Measurement of forward $t\bar{t}$, $W + b\bar{b}$ and $W + c\bar{c}$ production in $pp$ collisions at $\sqrt{s} = 8$ TeV

The LHCb collaboration

A R T I C L E   I N FO

Article history:
Received 26 October 2016
Received in revised form 12 January 2017
Accepted 20 January 2017
Available online 25 January 2017
Editor: L. Rolandi

A B S T R A C T

The production of $t\bar{t}$, $W + b\bar{b}$ and $W + c\bar{c}$ is studied in the forward region of proton-proton collisions collected at a centre-of-mass energy of 8 TeV by the LHCb experiment, corresponding to an integrated luminosity of $1.98 \pm 0.02$ fb$^{-1}$. The $W$ bosons are reconstructed in the decays $W \rightarrow \ell \nu$, where $\ell$ denotes muon or electron, while the $b$ and $c$ quarks are reconstructed as jets. All measured cross-sections are in agreement with next-to-leading-order Standard Model predictions.

1. Introduction

The production of $t\bar{t}$ pairs from proton–proton ($pp$) collisions in the forward region is of considerable interest, as it may be sensitive to physics beyond the Standard Model (SM) [1]. Furthermore, forward $t\bar{t}$ events can be used to constrain the gluon parton distribution function (PDF) at large momentum fraction [2]. The $t\bar{t}$ cross-section has been measured at ATLAS and CMS using several final states and at various centre-of-mass energies [3–5]. LHCb has also measured top quark production in the forward region in the $W + b$ final state [6].

Measurements of the production cross-sections of $W + b\bar{b}$ and $W + c\bar{c}$ in the forward region provide experimental tests of perturbative quantum chromodynamics (pQCD) [7–9], in a complementary phase space region to ATLAS and CMS. Previous studies of the $W + b\bar{b}$ final state have been performed by ATLAS [10] and CMS [11,12] at centre-of-mass energies $\sqrt{s} = 7$ TeV and 8 TeV. LHCb has previously performed measurements of the production cross-sections of a $W$ boson with at least one observed $b$ or $c$ jet [13] at 7 and 8 TeV, and a $Z$ boson with at least one $b$ jet at 7 TeV [14].

This Letter reports a study of events containing one isolated lepton (muon or electron) and two heavy-flavour tagged jets to measure the production cross-sections of $t\bar{t}$, $W^{+} + b\bar{b}$, $W^{-} + b\bar{b}$, $W^{+} + c\bar{c}$ and $W^{-} + c\bar{c}$. The study of $W + c\bar{c}$ is the first of its kind. Measurements are performed using a data sample corresponding to an integrated luminosity of $1.98 \pm 0.02$ fb$^{-1}$ of $pp$ collisions recorded at 8 TeV during 2012 by the LHCb experiment.

2. The LHCb detector and samples

The LHCb detector [15,16] is a single-arm forward spectrometer fully instrumented in 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, a silicon-strip detector located upstream of a dipole magnet with a bending power of about 4Tm, and three stations of silicon-strip detectors and straw drift tubes 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 (PV), the impact parameter (IP), is measured with a resolution of $(15 + 29/p_T) \mu m$, where $p_T$ is the component of the momentum transverse to the beam, 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 (SPD) and preshower (PRS) detectors, an electromagnetic calorimeter and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multilayer proportional chambers. The online event selection is performed by a trigger, which 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. The $W \rightarrow \mu \nu$ candidates are required to satisfy the hardware trigger requirement for muons, of having hits in the muon system corresponding to a high transverse momentum particle, and to satisfy the software trigger requirement of $p_T(\mu) > 10$ GeV. The $W \rightarrow e\nu$ candidates are required to satisfy the hardware trigger requirement for electrons of having an electromagnetic cluster of high transverse energy associated with signals in the PRS and SPD detectors, and the software trigger, which selects events with an
electron with $p_T(e) > 15$ GeV. A global event cut (GEC) on the number of hits in the SPD is applied in order to prevent high-multiplicity events from dominating the processing time of the reconstruction code.

Simulated event samples of $W +$ jets, $Z +$ jets, $\ell \ell$, single-top and diboson ($WZ$, $ZZ$) production are generated using Pythia 8 [17] with a specific LHCb configuration [18]. Event samples of $W + b\bar{b}$, $W + c\bar{c}$ and $Z + c\bar{c}$ production are generated with Alpgen [19], which includes tree-level contributions with up to four additional emissions of final state partons with respect to the leading-order diagram. Pythia 8 is used to perform the hadronisation for these samples. The cross-sections of the simulated processes are calculated at next-to-leading-order (NLO) including spin correlation effects with MCFM [20] using the CT10 PDF set [21]. Decays of hadronic particles are described by EvtGen [22], in which final-state radiation is generated using Photos [23]. The interaction of the generated particles with the detector, and its response, are implemented using the Geant4 toolkit [24] as described in Ref. [25]. Since neither showering nor hadronisation are included in MCFM, an overall correction is calculated to compare the measurements with the predicted cross-section at particle-level. This is done by generating $W + b\bar{b}$, $W + c\bar{c}$ and $\ell \ell$ events with Pythia 8 with the CT10 PDF set [21] where the same acceptance requirements are applied. The particle-level lepton momentum used here is the momentum after final-state radiation as implemented in Pythia 8.

3. Event selection

Events are selected by requiring the presence of either a high-$p_T$ muon or electron and two heavy-flavour tagged jets. The same fiducial definition for lepton and jets used in previous studies [6,13,26] is applied. The lepton must have $p_T(\ell) > 20$ GeV and $2.0 < \eta(\ell) < \eta_{\text{max}}$, where $\eta_{\text{max}}$ is 4.50 for a muon candidate, corresponding to the muon identification system acceptance, and is 4.25 for an electron candidate, corresponding to the electromagnetic calorimeter acceptance. The jets are required to have $p_T(j) > 12.5$ GeV and $2.2 < \eta(j) < 4.2$. Due to the limited sample size to validate the heavy-flavour tagging algorithm, for higher $p_T$ jets [27], only jets with $p_T(j) < 100$ GeV are considered. The lepton is required to be isolated from both jets using $\Delta R(\ell, j) > 0.5$, where $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$ is the distance between them in $\eta$-$\phi$ space and $\phi$ is the azimuthal angle. This requirement serves to remove the background formed by leptons coming from the same parton as the jets. The jets are also required to have $\Delta R(j_1, j_2) > 0.5$, where $j_1$ ($j_2$) is the highest (second highest) $p_T$ jet of the pair. Events with $p_T^{\text{miss}} < 15$ GeV, where $p_T^{\text{miss}}$ is the transverse component of $(\vec{p}(\ell) + \vec{p}(j_1) + \vec{p}(j_2))$, are removed to reduce the contamination from events not containing a $W$ boson. If more than one $(\ell + j_1 + j_2)$ candidate is found in the event, the candidate with highest $p_T^{\text{miss}}$ is selected.

Jets are reconstructed using a particle flow algorithm [28] and clustered using the anti-$k_T$ algorithm [29] with distance parameter $R = 0.5$ as implemented in the FastJet software package [30]. As in Ref. [28], the jet energy is corrected to the particle level, excluding neutrinos, and the same jet quality requirements are applied. Jets are heavy-flavour tagged, i.e. as originating from a $b$ or $c$ quark, by the presence of a secondary vertex (SV) with $\Delta R < 0.5$ between the jet axis and the direction of flight of the heavy-flavour hadron candidate, defined by the vector from the PV to the SV position.

The SV-tagger algorithm, described in detail in Ref. [27], uses two boosted decision trees (BDTs) [31,32]: one that separates heavy-flavour from light-parton jets (BDT(bc|uds)) and one that separates b jets from c jets (BDT(bc)). Both jets used in the analysis are required to have $\text{BDT}(bc|uds) > 0.2$, which gives a heavy-flavour tagging efficiency of about 50% (20%) for $b$ ($c$) jets and a misidentification probability of about 0.1% for light jets.

In order to suppress the $Z +$ jets background, events with an additional oppositely charged high-$p_T$ lepton that fulfills the lepton requirements described above are vetoed. Backgrounds from misidentified leptons or semileptonic decays of heavy-flavour hadrons are suppressed by two requirements applied to the lepton: $|p_T(\ell)/\sqrt{\text{MET}}| > 0.8$, where $j_\ell$ is defined as a reconstructed jet with relaxed quality criteria that contains the lepton.

4. Backgrounds

In both the electron and muon channels, the background processes include $Z + b\bar{b}$ and $Z + c\bar{c}$ production with $Z \rightarrow \mu\mu$ or $Z \rightarrow ee$, where one of the final state leptons is not reconstructed. $Z(\rightarrow \tau\tau)$ and $b\bar{b}$ production is also considered, where at least one $\tau$ decays to an electron or a muon. A small contribution of $Z \rightarrow \tau\tau$ produced in association with one $b$ or $c$ jet is also included. Other processes of $Z$ production associated to jets are negligible. Background contributions from $W(\rightarrow \ell v) +$ jets where the event does not contain two $b$ jets, and $W(\rightarrow \ell v) + b\bar{b}$ where $\ell$ decays to an electron or muon are also included. Single-top, $W(\rightarrow \ell v)Z(\rightarrow b\bar{b})$ and $Z(\rightarrow \ell\ell)Z(\rightarrow b\bar{b})$ production are considered as background processes. The expected yields of the background processes described above are obtained from NLO cross-sections. Weight factors are applied to compensate for residual differences between data and simulation for GEC, trigger and heavy-flavour tagging efficiencies. Further details about these factors and their uncertainties are given in Sec. 5.4.

The QCD multi-jet background, which includes lepton misidentification and semileptonic decays of a beauty or charm hadron, is estimated by using events which fail the $p_T(\ell)/\sqrt{\text{MET}} > 0.8$ requirement. The QCD multi-jet background normalisation is adjusted in order to describe the event yield at $|p_T(\ell) > 0.4$ mm, after subtracting the non-QCD backgrounds obtained from simulation.

5. Signal yield determination

5.1. Overview

The data sample is split into four subsamples, according to the flavour and charge of the lepton ($\mu^+ e^-$). A simultaneous fit to the distributions of four variables is performed to determine the $kT$, $W^+ + b\bar{b}$, $W^- + b\bar{b}$, $W^+ + c\bar{c}$, and $W^- + c\bar{c}$ yields in each sample. The four variables used in the fit are the invariant mass of the two jets ($m_{jj}$), the response of a multivariate classifier trained to distinguish between $\ell\ell$ and $W + b\bar{b}$ events and the multivariate discriminant classifier for each jet, $j_1$ BDT(b(c)) and $j_2$ BDT(b(c)), trained to discriminate between $b$ and $c$ jets. The expected background components are obtained from simulation, with the exception of the QCD multijet background. The fitted signal yields are converted into cross-sections using simulation and data-driven efficiencies and the measurement of the integrated luminosity [33]. The systematic uncertainties are included as nuisance parameters in the fit and propagated to the final result.

5.2. Fit variables

While $W + b\bar{b}$ and $W + c\bar{c}$ processes can be disentangled using the BDT(b(c)) variables for both jets, the separation between $\ell\ell$ and $W + b\bar{b}$ or $W + c\bar{c}$ is obtained by using the $m_{jj}$ variable and a multivariate discriminant, uGB, constructed such that its response is minimally correlated with $m_{jj}$ [34]. The variables $m_{jj}$, $j_1$ BDT(b(c)) and $j_2$ BDT(b(c)) are found to be uncorrelated. The uGB response is trained in simulation using 11 kinematic variables of...
The lepton and jets: \( p_T(\ell), \eta(\ell), p_T(j_1), p_T(j_2), m(j_1), m(j_2), \Delta R(j_1, j_2), \Delta R(j_1, j_1), \Delta R(j_2, j_2) \) and \( \cos(\theta_0(\ell)), \) where \( \theta_0(\ell) \) is the lepton scattering angle in the dijet rest frame and \( jj \) represents the dijet system. The muon and electron decay channels are trained separately. Fig. 1 shows the correlation between the uGB and the \( m_{jj} \) variables. In the fit all variables are treated as uncorrelated; the effect of the observed small correlations is taken into account in the systematic uncertainties of the results.

5.3. Signal determination

A binned maximum likelihood fit is performed to determine the yields of \( \mu \bar{\tau}, \ W^+ + b\bar{b}, \ W^- + b\bar{b}, \ W^+ + c\bar{c} \) and \( W^- + c\bar{c}. \) The simulated background yields are normalised to NLO predictions and they are allowed to vary in the fit within their uncertainties. The QCD multijet background is normalised from a data-driven method as explained in Section 4. The fit is performed assuming the four variables \( (m_{jj}, \text{uGB}, j_1 \text{BDT}(b/c) \) and \( j_2 \text{BDT}(b/c)) \) to be uncorrelated.

The free parameters in the fit are the normalisation factors with respect to the SM predicted yields \( K(i), \) where \( i = \mu \bar{\tau}, W^+ + b\bar{b}, W^- + b\bar{b}, W^+ + c\bar{c}, W^- + c\bar{c}. \) The \( K(\mu \bar{\tau}) \) parameter is fitted using all four samples, while the others are fitted in each corresponding sample. The projections of the fit in each of the four samples are shown in Figs. 2–5, while the fit results are given in Table 1.

5.4. Systematic uncertainties

Systematic effects can impact the results in two ways: by affecting signal and background yields, or by altering template shapes used in the fits. The efficiency of the GEC is measured in a \( Z + \text{jet} \) sample selected with a looser trigger requirement [26] and a 2% uncertainty is assigned to account for the final-state dependence of the GEC efficiency observed in simulation. The systematic uncertainty on the integrated luminosity is 1.16% [33].

The lepton reconstruction and trigger efficiencies are studied using data-driven methods in \( Z \to \ell^+ \ell^- \) [35,36]. These studies show that data and simulation agree within 1.0–5.0% depending on \( \eta(\ell) \) and \( p_T(\ell), \) which is taken as systematic uncertainty. The uncertainty of the lepton kinematic efficiency, which includes the effect of final-state radiation, is neglected. The method described in Ref. [27] is used to assess the systematic uncertainty due to the errors of the heavy-flavour tagging efficiency weight-factor described in Sec. 4, which amounts to 5–10% depending on \( p_T(j). \)
Fig. 3. Projections of the simultaneous 4D-fit results for the $\mu^-$ sample: a) the dijet mass; b) the $uGB$ response; the BDT(b,c) of the c) leading and d) sub-leading jets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Projections of the simultaneous 4D-fit results for the $e^+$ sample: a) the dijet mass; b) the $uGB$ response; the BDT(b,c) of the c) leading and d) sub-leading jets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
The LHCb collaboration / Physics Letters B 767 (2017) 110–120

The systematic uncertainty of the jet energy calibration includes possible biases due to flavour dependence (2%), tracks not associated to a real particle (1.2%), track momentum resolution (1%) and residual differences between simulation and data due to pile-up and calorimeter response (1%) as described in Refs. [26,28]. The jet energy resolution at LHCb is modelled in simulation to an accuracy of about 10% [28,13]. The uncertainties related to the jet reconstruction and quality selection efficiencies are found to be below 2%. The jet-related systematic uncertainties affect both the template shapes and the expected yields.

The simulated background normalisations are predicted at NLO and they are affected by uncertainties on the PDF ($\delta_{PDF}$), on the strong coupling constant $\alpha_s$ ($\delta_{\alpha_s}$) and on the renormalisation and factorisation scale ($\delta_{scale}$). The PDF uncertainty is evaluated following the procedure of Ref. [37]. The influence of the uncertainty on the strong coupling constant is evaluated by calculating the cross-sections with PDF sets [21] using values of $\alpha_s(M_Z)$: 0.117, 0.118 and 0.119. The scale uncertainty is evaluated by calculating the cross-sections varying the renormalisation and factorisation scale by a factor of two. The total uncertainty is taken as $\sqrt{\delta_{PDF}^2 + \delta_{\alpha_s}^2 + \delta_{scale}^2}$ as done in Ref. [6] which translates to relative uncertainties on the signal yields in the range 3–10%. These theoretical uncertainties are also considered in the signal yields in the experimental acceptance.

The systematic uncertainties in the normalisations due to the limited size of the simulated samples are between 1 and 7%. The uncertainty on the normalisation of the QCD multi-jet background, taken from data, is found to have a negligible effect.

Possible correlation effects between the fitted variables are studied by using templates generated randomly from the analysis templates with or without correlations found in simulation. It is found that the correlation and the fit procedure can affect the final yields by up to 10%.

All significant systematic uncertainties are correlated between the four samples except for the uncertainty due to the finite size of the simulated samples, which affects each sample and process independently.

6. Results and conclusions

The production cross-sections for $t\bar{t}$, $W^+ + b\bar{b}$, $W^- + b\bar{b}$, $W^+ + c\bar{c}$ and $W^- + c\bar{c}$ are measured for $pp$ collisions at a centre-of-mass energy of 8 TeV corresponding to an integrated luminosity of 1.98 ± 0.02 fb$^{-1}$ of data collected in 2012 by the LHCb experiment. These production cross-sections are obtained as the product of the normalisation factors shown in Table 1 and the expected SM cross-sections. The muons (electrons) coming from the $W$ boson are required to have $2.0 < \eta(\ell) < 4.5$ ($2.0 < \eta(\ell) < 4.25$) and $p_T(\ell) > 20$ GeV, while the jets are required to have $2.2 < \eta(j) < 4.2$ and $p_T(j) > 12.5$ GeV. In addition, the transverse component of
\[ \sigma(W^+ + b\bar{b}) \]
\[ \sigma(W^- + b\bar{b}) \]
\[ \sigma(W^+ + c\bar{c}) \]
\[ \sigma(W^- + c\bar{c}) \]
\[ \sigma(t\bar{t}) \]

Fig. 6. Graphical representation of Table 2. The outer bars (light yellow) correspond to the total uncertainties of the measured cross-sections and the inner bars (dark yellow) correspond to the statistical uncertainties. Theoretical prediction is represented by the black markers and error bars, where inner and outer uncertainties represent the scale and the total errors respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

<table>
<thead>
<tr>
<th>Process</th>
<th>Expected [pb]</th>
<th>Observed [pb]</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W^+ + b\bar{b}$</td>
<td>0.081$^{+0.022+0.040}_{-0.015-0.018}$</td>
<td>0.121$^{+0.019+0.029}_{-0.018-0.020}$</td>
<td>7.1σ</td>
</tr>
<tr>
<td>$W^- + b\bar{b}$</td>
<td>0.056$^{+0.014+0.018}_{-0.016-0.020}$</td>
<td>0.093$^{+0.018+0.023}_{-0.018-0.020}$</td>
<td>5.6σ</td>
</tr>
<tr>
<td>$W^+ + c\bar{c}$</td>
<td>0.121$^{+0.014+0.060}_{-0.020-0.027}$</td>
<td>0.24$^{+0.088+0.088}_{-0.070-0.04}$</td>
<td>4.7σ</td>
</tr>
<tr>
<td>$W^- + c\bar{c}$</td>
<td>0.084$^{+0.021+0.027}_{-0.015-0.020}$</td>
<td>0.133$^{+0.073+0.050}_{-0.062-0.022}$</td>
<td>2.5σ</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>0.045$^{+0.008+0.012}_{-0.007-0.010}$</td>
<td>0.05$^{+0.02+0.02}_{-0.01-0.01}$</td>
<td>4.9σ</td>
</tr>
</tbody>
</table>

Table 3

Correlation matrix for the measured cross sections. The correlations are given in %.

<table>
<thead>
<tr>
<th>Process</th>
<th>$t\bar{t}$</th>
<th>$W^+ + b\bar{b}$</th>
<th>$W^- + b\bar{b}$</th>
<th>$W^+ + c\bar{c}$</th>
<th>$W^- + c\bar{c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>$W^+ + b\bar{b}$</td>
<td>39.02</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>$W^- + b\bar{b}$</td>
<td>35.10</td>
<td>58.62</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>$W^+ + c\bar{c}$</td>
<td>31.26</td>
<td>30.87</td>
<td>37.65</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>$W^- + c\bar{c}$</td>
<td>19.06</td>
<td>31.97</td>
<td>20.16</td>
<td>22.99</td>
<td>100.00</td>
</tr>
</tbody>
</table>

We acknowledge support from the national agencies: CAPES, CNPq, FAPERJ and FINPEP (Brazil); NSFC (China); CNRS/IN2P3 (France); BMBF, DFG and MPG (Germany); INFN (Italy); FOM and NWO (The Netherlands); MINISW and NCN (Poland); MEN/IFA (Romania); MinES and FASO (Russia); MINECO (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); NSF (USA). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (The Netherlands), PIC (Spain), GridPP (United Kingdom), RRCKI and Yandex LLC (Russia), CSCS (Switzerland), IFIN-HH (Romania), CBPF (Brazil), PL-GRID (Poland) and OSC (USA). We are indebted to the communities behind the multiplicity of open source software packages on which we depend. Individual groups or members have received support from AvH Foundation (Germany), EPLANET, Marie Skłodowska-Curie Actions and ERC (European Union), Consiglione Regionale di Trento (Italy), open source libraries (many), and the LHCb collaboration. We thank the technical and administrative staff at the LHCb institutes. We acknowledge the support from CERN and from the national agencies: CAPES, CNPq, FAPERJ and FINPEP (Brazil); NSFC (China); CNRS/IN2P3 (France); BMBF, DFG and MPG (Germany); INFN (Italy); FOM and NWO (The Netherlands); MINISW and NCN (Poland); MEN/IFA (Romania); MinES and FASO (Russia); MINECO (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); NSF (USA). We acknowledge the computing resources provided by the LHCb collaboration, an international collaboration supported by 100 collaborating institutes in 15 countries. The LHCb collaboration / Physics Letters B 767 (2017) 110–120

References

LHCb collaboration

Associated to Center for High Energy Physics, Tsinghua University, Beijing, China.

Associated to LPNHE, Université Pierre et Marie Curie, Université Paris Diderot, CNRS/IN2P3, Paris, France.

Associated to Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany.

Associated to Institute of Theoretical and Experimental Physics (ITEP), Moscow, Russia.

Associated to ICCUB, Universitat de Barcelona, Barcelona, Spain.

Associated to Nikhef National Institute for Subatomic Physics, Amsterdam, The Netherlands.

† Deceased.