CMS CR -2013/440

The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

29 November 2013 (v3, 02 December 2013)

Measurement of Jet Multiplicity Distributions in Semileptonic Top Quark Pair Events

Shawn Darrell Williamson for the CMS Collaboration

Abstract

The normalized differential cross section of $t\bar{t}$ events in the semileptonic decay channel in jet multiplicity bins is presented, measured in proton-proton collisions using $5.0\,\mathrm{fb}^{-1}$ of data collected at $\sqrt{s} = 7\,\mathrm{TeV}$. Additionally, the normalized differential cross-section of top quark pair production in bins of the number of radiated additional hard partons is measured. These measurements are sensitive to additional radiation at the top quark scale. Comparisons of data with predictions from different Monte Carlo generators or using different scale values for the predictions are provided. Good agreement is observed between data and the prediction from MADGRAPH + PYTHIA.

Presented at *TOP2013 6th International Workshop on Top Quark Physics*

Measurement of jet multiplicity distributions in semi-leptonic top quark pair events

Shawn Williamson on behalf of the CMS Collaboration

KIT, Wolfgang-Gaede-Str. 1, 76131 Karlsruhe, Germany

DOI: `will` be assigned

The inclusive $t\bar{t}$ production cross-section at a center of mass energy of 7 TeV has been measured by the ATLAS [1, 2] and CMS [3, 4] collaborations. The large amount of data recorded by the CMS detector [5] in 2011 enables an exclusive measurement of the jet multiplicity in $t\bar{t}$ production. Providing a test of perturbative QCD on the energy scale of top quark production is an important feature of this measurement. Furthermore, top quark pair production events including additional jets are an important background for many Higgs analysis and beyond the Standard Model searches. Anomalous production of additional jets in $t\bar{t}$ events could be a sign of physics beyond the Standard Model [6].

Two separate measurements of normalized differential cross sections accounting for the number of jets in top quark pair production events are presented [7]. The determination of the differential $t\bar{t}$ cross section as a function of the jet multiplicity has been performed in the semileptonic electron + jets and muon + jets channels. The measurement of the differential cross section with respect to the number of jets produced in addition to the top quark pairs is shown for the semileptonic muon + jets channel.

Top pair production events feature a top quark and a top anti-quark which, according to the Standard Model (SM), decay into a bottom quark and a W boson at almost 100 %. The produced W bosons can either decay leptonically into a charged lepton and a neutrino or hadronically into two light quarks. The semileptonic decay channel of the top quark pair features one leptonically decaying W boson and one decaying hadronically W boson. The corresponding signature contains an energetic charged lepton, missing transverse energy associated with a neutrino, and at least four jets, two originating from b quarks.

The full 2011 CMS data sample has been analyzed, corresponding to an integrated luminosity of $5.0\,\mathrm{fb}^{-1}$. The triggers used for this analysis demand a single electron ($p_T > 25\,\mathrm{GeV}$) and three jets ($p_T > 30\,\mathrm{GeV}$) in the electron + jets channel and a single isolated muon ($p_T > 24\,\mathrm{GeV}$) in the muon + jets channel. The signal process and smaller background processes rely on the prediction by MC simulation while the main backgrounds are estimated using data-driven methods. The $t\bar{t}$ signal events, the W + jets, and the Z + jets background events have been generated with the MadGraph v5.1.1 matrix element generator [8] interfaced to Pythia 6.424 for parton showering [9]. For the simulation of the single top background events a combination of Powheg v1.0 and Pythia has been used [10]. The diboson background events have been generated with Pythia. All generated samples are passed to a full detector simulation using Geant4 [11] which incorporates the generation of pileup events with a multiplicity matching the one observed in data.

The recorded and simulated events are reconstructed using the particle flow (PF) algorithm [12]. It identifies and reconstructs muons, electrons, photons, charged hadrons, and neutral hadrons produced by the proton-proton collisions including the requirement for charged particles to originate from the primary collision vertex. The reconstructed objects are clustered into jets by applying the anti-k_T algorithm with a distance parameter of 0.5 [13, 14]. Jet energy corrections are applied in order to maintain a relative uniform response of the calorimeter and a calibrated absolute response. The event selection applied matches the signature of the semileptonic decay channel demanding one isolated lepton (electron/muon) with $p_T > 30\,\mathrm{GeV}$ and $|\eta| < 2.5/2.1$. Further at least three reconstructed jets with $p_T > 35\,\mathrm{GeV}$ and $|\eta| < 2.4$ are required for the jet multiplicity measurement while the differential cross section measurement with respect to the number of additional jets requires at least four reconstructed jets with $p_T > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$. Two or more of the selected jets have to be tagged as originating from a bottom quark by the Combined Secondary Vertex (CSV) tagger, which combines reconstructed secondary vertices and track-based lifetime information, at a medium working point (1 % misidentification rate) [15].

The prediction for QCD multijet events is extracted from data based on the fact that QCD objects reconstructed as leptons are less isolated. Applying a modified event selection with an inverted lepton isolation cut provides a pure sample of QCD multijet events. The proper normalization of this sample is acquired by performing a maximum likelihood fit in a sideband region where fewer than two b jets are identified. The normalization factor resulting from the fit is used to scale the QCD multijet sample extracted from data with at least two b tagged jets.

The normalization of the W + jets background is modified using information from data. First its normalization is estimated making use of the charge asymmetry of this process at the LHC [16]. The valence quark composition of the colliding protons featuring two up quarks and one down quark favors the production of positively charged W bosons. The contributions of other processes to the charge asymmetry are negligible therefore the difference in the number of events with positively charged leptons (N^+) and negatively charged leptons (N^-) is proportional to the number of W + jets events.

Besides the normalization, a correction is applied improving the modeling of the heavy flavor fraction in W + jets events. This correction includes a reweighting of W + bX events with a factor 2 ± 1 and W + cX events with factor $1^{+1}_{-0.5}$ [17]. A subsequent renormalization step cancels any event yield changes caused by this last corrections.

The evaluation of systematic uncertainties on the measurement is performed by varying the process samples according to the respective uncertainties. The subsequent treatment differs for the two differential cross-section measurements. For the measurement concerning the jet multiplicity the cross-section calculation is repeated for all systematic variations. The measurement of the production of additional jets makes use of pseudo experiments randomly generating pseudo data sets from the varied samples. The following sources of systematic uncertainties are considered for this analysis:

Jet Energy Variation of the jet energy scale (JES) and of the jet energy resolution (JER) by their uncertainties.

Background

W + jets Variation of the W + jets corrections by the uncertainties of the heavy flavor correction and the normalization method.

QCD Variation of the estimated QCD multijet sample due to the choice of the lepton isolation cut and the remaining contamination.

Other backgrounds Variation of sample normalization by a conservative estimation of the uncertainty on the prediction (±30 %).

$\mathbf{Q^2}$ Scale Independent variation of the renormalization and factorization scale in MadGraph generation of $t\bar{t}$ and W + jets events.

ME/PS Matching Independent variation of the matrix-element/parton-showering jet-matching threshold in MadGraph generation of $t\bar{t}$ and W + jets events.

Parton Distribution Function (PDF) Reweighting of the default MadGraph $t\bar{t}$ sample according to the 44 CTEQ66 error PDF sets.

Remaining Systematic Uncertainties Uncertainties from the integrated luminosity estimation, from the trigger efficiency in MC, of the lepton identification and isolation, of the b-tagging efficiency, and of pileup according to the 2011 data scenario are considered.

The differential cross section with respect to the jet multiplicity is defined as

$$\frac{d\sigma_{t\bar{t}}^{\text{measured}}}{dN_{\text{jets}}} = \frac{N_{\text{data}}^{i} - N_{\text{bkg}}^{i}}{\epsilon^{i} \cdot \mathcal{L}} \, . \tag{1}$$

N_{data}^{i} (N_{bkg}^{i}) is the number of data (background) events containing i jets and fulfilling the full event selection. The factor $\epsilon^{i} = \frac{N_{\text{rec}}^{i}}{N_{\text{gen}}^{i}}$ corrects for bin-to-bin migration effects between the generator jets and the reconstructed jets. Generator jets are obtained by clustering stable particles applying the anti-k_T algorithm with a distance parameter of 0.5. N_{rec}^{i} is the number of $t\bar{t}$ signal events with i reconstructed jets and passing the full event selection while N_{gen}^{i} represents the number $t\bar{t}$ signal events with i generator jets with $p_T > 35\,\text{GeV}$ and $|\eta| < 2.4$. Dividing the differential cross section value for each jet multiplicity i by the measured inclusive cross section provides the normalized differential cross section

$$\frac{d\sigma}{\sigma_{t\bar{t}} \cdot dN_{\text{jets}}} = \frac{1}{\sigma_{t\bar{t}}^{\text{measured}}} \times \frac{d\sigma_{t\bar{t}}^{\text{measured}}}{dN_{\text{jets}}} \qquad \text{with} \ \ \sigma_{t\bar{t}}^{\text{measured}} = \sum^{i} \frac{d\sigma_{t\bar{t}}^{\text{measured}}}{dN_{\text{jets}}} \, . \tag{2}$$

The normalization of the differential cross section cancels out the systematic uncertainties correlated over all jet multiplicities. The differential cross section values for each jet multiplicity are calculated for both lepton channels and combined with the BLUE (Best Linear Unbiased Estimator) method [18]. All systematic uncertainties except for the ones concerning the lepton are assumed to be 100 % correlated for the combination. Figure 1 shows the results of this measurement compared to the prediction of different MC generator combinations (MadGraph+Pythia, Powheg+Pythia, and MC@NLO+Herwig) and different parameter variations of the nominal generator (Q^2 scale and ME/PS matching).

The measurement of jets produced in addition to top quark pairs is an alternative method to extract information about additional radiation. The determination of the $t\bar{t}$ production cross section differential in the number of additional jets is performed for the μ+jets channel. It starts by classifying the simulated $t\bar{t}$ events depending on the number of additional jets using

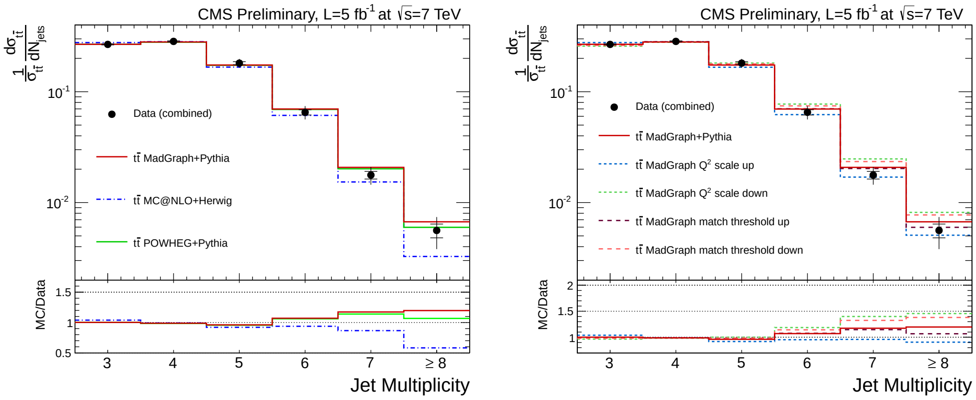

Figure 1: Measured normalized differential cross section of the $t\bar{t}$ production process as a function of the number of generator jets compared to MC expectations from different generators (left) and generator parameter variations (right).

MC truth information. Additional jets are defined by the absence of a top decay parton within the angular distance of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.5$ of the respective generator jet fulfilling the jet selection criteria ($p_T > 30\,\mathrm{GeV}$, $|\eta| < 2.4$). The matching of the top decay partons to generator jets enables the identification of hard additional radiation while low angle final state radiation is assigned to the original parton. The simulated $t\bar{t}$ sample is divided into three categories, $t\bar{t}$ + 0, 1, and ≥ 2 additional jets, according to the number of generator jets in the event matching the additional jet definition. The corresponding cross sections are extracted by performing a maximum likelihood fit of the signal and the background templates to data. The fit variable

$$\chi = \sqrt{\left(\frac{m^{\mathrm{rec}}_{W^{\mathrm{had}}} - m^{\mathrm{true}}_{W^{\mathrm{had}}}}{\sigma_{W^{\mathrm{had}}}}\right)^2 + \left(\frac{m^{\mathrm{rec}}_{t^{\mathrm{had}}} - m^{\mathrm{true}}_{t^{\mathrm{had}}}}{\sigma_{t^{\mathrm{had}}}}\right)^2 + \left(\frac{m^{\mathrm{rec}}_{t^{\mathrm{lep}}} - m^{\mathrm{true}}_{t^{\mathrm{lep}}}}{\sigma_{t^{\mathrm{lep}}}}\right)^2} \tag{3}$$

is based on the full reconstruction of the $t\bar{t}$ system. The reconstructed masses of the hadronic W boson and both top quarks are calculated for every possible combination of assigning jets to top decay partons with the restriction of always matching b-tagged jets to the bottom quarks. Subsequently the hypothesis with the lowest χ value is kept for each event. The true masses m^{true} and the widths $\sigma_{t^{had}}$, $\sigma_{t^{lep}}$ and $\sigma_{W^{had}}$ are determined from the $t\bar{t}$ MC simulation. Due to the sensitivity of the event reconstruction to the relation between the number of reconstructed jets and the number of additional jets, the variable χ provides good discrimination power between the $t\bar{t}$ + 0, 1, and ≥ 2 additional jets events containing the same number of reconstructed jets. Therefore performing a simultaneous fit in three exclusive jet multiplicity bins with events with 4, 5, and ≥ 6 reconstructed jets increases the sensitivity of this measurement. The normalization of the $t\bar{t}$ + 0, 1, and ≥ 2 additional jets signal samples are free parameters in the fit while Gaussian constraints corresponding to the uncertainties of the prediction are applied to the normalizations of the backgrounds.

The outcome of the maximum likelihood fit is shown in Figure 2. The extracted cross sections including a comparison with the prediction of different MC generators and different

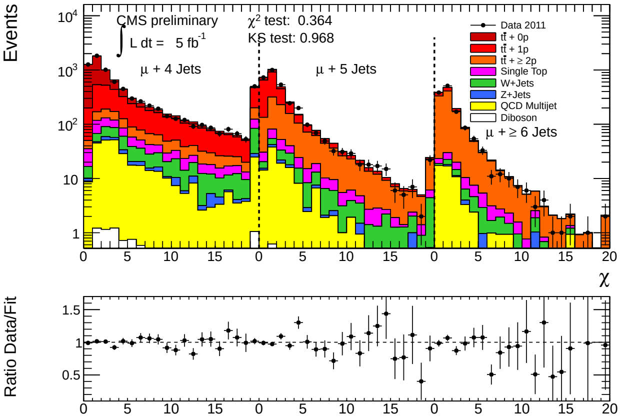

Figure 2: Result of cross section extraction fit simultaneous in three reconstructed jet multiplicity bins. All templates are scaled by the resulting fit parameters.

parameter variations of the nominal generator are displayed in Figure 3.

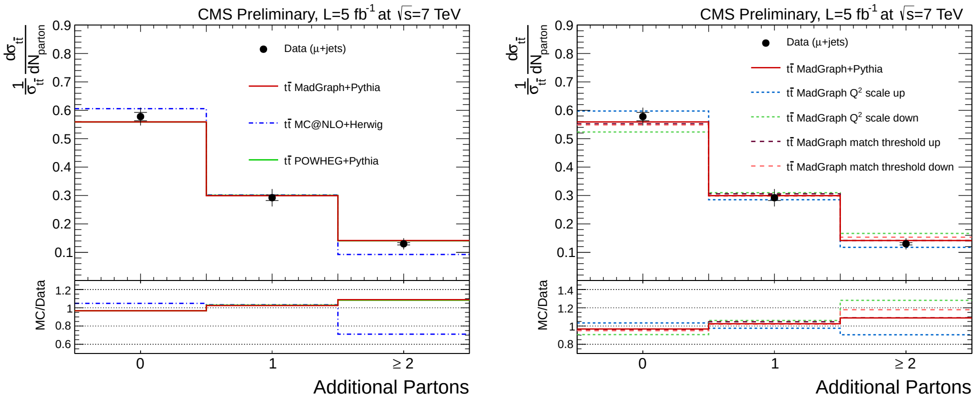

Figure 3: Measured normalized cross section of $t\bar{t}$ production processes with additional jets in the μ+jets channel compared with MC expectations from different generators (left) and generator parameter variations (right).

The comparison of the two differential cross section measurements shows a similar behavior of the results. Good agreement between the measured values and the predictions of the MC simulations can be found for the lower jet multiplicities. A slight discrepancy between the measurement and the predictions from MC@NLO+Herwig and the down variation of Q^2 and matching parameters in MadGraph+Pythia is observed for high jet multiplicities. Specific modeling configurations cannot be excluded due to the large uncertainties with main contributions coming from the jet energy scale uncertainty, MC generator parameters (Q^2 scale & ME/PS matching), and the PDF used for the signal simulation.

References

[1] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **711** (2012) 244 [arXiv:1201.1889 [hep-ex]].

[2] G. Aad *et al.* [ATLAS Collaboration], JHEP **1205** (2012) 059 [arXiv:1202.4892 [hep-ex]].

[3] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Rev. D **84** (2011) 092004 [arXiv:1108.3773 [hep-ex]].

[4] S. Chatrchyan *et al.* [CMS Collaboration], JHEP **1211** (2012) 067 [arXiv:1208.2671 [hep-ex]].

[5] S. Chatrchyan *et al.* [CMS Collaboration], JINST **3** (2008) S08004.

[6] M. I. Gresham, I. -W. Kim and K. M. Zurek, Phys. Rev. D **84** (2011) 034025 [arXiv:1102.0018 [hep-ph]].

[7] [CMS Collaboration], CMS-PAS-TOP-12-018.

[8] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, JHEP **1106** (2011) 128 [arXiv:1106.0522 [hep-ph]].

[9] T. Sjöstrand, S. Mrenna and P. Z. Skands, JHEP **0605** (2006) 026 [hep-ph/0603175].

[10] S. Alioli, P. Nason, C. Oleari and E. Re, JHEP **1006** (2010) 043 [arXiv:1002.2581 [hep-ph]].

[11] S. Agostinelli *et al.* [GEANT4 Collaboration], Nucl. Instrum. Meth. A **506** (2003) 250.

[12] [CMS Collaboration], CMS-PAS-PFT-10-002.

[13] M. Cacciari and G. P. Salam, Phys. Lett. B **641** (2006) 57 [hep-ph/0512210].

[14] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804** (2008) 063 [arXiv:0802.1189 [hep-ph]].

[15] CMS Collaboration [CMS Collaboration], CMS-PAS-BTV-11-004.

[16] S. Chatrchyan *et al.* [CMS Collaboration], JHEP **1201** (2012) 010 [arXiv:1110.3226 [hep-ex]].

[17] CMS Collaboration [CMS Collaboration], CMS-PAS-TOP-11-003.

[18] L. Lyons, D. Gibaut and P. Clifford, Nucl. Instrum. Meth. A **270** (1988) 110.