. All $W \to \mu\nu$ yields have a small contamination from $W \to \tau \to \mu$ decays that cancels in all cross section ratios except for the $W/Z$ ratios. A scaling factor of 0.975, obtained from simulation, is applied to the $W$ boson yields. A 1% uncertainty is assigned to the scale factor, which is obtained from the difference between the correction factor from simulation and a data-driven study of this background [16] for inclusive $W \to \mu\nu$ production.

The trigger, reconstruction and selection requirements are consistent with being charge symmetric [16], which results in negligible uncertainty on $A(Wb)$ and $A(Wc)$. Unfolding of the jet $p_T$ detector response is performed independently for $W^+$ and $W^-$ bosons, with the statistical uncertainties on the corrections to the charge asymmetries assigned as systematic uncertainties. The uncertainty on the $W+b$ and $W+c$ yields from the BDT templates is included in the charge asymmetry uncertainty due to the fact that the fractional jet content of the SV-tagged samples is charge dependent. The uncertainty on the charge asymmetries due to determination of the $W$ boson yields is evaluated using an alternative method for obtaining the charge asymmetries. The raw charge asymmetry in the $b$-jet and $c$-jet yields in the $p_T(\mu)/p_T(j_\mu) > 0.9$ region is obtained from the SV-tagger BDT fits. The $Z$+jet and dijet backgrounds are charge symmetric at the percent level and contribute at most to 20% of the events in this $p_T(\mu)/p_T(j_\mu)$ region. Therefore, $A(Wb)$ and $A(Wc)$ are approximated by scaling the raw asymmetries by the inverse of the $W$ boson purity in the $p_T(\mu)/p_T(j_\mu) > 0.9$ region. A small correction must also be applied
Table 3: Summary of the results and SM predictions. For each measurement the first uncertainty is statistical, while the second is systematic. All results are reported within a fiducial region that requires a jet with $p_T > 20$ GeV in the pseudorapidity range $2.2 < \eta < 4.2$, a muon with $p_T > 20$ GeV in the pseudorapidity range $2.0 < \eta < 4.5$, $p_T(\mu + j) > 20$ GeV, and $\Delta R(\mu, j) > 0.5$. For $Z$+jet events both muons must fulfill the muon requirements and $60 < M(\mu\mu) < 120$ GeV; the $Z$+jet fiducial region does not require $p_T(\mu + j) > 20$ GeV.

<table>
<thead>
<tr>
<th>Source</th>
<th>$\sigma(Wb)/\sigma(Wj)$</th>
<th>$\sigma(Wc)/\sigma(Wj)$</th>
<th>$\sigma(Wj)/\sigma(Zj)$</th>
<th>$A(Wb)$</th>
<th>$A(Wc)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muon trigger and selection</td>
<td>–</td>
<td>–</td>
<td>2%</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>GEC</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Jet reconstruction</td>
<td>2%</td>
<td>2%</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Jet $p_T$</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$(b, c)$-tag efficiency</td>
<td>10%</td>
<td>10%</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SV-tag BDT templates</td>
<td>5%</td>
<td>5%</td>
<td>–</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$p_T(\mu)/p_T(j_\mu)$ templates</td>
<td>10%</td>
<td>5%</td>
<td>4%</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Top quark</td>
<td>13%</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.02</td>
</tr>
<tr>
<td>$Z \rightarrow \tau\tau$</td>
<td>–</td>
<td>3%</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Other electroweak</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$W \rightarrow \tau \rightarrow \mu$</td>
<td>–</td>
<td>1%</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>20%</td>
<td>13%</td>
<td>5%</td>
<td>0.09</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Table 2:** Systematic uncertainties. Relative uncertainties are given for cross section ratios and absolute uncertainties for charge asymmetries.

<table>
<thead>
<tr>
<th>Source</th>
<th>$\sigma(Wb)/\sigma(Wj)$</th>
<th>$\sigma(Wc)/\sigma(Wj)$</th>
<th>$\sigma(Wj)/\sigma(Zj)$</th>
<th>$A(Wb)$</th>
<th>$A(Wc)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muon trigger and selection</td>
<td>–</td>
<td>–</td>
<td>2%</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>GEC</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Jet reconstruction</td>
<td>2%</td>
<td>2%</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Jet $p_T$</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$(b, c)$-tag efficiency</td>
<td>10%</td>
<td>10%</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SV-tag BDT templates</td>
<td>5%</td>
<td>5%</td>
<td>–</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$p_T(\mu)/p_T(j_\mu)$ templates</td>
<td>10%</td>
<td>5%</td>
<td>4%</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Top quark</td>
<td>13%</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.02</td>
</tr>
<tr>
<td>$Z \rightarrow \tau\tau$</td>
<td>–</td>
<td>3%</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Other electroweak</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$W \rightarrow \tau \rightarrow \mu$</td>
<td>–</td>
<td>1%</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>20%</td>
<td>13%</td>
<td>5%</td>
<td>0.09</td>
<td>0.04</td>
</tr>
</tbody>
</table>

The difference between the asymmetries from this method and the nominal method is assigned as a systematic uncertainty from $W$ boson signal determination. The uncertainty on $A(Wb)$ due to top quark production is taken from Ref. [22].
6 Results

The results for $\sqrt{s} = 7$ and 8 TeV are summarized in Table 3. Each result is compared to SM predictions calculated at NLO using MCFM [41] and the CT10 PDF set [42] as described in Sect. 2. Production of $W+\text{jet}$ events in the forward region requires a large imbalance in $x$ of the initial partons. In the four-flavor scheme at leading order, $W+b$ production proceeds via $q\bar{q} \to Wg(b\bar{b})$, where the charge of the $W$ boson has the same sign as that of the initial parton with larger $x$. Therefore, $A(Wb) \approx +1/3$ is predicted due to the valence quark content of the proton. The dominant mechanism for $W+c$ production is $gs \to Wc$, which is charge symmetric assuming symmetric $s$ and $\bar{s}$ quark PDFs. However, the Cabibbo-suppressed contribution from $gd \to Wc$ leads to a prediction of a small negative value for $A(Wc)$.

The $\sigma(Wb)/\sigma(Wj)$ ratio in conjunction with the $W+b$ charge asymmetry is consistent with MCFM calculations performed in the four-flavor scheme, where $W+b$ production is primarily from gluon splitting. This scheme assumes no intrinsic $b$ quark content in the proton. The data do not support a large contribution from intrinsic $b$ quark content in the proton but the precision is not sufficient to rule out such a contribution at $O(10\%)$. The ratio $[\sigma(Wb) + \sigma(\text{top})]/\sigma(Wj)$ is measured to be $1.17 \pm 0.13 \text{ (stat)} \pm 0.18 \text{ (syst)\%}$ at $\sqrt{s} = 7$ TeV and $1.29 \pm 0.08 \text{ (stat)} \pm 0.19 \text{ (syst)\%}$ at $\sqrt{s} = 8$ TeV, which agree with the NLO SM predictions of $1.23 \pm 0.24\%$ and $1.38 \pm 0.26\%$, respectively.

The $\sigma(Wc)/\sigma(Wj)$ ratio is much larger than $\sigma(Wb)/\sigma(Wj)$, which is consistent with $Wc$ production from intrinsic $s$ quark content of the proton. The measured charge asymmetry for $W+c$ is about $2\sigma$ smaller than the predicted value obtained with CT10, which assumes symmetric $s$ and $\bar{s}$ quark PDFs. This could suggest a larger than expected contribution from scattering off of strange quarks or a charge asymmetry between $s$ and $\bar{s}$ quarks in the proton. The ratio $\sigma(W^+j)/\sigma(Zj)$ is consistent within $1\sigma$ with NLO predictions, while the observed $\sigma(W^-j)/\sigma(Zj)$ ratio is higher than the predicted value by about $1.5\sigma$.

Acknowledgments

We express our gratitude to our colleagues in the CERN accelerator departments for the excellent performance of the LHC. We thank the technical and administrative staff at the LHCb institutes. We acknowledge support from CERN and from the national agencies: CAPES, CNPq, FAPERJ and FINEP (Brazil); NSFC (China); CNRS/IN2P3 (France); BMBF, DFG, HGF and MPG (Germany); INFN (Italy); FOM and NWO (The Netherlands); MNISW and NCN (Poland); MEN/IFA (Romania); MinES and FANO (Russia); MinECo (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); NSF (USA). The Tier1 computing centres are supported by IN2P3 (France), KIT and BMBF (Germany), INFN (Italy), NWO and SURF (The Netherlands), PIC (Spain), GridPP (United Kingdom); NSF (USA). The Tier1 computing centres are supported by IN2P3 (France), KIT and BMBF (Germany), INFN (Italy), NWO and SURF (The Netherlands), PIC (Spain), GridPP (United Kingdom). We are indebted to the communities behind the multiple open source software packages on which we depend. We are also thankful for the computing resources and the access to software R&D tools provided by Yandex LLC.
(Russia). Individual groups or members have received support from EPLANET, Marie Skłodowska-Curie Actions and ERC (European Union), Conseil général de Haute-Savoie, Labex ENIGMASS and OCEVU, Région Auvergne (France), RFBR (Russia), XuntaGal and GENCAT (Spain), Royal Society and Royal Commission for the Exhibition of 1851 (United Kingdom).

References


LHCb collaboration


1 Centro Brasileiro de Pesquisas Físicas (CBPF), Rio de Janeiro, Brazil
2 Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
3 Center for High Energy Physics, Tsinghua University, Beijing, China
4 LAPP, Université Savoie Mont-Blanc, CNRS/IN2P3, Annecy-Le-Vieux, France
5 Clermont Université, Université Blaise Pascal, CNRS/IN2P3, LPC, Clermont-Ferrand, France
6 CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France
7 LAL, Université Paris-Sud, CNRS/IN2P3, Orsay, France
8 LPNHE, Université Pierre et Marie Curie, Université Paris Diderot, CNRS/IN2P3, Paris, France
9 Fakultät Physik, Technische Universität Dortmund, Dortmund, Germany
10 Max-Planck-Institut für Kernphysik (MPIK), Heidelberg, Germany
11 Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany
12 School of Physics, University College Dublin, Dublin, Ireland
13 Sezione INFN di Bari, Bari, Italy
14 Sezione INFN di Bologna, Bologna, Italy
15 Sezione INFN di Cagliari, Cagliari, Italy
16 Sezione INFN di Ferrara, Ferrara, Italy
17 Sezione INFN di Firenze, Firenze, Italy

22
Universidade Federal do Triângulo Mineiro (UFTM), Uberaba-MG, Brazil
P.N. Lebedev Physical Institute, Russian Academy of Science (LPI RAS), Moscow, Russia
Università di Bari, Bari, Italy
Università di Bologna, Bologna, Italy
Università di Cagliari, Cagliari, Italy
Università di Ferrara, Ferrara, Italy
Università di Firenze, Firenze, Italy
Università di Urbino, Urbino, Italy
Università di Modena e Reggio Emilia, Modena, Italy
Università di Genova, Genova, Italy
Università di Milano Bicocca, Milano, Italy
Università di Roma Tor Vergata, Roma, Italy
Università di Roma La Sapienza, Roma, Italy
Università della Basilicata, Potenza, Italy
AGH - University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland
LIFAELS, La Salle, Universitat Ramon Llull, Barcelona, Spain
Hanoi University of Science, Hanoi, Viet Nam
Università di Padova, Padova, Italy
Università di Pisa, Pisa, Italy
Scuola Normale Superiore, Pisa, Italy
Università degli Studi di Milano, Milano, Italy
Politecnico di Milano, Milano, Italy

† Deceased