Measurement of the $t\bar{t}$ Production Cross Section at $\sqrt{(s)} = 7$ TeV in Lepton + Jets Events using b-quark Jet Identification Techniques

Sadia. Khalil
Department of Physics, 116 Cardwell Hall, Kansas State University, Manhattan, Kansas, USA

An updated measurement of the production cross section for $pp \dashrightarrow t\bar{t}$ at a center-of-mass energy of 7 TeV using data collected by the CMS detector at the Large Hadron Collider is presented. Top quark pair production candidate events are selected based on the presence of an isolated muon or electron of high transverse momentum, large missing transverse energy and hadronic jets. At least one jet is required to be consistent with originating from a b-quark. The analyzed dataset corresponds to an integrated luminosity of 0.8 (1.1) fb^{-1} for the electron (muon) sample. The cross section is extracted with a profile likelihood method using a fit to the number of reconstructed jets, the number of b-tagged jets, and the secondary vertex mass distribution. The measured cross section is 164.4 $\pm$ 2.8 (stat.) $\pm$ 11.9 (syst.) $\pm$ 7.4 (lum.) pb, consistent with higher order QCD calculations.

1 Introduction

The top quark discovered at Tevatron [1,2] has a great importance not only as a standard Model (SM) process but also as an important background to many new physics (NP) searches. In SM, the top quark decays almost 100% by the week process $t \to Wb$. This article describes the $t\bar{t}$ decays where one of the two W bosons decays hadronically and other decays leptonically. The final state we consider contains one lepton(an electron or a muon), a neutrino and at least one jet. Further we require at least one jet to be identified as originating from a b-decay. This presentation is the update of the earlier analysis [3] based on 36 pb^{-1}.

2 Datasets and Event Selection

The data used in this analysis corresponds to an integrated luminosity of 804 (1087) pb^{-1} for electrons (muons), recorded by the CMS experiment [4] between March and July 2011. Various generators are used to simulate the $t\bar{t}$ signal and standard model (SM) backgrounds. Table 1 summarizes the NNLO cross-sections and generators used to normalize signal and various SM backgrounds. The decay of these generated particles is performed using PYTHIA [6]. Events are then simulated using a GEANT4-based model [7] of the CMS detector, and finally reconstructed and analyzed with the same software used to process collision data. For the electron+jets channel, the QCD multi-jets background shape is taken from the simulation using only PYTHIA, however normalization is obtained using the fit to the missing transverse energy distribution in data. The QCD multi-jets background for the muon+jets channel is purely data driven.

The triggers used to collect the data requires at least one charged lepton with transverse momentum (p_T) threshold of 27-42 GeV for electrons (to cope with the high trigger rates) and

Table 1: The NLO cross-sections used to normalize the $t\bar{t}$ and SM backgrounds computed with MCFM.

process	cross section	Generator
$t\bar{t}$	157 pb	MADGRAPH [5]
single t	85 pb	POWHEG
W+jets	31 nb	MADGRAPH
Z+jets	3.1 nb	MADGRAPH

30 GeV for muons. To retain the maximum efficiency, we select the offline isolated electron (muon) with p_T >45 (35) GeV and pseudo-rapidity (η) < 2.5(2.1) and relative isolation w.r.t lepton p_T, I_{rel} < 0.1(0.125) respectively. The detector transition region of barrel and forward calorimeters, 1.4442 < $|\eta_C|$ < 1.5666, where η_C is pseudo-rapidity of electromagnetic cluster, has been also excluded for the electron channel.

The observable particles: muons, electrons, photons, charged and neutral hadrons are reconstructed using CMS particle flow algorithm [8]. The electron candidate is reconstructed by matching the energy clusters deposits in the electromagnetic calorimeter to a track in the pixel and strip tracker. The muon candidate is reconstructed by a fit to hits in the tracker and in the muon system. After these particles are reconstructed, jets are clustered using the anti-k_T jet clustering algorithm [13], with a cone size parameter $\Delta R = 0.5$, as implemented in the Fastjet software package version 2.4 [9]. Small residual jet energy corrections [10] are applied as a function of jet η and p_T. The b-quark jets are identified by a technique using the information about impact parameter significance of the displaced tracks to reconstruct the secondary vertex [11]. The missing transverse momentum (MET) in an event is defined as the negative vector sum of the transverse momenta of all objects from the particle-flow algorithm.

3 Cross Section Measurements

To measure the $t\bar{t}$ cross section a binned Poisson maximum likelihood fit is performed to the number of reconstructed jets ($j = 1, 2, 3, 4, \geq 5$), the number of b-tagged jets ($i = 1, \geq 2$), and the secondary vertex mass distribution in the data. The secondary vertex mass, defined as the mass of the sum of the four-vectors of the tracks associated to the secondary vertex, is found to be a good discriminator between the contributions from heavy flavor (HF) and light flavor (LF) quark production [3].

The signal and backgrounds templates for the fit are normalized to the expected event yields for 0.8 (1.1) fb^{-1} for the electron (muon)+jets events. The W+HF (W_{bx}, W_{cx}), W+LF, and Z+jets backgrounds are all floated independently in the fit. During a simultaneous likelihood maximization, the normalization of each of these components are extracted in-situ along with the $t\bar{t}$ cross section. The QCD multi-jet kinematic distribution of the secondary vertex mass are obtained using events in the sideband data regions (I_{rel} > 0.2, MET < 20 GeV) for muon+jets channel and from simulation with a relatively looser electron identification and isolation of for the electron+jets channel. The initial normalization is estimated from a fit to the MET spectrum. The fit is then constrained to ± 100% of this normalization.

The jet multiplicity distribution (N_j) is sensitive to the jet energy scale (JES), since N_j depends on the choice of the jet p_T threshold. In addition, N_j is affected by variation of Q^2 scale. A larger b-tag efficiency will result in events moving from 1-tag to 2-tag samples, where as, an overall increase in all tag bins together would indicate an increase in the $t\bar{t}$ cross section. JES, W+jets Q^2-scales, and the b-tag efficiency are expected to be the sources of our largest systematic uncertainties. Therefore, we treat them as nuisance parameters in the profile likelihood fit, to take account the correlations between them in order to minimize the total uncertainty.

The Eqs. 1, describes the number of predicted events, for the $t\bar{t}$ signal and two of the W+jets

backgrounds (W+b-jets and W+LF). Similarly, there are other W+jets, the single top, and the QCD backgrounds terms.

$$\begin{aligned} N_{t\bar{t}}^{\rm pred}(i,j) &= \sigma_{t\bar{t}} \cdot N_{t\bar{t}}^{\rm MC}(i,j) \cdot \\ &\quad P^{\text{b tag}}(i,j,R_{\text{b tag}}) \cdot P^{\rm mistag}(i,j,R_{\rm mistag}) \cdot P^{\rm JES}(i,j,R_{\rm JES}) \\ N_{Wb\bar{b}}^{\rm pred}(i,j) &= K_{Wb\bar{b}} \cdot N_{Wb\bar{b}}^{\rm MC}(i,j) \cdot \\ &\quad P^{\text{b tag}}(i,j,R_{\text{b tag}}) \cdot P^{\rm mistag}(i,j,R_{\rm mistag}) \cdot P^{\rm JES}(i,j,R_{\rm JES}) \cdot P^{Q^2}(i,j,R_{Q^2}) \\ N_{Wq\bar{q}}^{\rm pred}(i,j) &= K_{Wq\bar{q}} \cdot N_{Wq\bar{q}}^{\rm MC}(i,j) \cdot \\ &\quad P^{\rm mistag}(i,j,R_{\rm mistag}) \cdot P^{\rm JES}(i,j,R_{\rm JES}) \cdot P^{Q^2}(i,j,R_{Q^2}) \end{aligned} \tag{1}$$

where $\sigma_{t\bar{t}}$ is the fitted cross section of $t\bar{t}$; i and j run over tags and jets, respectively; $K_{Wb\bar{b}}$ is the fitted scale factor for the NNLO prediction for $Wb\bar{b}$ (etc); $N_x^{\rm MC}(i,j)$ is the number of events expected for sample X, derived from MC. The $P^X(i,j,R_X)$ factors are multiplicative functions accounting for the relative differences with respect to the input expected yield, as a function of the assumed value R_X of nuisance parameter X. These are interpolated from various configurations in the simulation with polynomials. The convention chosen is that the nominal event yield is at $R_X = 0$ (i.e., no variation in parameter X), and therefore $P(i,j,R_X) = 1.0$. The "$+1\sigma$" variation is at $R_X = 1$, and the "-1σ" variation is at $R_X = -1$.

The fit minimizes the negative log likelihood, summing over the histogram bins (k) of the secondary vertex mass, the number of jets (j), and the number of tags (i). The various Gaussian constraints (described above) are included, and are represented by C_X. The full profile likelihood expression is

$$-2\ln L = -2\left\{\sum_{i,j}^{\text{tag,jet}} \sum_{k}^{\text{bins}} (\ln \mathcal{P}(N_k^{\rm obs}(i,j), N_k^{\rm exp}(i,j))) - \tfrac{1}{2}\sum_{l}^{\text{constraints}} \frac{(C_X - \hat{C}_X)^2}{\sigma_{C_X}^2}\right\} \tag{2}$$

where $\mathcal{P}$ is a Poisson probability that the predicted yield in each tag/jet bin i,j, given by

$$\ln \mathcal{P}(x,y) = x\ln y - y - \ln\Gamma(x+1) \tag{3}$$

where $\Gamma(x)$ is the Gamma function.

Table 2 shows a summary of all of the inputs to the profile likelihood, as well as the constraints.

Table 2: Inputs to the profile likelihood, along with constraints. All values are in percent.

Quantity	Constraint (%)
b-tag Efficiency Scale Factor	100 ± 10
b-tag Mistag Scale Factor	100 ± 10
Jet energy scale relative to nominal	100 ± 3 (η,p_T dependent)
W+jets renormalization/factorization scales	100^{+100}_{-50}
W+jets background normalization	unconstrained
QCD background normalization	100 ± 100
Single-top background normalization	100 ± 30
Z+jets background normalization	100 ± 30

4 Systematic Uncertainties

Table 3 summarizes the systematic uncertainties. The upper portion of which details the uncertainties that are not included in the profile likelihood and are assessed by the differences in the

acceptance. The lower half portion describes the relative uncertainty due to nuisance parameters. It is determined by fixing all of the other parameters of the likelihood and only allowing the chosen term to vary source that is accounted into fit. The combined number is not the sum of the squares of the contributions, since the fit takes care of all the correlations between them.

Table 3: List of systematic uncertainties for the muon+jet, electron+jet, and combined analyses. Due to the correlation between parameters in the fit, the combined number is not the sum of the squares of the contributions.

Source	Muon Analysis	Electron Analysis	Combined Analysis
Quantity	Uncertainty (%)		
Lepton ID/reco/trigger	3.4	3	3.4
MET resolution due to unclustered energy	< 1	< 1	< 1
$t\bar{t}$+jets Q^2 scale	2	2	2
ISR/FSR	2	2	2
ME to PS matching	2	2	2
Pile-up	2.5	2.6	2.6
PDF	3.4	3.4	3.4
Profile Likelihood Parameter	Uncertainty (%)		
Jet energy scale and resolution	4.2	4.2	3.1
b-tag efficiency	3.3	3.4	2.4
W+jets Q^2 scale	0.9	0.8	0.7
Combined	7.8	7.8	7.3

5 Results

Both channels are fitted simultaneously to the data to extract the cross section. Fig 1 shows the fitted vertex mass distributions.

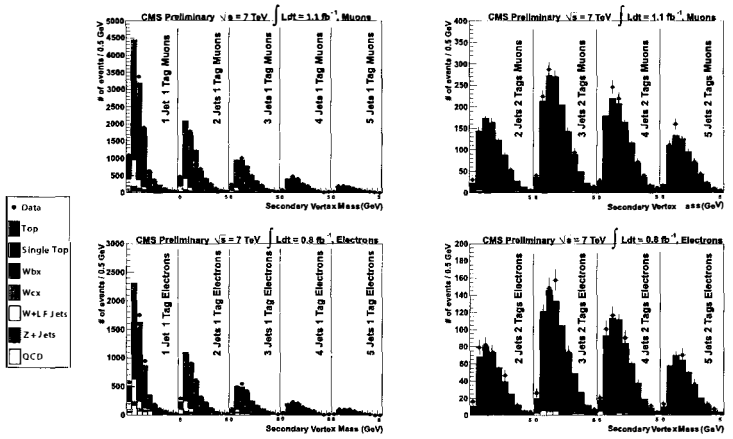

Figure 1: Results of the combined muon and electron channel fit. The top and bottom plots are for the muon and electron channels, respectively. The plots of the left are for single b-tags and those on the right are for $\geq$ 2b-tags. The histograms within each panel correspond to events with 1-, 2-, 3-, 4- and $\geq$5-jets,respectively.

Internal cross check with electron only and muon only fits were also performed. The resulted

cross section measurement is summarized in Eq. 4.

$$\begin{aligned} \sigma_{t\bar{t}}(\mu + jets) &= 163.2 \pm 3.4(\text{stat.}) \pm 12.7(\text{syst.}) \pm 7.3(\text{lum.}) \text{ pb} \\ \sigma_{t\bar{t}}(e + jets) &= 163.0 \pm 4.4(\text{stat.}) \pm 12.7(\text{syst.}) \pm 7.3(\text{lum.}) \text{ pb} \\ \sigma_{t\bar{t}}(l + jets) &= 164.4 \pm 2.8(\text{stat.}) \pm 11.9(\text{syst.}) \pm 7.4(\text{lum.}) \text{ pb} \end{aligned} \quad (4)$$

This result is in good agreement with the QCD predictions of 164^{+6}_{-10}pb [12], [13], 163^{+11}_{-10} pb [14] and 149±11 pb [15] that are based on the full NLO matrix elements and the resummation of the leading and next-to-leading soft logarithms. The fit provides in-situ measurements of the scale factors for both b-tagging and the jet energy scale. We obtain a result of $97 \pm 1\%$ for the b-tagging scale factor which agrees well with the result obtained by the CMS b-tagging group [11]. For the jet energy scale we obtain a result of $99 \pm 2\%$ in agreement with 1. The scale factors for the W+b-jets and W+c-jets components indicate that the contributions to the data may be larger than what is predicted. For the W+b-jets contribution we find cross section scale-factors of 1.2 ± 0.3 and for the W+c-jets contribution of 1.7 ± 0.1. These results are consistent with the scale factors obtained by the individual lepton flavor analyses.

Acknowledgments

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC machine. We thank the technical and administrative staff at CERN and other CMS institutes, and acknowledge support from NSF Grant No. XXX and U.S. DOE under Contract No. XXX.

References

1. Abe, F. et al., CDF collaboration, *Phys. Rev. Lett.* **74**, 2626 (1995).
2. Abachi, S. et al., D0 collaboration, *Phys. Rev. Lett.* **74**, 2632 (1995).
3. Chatrchyan, S. et al., CMS collaboration, *Phys. Rev.* D **84**, 092004 (2011).
4. Adolphi, R. et al., CMS Collaboration, *JINST* **3**, S08004 (2008).
5. Alwall, J. et al., *JHEP* **09**, 028 (2007).
6. Sjostrand. et al., *JHEP* **05**, 026 (2006).
7. Agostineeli, S. et al., *Nucl. Instrum. Methods* A **506**, 250 (2003).
8. CMS collaboration., *CMSPAS*, **PFT-10-002** (2010).
9. Cacciari, F. et al., *Phys. Lett.* B **641**, 57-61 (2006).
10. Chatrchyan, S. et al., CMS Collaboration, *JINST*, **6** P110022011.
11. CMS collaboration, *CMSPAS*, **BTV-11-001** (2011).
12. Aliev, M. et al., *Comput.Phys.Commun*, **182** 1034-1046 (2011).
13. Langenfeld, U. et al., *Phys. Rev.* D **80**, 054009 (2009).
14. Kidonakis, N. et al., *Phys. Rev.* D **82**, 114030 (2010).
15. Ahrens, V. et al., *JHEP* **1009**, 097 (2010).