W+Heavy Flavour Jet Measurements at CMS

Isobel Ojalvo[*†]
Author affiliation
E-mail: ojalvo@wisc.edu

The production of W bosons in association with b quarks is studied using proton-proton collisions at $\sqrt{s} = 7$ TeV in a data sample collected with the CMS experiment at the LHC corresponding to an integrated luminosity of 5.0 fb^{-1}. The $\mathrm{W} + \mathrm{b\bar{b}}$ events are selected in the $\mathrm{W} \to \mu\nu$ decay mode by requiring a muon with transverse momentum $p_{\mathrm{T}} > 25$ GeV and pseudorapidity $|\eta| < 2.1$, and exactly two b-tagged jets with $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.4$. The measured $\mathrm{W} + \mathrm{b\bar{b}}$ production cross section in the fiducial volume, calculated at the level of final-state particles, is 0.53 ± 0.05 (stat.) ± 0.09 (syst.) ± 0.06 (th.) ± 0.01 (lum.) pb, in agreement with the standard model prediction. In addition, kinematic distributions of the $\mathrm{W} + \mathrm{b\bar{b}}$ system are measured and found to be in agreement with the predictions of a simulation using MADGRAPH and PYTHIA.

The European Physical Society Conference on High Energy Physics -EPS-HEP2013
18-24 July 2013
Stockholm, Sweden

[*]Speaker.
[†]A footnote may follow.

PoS(EPS-HEP 2013)242

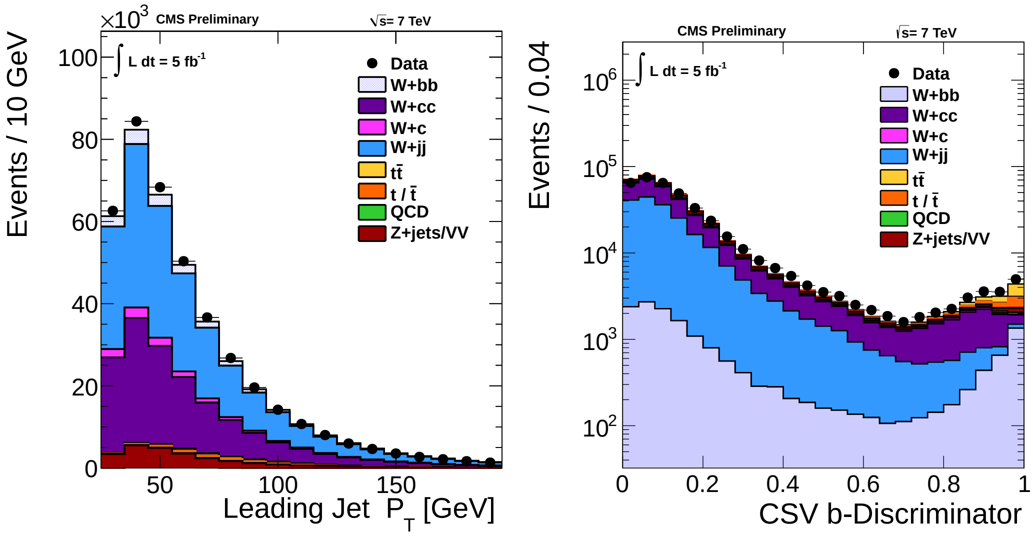

Figure 1: (left) The highest-p_T jet (J_1) before applying b-tagging. (right) The CSV b-discriminator for J_1.

This document describes a study of W boson production in association with two b quarks, where W bosons are observed via their decays to muons, and b quarks are identified as two separated jets and the study of a W boson plus one c Jet. Understanding the production of $W + b\bar{b}$ events is also crucial as they constitute a major background in the study of processes with two separated and well identified b jets in the final state, one of which is the search for the standard model Higgs boson (H) decaying to $b\bar{b}$ when produced in association with a weak vector boson.

These analyses use a sample of proton-proton collisions at a center-of-mass energy of $\sqrt{s}$=7 TeVcollected in 2011 with the CMS experiment at the LHC, corresponding to an integrated luminosity of $5.0 fb^{-1}$. While the CMS detector is described in detail elsewhere [4]. A number of Monte Carlo (MC) event generators are used to simulate the signal and backgrounds. The events with W or Z boson production, or with $t\bar{t}$ production, are generated at leading order (LO) with MADGRAPH5.1 [3], which is interfaced with PYTHIA6.4 [7] (also LO) for hadronization. Single top samples are generated at next-to-leading order (NLO) with POWHEG2.0 [2, 6, 5]. Diboson (WW, WZ, ZZ) and multijet samples are generated with PYTHIA6.4 [7].

Secondary vertices (SV) are reconstructed inside each jet. This study makes use of the combined secondary vertex (CSV) b-tagging algorithm [8]; this algorithm makes use of the long lifetime and high mass of b hadrons to provide optimized b-quark jet discrimination, by combining information about impact parameter significance, secondary vertex, and jet kinematics in a likelihood ratio technique. B-tagged jets are selected by imposing a minimum threshold on the CSV discriminator value. The analysis is based on a CSV discriminator threshold which provides an efficiency of approximately 50% for identifying jets containing b-flavored hadrons while reducing the misidentification probability for light-quark jets to 0.1% [1]

The $W + b\bar{b}$ selected events are required to have an isolated muon with $I^{\rm rel} < 0.12$, $p_T >$ 25GeV, $|\eta| < 2.1$, exactly two jets with $p_T >$ 25GeV and $|\eta| < 2.4$, where both selected jets must contain a secondary vertex and pass the b-tagging CSV requirement. To reduce the contribution from Z-boson production, the event is rejected if there is a second muon, without any requirements on the isolation and p_T, which builds with the isolated muon a dimuon system with invariant mass $m_{\mu\mu} > 60$GeV. The $t\bar{t}$ background is reduced by requiring that there are no additional isolated electrons or muons with $p_T > 20$GeV in the event and no jets with $p_T > 25$GeV and $2.4 < |\eta| < 4.5$. To reduce the contribution from QCD multijet events $M_T > 45$GeV is also required.

After all the selection requirements the significant background contributions are: $t\bar{t}$, single top, W+jets (u,d,c,s,g), Z+jets (u,d,c,s,b,g) and QCD multijet. Contributions of all these backgrounds are computed in a simultaneous fit, which provides a final estimate for the signal and background yields.

With the exception of QCD, the shapes of the background distributions are taken from simulation. A shape for the QCD contribution is obtained directly from data from a multijet-enriched control region defined by all the selection requirements, but requiring the muon to be non-isolated ($I^{rel} > 0.2$). The QCD uncertainty in the final fit is taken to be $\pm 50\%$. This uncertainty is large enough to provide coverage for normalization and shape mismodelings of the small QCD contribution in the final sample. The initial yields are taken either from data, in estimates based on the control regions, or from simulation, normalized to the NNLO predictions. The shapes and normalizations of the background distributions are validated in data with a set of control regions.

The W+jets (u,d,c,s,g) process, where the jets are not initiated by b quarks, is the dominant background before applying the selection requirements on the secondary vertex and b-tagging. Figure 1 (left) shows the p_T of the leading jet at this preselected stage. The CSV algorithm working point which provides maximum reduction of W+jets (u,d,c,s,g) is used. The CSV b-tagging discriminant for the leading jet is shown in Figure 1 (right). The presence of light and charm jets in the sample is very small at the higher values of the discriminant. Furthermore, to increase the purity of the sample a secondary vertex is required to be reconstructed in each of the selected jets. These selection requirements have been validated in the $t\bar{t}$ and Z+jets control regions described below.

Figure 2 (right) shows the invariant mass of the two highest-p_T additional jets (3rd and 4th highest-p_T in the event, $m_{J_3J_4}$).

The final yields are extracted via a binned maximum likelihood fit. To constrain the most prominent backgrounds and reduce the final systematic uncertainty the fit is performed simultaneously on the p_T of the leading jet (J_1) in the signal region after all selection requirements have been applied, and on the $m_{J_3J_4}$ distribution obtained from the $t\bar{t}$ control region. The J_1 p_T is chosen as the final fit variable due to its discrimination power against top-related backgrounds. Figure 2 shows the two fitted distributions, ${p_T}^{J_1}$ in the signal region (left) and $m_{J_3J_4}$ in the $t\bar{t}$ control region (right), normalized to the results of the fit.

The observed number of events in data after selection in the signal region is $N(S+B)_{data} = 1230 \pm 35$. The number of signal events obtained in the binned maximum likelihood fit is 300 ± 60.

To show the robustness of the $W+b\bar{b}$ fit result a separate study was performed with two selected b-tagged jets that require each jet to fulfill a looser CSV b-tagging criterion, corresponding to an efficiency of 70% for jets containing b-flavored hadrons, while the misidentification probability for light-quark jets is 1%. With the exception of the modification to the CSV threshold, all other selections for the signal and control region remain unchanged. The $W+c\bar{c}$ contribution is non-negligible with this selection, therefore, the sum of the invariant mass of the secondary vertex found in each selected jet is used to distinguish between $W+b\bar{b}$ and $W+c\bar{c}$. The scalar sum of

PoS(EPS-HEP 2013)242

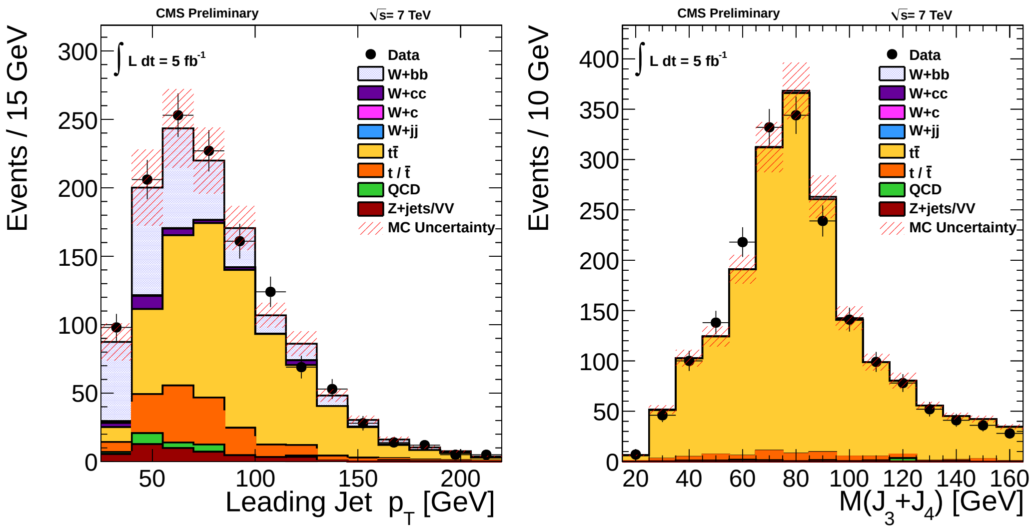

Figure 2: (left) The p_T distribution of the highest-p_T jet in the signal region, normalized to the result of the binned maximum likelihood fit. (right) The invariant mass of the two additional light jets in the $t\bar{t}$ control region, also normalized to the results of the fit.

Process	Prediction	Fitted Yield
$W+b\bar{b}$	332 ± 66	300 ± 60
$W+c$, $W+c\bar{c}$	21 ± 4	20 ± 4
W+usdg	1.5 ± 0.2	1 ± 1
Z+jets	31 ± 3	32 ± 3
$t\bar{t}$	596 ± 35	647 ± 52
Single top	160 ± 13	170 ± 13
WW, WZ	19 ± 3	17 ± 3
QCD	33 ± 17	33 ± 16
Total	1194 ± 78	1220 ± 82
Observed Events	1230 ± 35	

Table 1: Comparison of the expected (before the fit) and measured (after the fit) yields for each of the processes. The uncertainty on the Monte Carlo prediction takes into account the variation allowed to the nuisance parameters in the fit. The uncertainty in the fitted yields corresponds to the full uncertainty after the fit.

the transverse momenta of the muon, the $\vec{E}_T^{miss}$ and the jets, H_T, is used to distinguish W+jets from top contributions. The $W+b\bar{b}$ signal is extracted via a two dimensional fit of H_T versus the sum of the the secondary vertex masses of the highest- (J_1) and second-highest-p_T (J_2) jets. An equivalent $t\bar{t}$ control region to the one described in the tighter selection, based on the reconstruction of the W mass using two light jets, is also used in this case. The variables J_1 SV mass + J_2 SV mass and H_T are shown in Fig. 3, with yields normalized to the results of the fit. The cross section value computed with this alternative method is found to be consistent with the primary fit results quoted

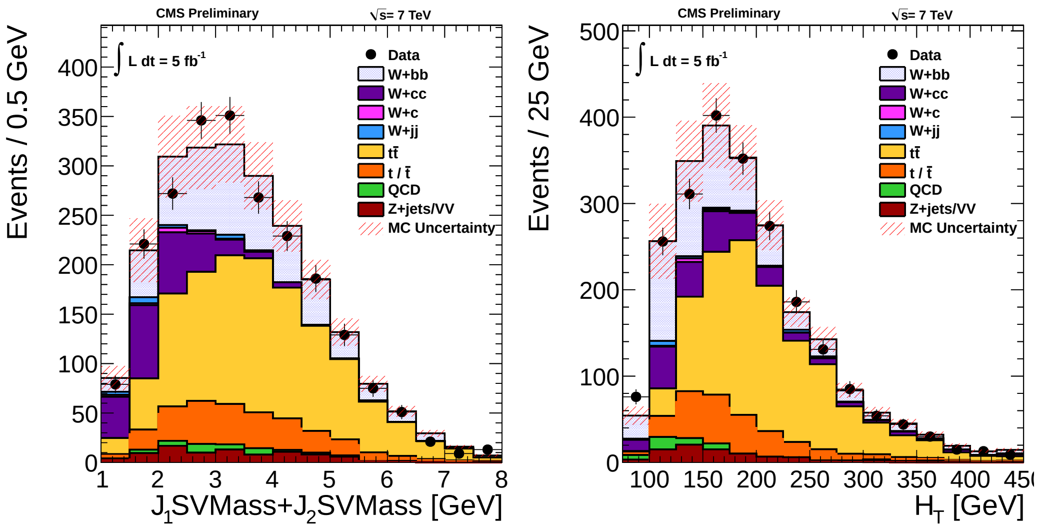

Figure 3: The distribution of the sum of the masses of the two secondary vertices (J_1 SV mass + J_2 SV mass) (left) and H_T of the system (right) in the alternative medium b-tag selection, normalized to the results of the cross-check fit.

above.

The $\mathrm{W}+\mathrm{b\bar{b}}$ cross section within the reference fiducial phase space is determined using the following equation:

$$\sigma(pp \to \mathrm{W}+\mathrm{b\bar{b}}) = \frac{N_{Data} - N_{Bckg}}{\mathscr{L}_{\mathrm{int}}\ \varepsilon_{sel}},$$

where the efficiency of the selection requirements $\varepsilon_{sel} = 10.4 \pm 1.0\%$ is estimated with MADGRAPH. The associated errors correspond to PDFs and to the factorization/matching scales.

The measured cross section is

$$\sigma(pp \to \mathrm{W}+\mathrm{b\bar{b}}, p_T^{\mathrm{b}} > 25\ \mathrm{GeV}, |\eta^{\mathrm{b}}| < 2.4) \times \mathscr{B}(\mathrm{W} \to \mu\nu, p_T^{\mu} > 25\ \mathrm{GeV}, |\eta^{\mu}| < 2.1) = \\ = 0.53 \pm 0.05\,(\mathrm{stat.})\ \pm 0.09\,(\mathrm{syst.})\ \pm 0.06\,(\mathrm{theo.})\ \pm 0.01\,(\mathrm{lum.})\ \mathrm{pb}.$$

This cross section is calculated at the level of final-state particles, by requiring a muon with $p_T > 25$GeV and $|\eta| < 2.1$ and exactly two jets, reconstructed using the anti-k_T jet algorithm with distance parameter 0.5, with $p_T > 25$GeV and $|\eta| < 2.4$ and each containing at least one b hadron with $p_T > 5$GeV. Events with extra jets are vetoed. A hadronization correction factor 0.92 ± 0.01 to extrapolate from the final-state particle jets to the parton-level cross section is estimated with MADGRAPH+PYTHIA. At the parton level the events are required to have a muon with $\mathrm{p_T} > 25$GeV and $|\eta| < 2.1$ and exactly two parton-jets with $\mathrm{p_T} > 25$GeV and $|\eta| < 2.4$ each containing a b parton just before hadronization. This factor has been computed in the 4-flavor and 5-flavor schemes, and the difference between said calculations is assigned as systematic uncertainty to the measurement. The simulated partons include double parton scattering and multiple parton interaction production of $\mathrm{b\bar{b}}$ pairs, which have been found to reproduce adequately these processes observed by CMS measurements.

Measurements were also presented of the associated production of a W boson and a charm-quark jet (W + c) in pp collisions at a center-of-mass energy of 7TeV. The analysis is conducted with a data sample corresponding to a total integrated luminosity of $5fb^{-1}$, collected by the CMS detector at the LHC. W boson candidates are identified by their decay into a charged lepton (muon or electron) and a neutrino. The W + c measurements are performed for charm-quark jets in the kinematic region $\mathrm{p_T}^{\mathrm{jet}} > 25\mathrm{GeV}$, $|\eta^{jet} < 2.5|$, for two different thresholds for the transverse momentum of the lepton from the W-boson decay, and in the pseudorapidity range $|\eta| < 2.1$. Hadronic and inclusive semileptonic decays of charm hadrons are used to measure the following total cross sections: $\sigma(\mathrm{pp} \to \mathrm{W} + \mathrm{c} + \mathrm{X}) \times \mathscr{B}(\mathrm{W} \to l\nu) = 107.7 \pm 3.3\,(\mathrm{stat.})\ \pm 6.9\,(\mathrm{syst.})\ pb$ $(\mathrm{p_T}^l > 25\mathrm{GeV})$ and $\sigma(\mathrm{pp} \to \mathrm{W} + \mathrm{c} + \mathrm{X}) \times \mathscr{B}(\mathrm{W} \to l\nu) = 84.1 \pm 2.0\,(\mathrm{stat.})\ \pm 4.9\,(\mathrm{syst.})\ pb$ $(\mathrm{p_T}^l > 35\mathrm{GeV})$, and the cross section ratios $\sigma(\mathrm{pp} \to \mathrm{W} + \mathrm{c} + \mathrm{X})/\sigma(\mathrm{pp} \to \mathrm{W} + \mathrm{c} + \mathrm{X}) = 0.954 \pm 0.025\,(\mathrm{stat.})\ \pm 0.004\,(\mathrm{syst.})$ and $\sigma(\mathrm{pp} \to \mathrm{W} + \mathrm{c} + \mathrm{X})/\sigma(\mathrm{pp} \to \mathrm{W} + \mathrm{c} + \mathrm{X}) = 0.938 \pm 0.019\,(\mathrm{stat.})\ \pm 0.006\,(\mathrm{syst.})$. Cross sections and cross section ratios are also measured differentially with respect to the absolute value of the pseudorapidity of the lepton from the W-boson decay. These are the first measurements from the LHC directly sensitive to the strange quark and antiquark content of the proton. Results are compared with theoretical predictions and are consistent with the predictions based on global fits of parton distribution functions.

References

[1] Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.

[2] S. Alioli, P. Nason, C. Oleari, and E. Re. NLO vector-boson production matched with shower in POWHEG. *JHEP*, 07:060, 2008.

[3] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer. Madgraph 5: going beyond. *JHEP*, 06:128, 2011.

[4] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[5] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[6] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[7] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.

[8] Christian Weiser. A combined secondary vertex based b-tagging algorithm in cms. (CMS-NOTE-2006-014), Jan 2006.

PoS(EPS-HEP 2013)242