

Measurement of the $t\bar{t}$ production cross section in the fully hadronic decay channel in pp collisions at 7 TeV

A. Tropiano^{1,a}

Università degli Studi di Firenze,
e-mail: antonio.tropiano@cern.ch

Abstract. The first measurement of the top quark pair production cross section in the fully hadronic channel at a center-of-mass energy of $\sqrt{s} = 7$ TeV is presented. The measurement has been performed using an integrated luminosity of 1.09 fb^{-1} , collected with the CMS detector. The reconstruction of the $t\bar{t}$ candidates is performed after a cut-based event selection using a kinematic fit. A data-driven technique is used to estimate the dominant background from QCD multijet production. The cross section is determined from a fit to the top quark mass. The cross section measurement yields $\sigma(t\bar{t}) = 136 \pm 20(\text{stat.}) \pm 40(\text{sys.}) \pm 8(\text{lumi.}) \text{ pb}$. This result is consistent with an independent measurement in the same channel, with the measurements in other decay channels and with the Standard Model predictions.

1 Introduction

Measurements of the top quark pair production cross section in proton-proton collisions at the LHC provide important tests of our understanding of the top quark production mechanism and can also be used in new physics searches. Recent CMS [1] measurements, making use of the full 2011 dataset corresponding to an integrated luminosity greater than 1 fb^{-1} , have been performed in the dilepton channel, including τ leptons, and in the lepton+jets channel [2–4].

Here the first cross section measurement in the fully hadronic decay channel with the CMS detector is presented [5]. The fully hadronic decay channel has the largest signal yield compared to the dilepton and lepton+jets channels. However, with only jets in the final state, this channel is dominated by the nearly identical QCD multijet background. This measurement demonstrates the good understanding of the CMS detector and is, therefore, an important milestone. This measurement was combined with the more recent measurements in the other channels, providing the most precise CMS measurement obtained to date.

2 Event Selection

To select fully hadronically decaying top quark pairs the candidate events are required to have at least four jets with $p_T > 60 \text{ GeV}/c$, a fifth jet with $p_T > 50 \text{ GeV}/c$, and a sixth jet with $p_T > 40 \text{ GeV}/c$. All jets are required to be within an absolute pseudorapidity $|\eta|$ of 2.4, in the tracker acceptance. Additional jets are considered only if they have a transverse momentum of more than $30 \text{ GeV}/c$.

Tagging jets originating from bottom quarks (b -tagging) is achieved with a robust b -tagging algorithm based on the reconstruction of secondary vertices [6]. This algorithm yields an efficiency of $38 \pm 4\%$ while having a mis-tag rate

of $0.12 \pm 0.02\%$. At least two b -tagged jets are required in each event.

2.1 Kinematical Fit

For the final selection of candidate top quark events a kinematic (least-squares) fit is applied. It exploits the characteristic topology of top quark events: two W bosons with a mass of $80.4 \text{ GeV}/c^2$ that can be reconstructed from the untagged jets and two top quarks that can be reconstructed from the W bosons and the b -tagged jets. The masses of the two top quarks are assumed to be equal, but not fixed, in order to avoid any bias.

To find the right combination of jets, the fit procedure is repeated for every distinguishable jet permutation. This is done using all (six or more) jets that passed the selection criteria. All b -tagged jets are taken as bottom quark candidates, the remaining jets are taken as light quark candidates. If the fit converges for more than one of the possible jet permutations, the one with the minimum χ^2 is chosen. After the kinematic fit, all events with a fit probability $P(\chi^2) > 0.01$ are accepted. In the kinematic fit gaussian resolutions are used for the jets.

The number of events in data passing each selection step and the expected signal fraction using a Standard Model $t\bar{t}$ production cross section of 163 pb [7] are given in Table 1.

3 Signal Extraction

The number of signal events after the final selection is determined via an unbinned maximum likelihood fit to the reconstructed top quark mass distribution, obtained from the kinematic fit. The shapes used in the fit for the signal and background distributions are derived from simulation and a data-driven estimate, respectively. The resulting distribution is shown in Figure 1.

^a on behalf of the CMS Collaboration

Table 1. Number of events and the expected signal fraction in the data sample after each selection step. The expected signal fraction is taken from the simulation, assuming a cross section of 163 pb.

Selection Step	Events	Signal Fraction
At least 6 jets	248109	2%
At least 2 b -tags	6905	17%
Kinematical Fit	1620	32%

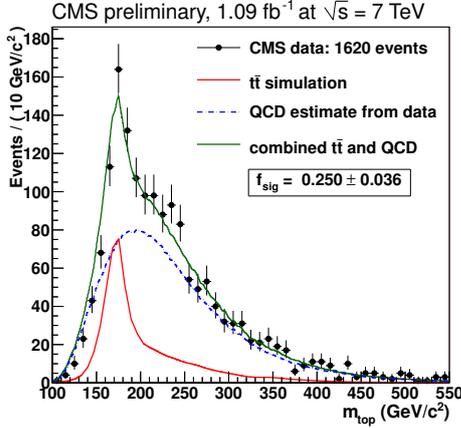


Fig. 1. Result of the fit of the reconstructed top quark mass for the $t\bar{t}$ simulation (solid line) and the multi-jet QCD estimated from data (dashed line). The uncertainty stated on the signal fraction f_{sig} is only statistical.

4 QCD Background Estimation

The background from QCD multijet events is estimated from events with six or more jets of which exactly zero are b -tagged. In this region of the phase space the signal contribution is below 1%. As the kinematics of b -tagged jets and non- b -tagged jets is different, the events that do not contain any b -tag are weighted such that they reproduce the kinematics of the b -tagged jets. This is done by deriving a ratio:

$$R(p_T, |\eta|) = \frac{N(p_T, |\eta|, b)}{N(p_T, |\eta|, not\ b)} \quad (1)$$

where $N(p_T, |\eta|, b)$ is the number of b -tagged jets in events with at least one jet coming from b quarks and $N(p_T, |\eta|, not\ b)$ is the number of untagged jets from all events in data. The ratio R gives the probability of b -tagging a jet as a function of its transverse momentum p_T and absolute pseudorapidity $|\eta|$. The kinematic fit is performed on all events with zero b -tagged jets, assuming every jet as bottom quark candidate. For all permutations with $P(\chi^2) \geq 0.01$ an event weight

$$w = R(p_T^b, |\eta|^b) \times R(p_T^{\bar{b}}, |\eta|^{\bar{b}}) \quad (2)$$

is calculated depending on the p_T and $|\eta|$ of the two jets assigned to the bottom quarks in the kinematic fit hypothesis. The weighted events are used for the estimation of the QCD multijet background that passes the event selection criteria with two b -tagged jets.

5 Results

The inclusive cross section for top quark pair production in the fully hadronic decay channel has been derived from

$$\sigma_{t\bar{t}} = \frac{f_{sig} \cdot N}{\epsilon \cdot L_{int}} \quad (3)$$

The fraction of signal events, f_{sig} , is estimated by the fit of the top quark mass distribution and is found to be 25.0%. The number of events N seen in the data after the full selection is 1620 and the efficiency for selecting $t\bar{t}$ events ϵ , which is determined from simulation, is 0.22%. The latter includes the correction factors for the b -tagging efficiency, mis-tag rate and trigger efficiency found in data. The integrated luminosity of the data sample is $L_{int} = 1.09 \text{ fb}^{-1}$. With these numbers the $t\bar{t}$ production cross section $\sigma_{t\bar{t}} = 136 \pm 20(stat.) \pm 40(sys.) \pm 8(lumi.) \text{ pb}$ is obtained. The total relative uncertainty corresponds to $\pm 33\%$.

To determine the systematic uncertainties on the cross section measurement, samples of simulated events with modified parameters have been used. For each variation the change in the efficiencies and the fitted signal fraction, and hence the cross section, has been studied using pseudo-experiments. An overview of the main sources of systematic uncertainties is given in Table 2.

Table 2. Summary of the main sources of systematic uncertainties with their values.

Systematic Error	Relative Uncertainty
b -tagging	15.7%
Jet Energy Scale	13.5%
Background	12.2%
Q^2 scale	8.7%

6 Cross Check Analysis

A second measurement has been performed as a cross check. For this analysis the kinematic properties of signal and background events are characterized to develop a kinematic selection based on a neural network. Moreover, the QCD multijet background is estimated from data using the information from the kinematic variables of dijet pairs that passed loose b -tagging criteria and extrapolating to the signal region with tighter b -tagging criteria and higher jet multiplicities.

The neural network has been trained using in total eight kinematic variables. These variables are sphericity, centrality, aplanarity, the sum of transverse jet energies (including and without the first two jets), the transverse jet energies weighted with $\sin^2 \theta$ (θ being the angle enclosed by the jet and the beam-axis in the center-of-mass frame of the multijet final state) for the first and second jets, as well as the mean of the remaining jets. The events are all required to have a jet multiplicity of $6 \leq N_{jet} \leq 8$, where the jets selection criteria are the same as for the main analysis. The outcome of the training process is shown in Figure 2. The selected events are required to have a neural network output $NN_{out} > 0.65$.

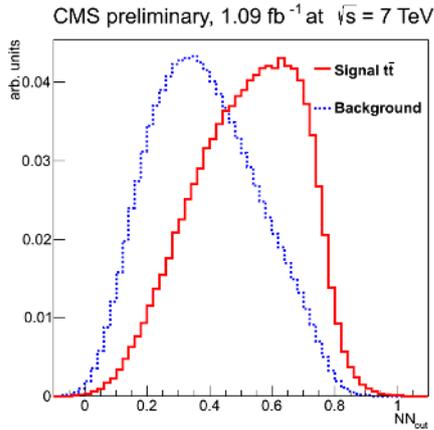


Fig. 2. Neural network output for simulated $t\bar{t}$ events and multi-jet events from data shown normalized to unity.

To improve the purity of the selected sample two b -tags are required using an algorithm different from the one used in the main analysis, with an efficiency of $63 \pm 4\%$ and mis-tag rate of $1.8 \pm 0.2\%$. The assumption for the data-driven QCD multijet background estimate is that the $b\bar{b}$ dynamics in events with six or more jets can be appropriately inferred from a control sample of events with four or five jets. Three variables are used to parametrize the probability that both jets in a given dijet pair are both b -tagged: the average transverse momentum $\langle p_T \rangle$, the average pseudorapidity $\langle |\eta| \rangle$, and the opening angle between the two jets ΔR .

Events with six or more jets have by definition more dijet combinations than events with four or five jets. In order to avoid a bias due to this fact the probability that a given jet pair with both jets having a loose b -tag also has both jets with a medium b -tag was parametrized. With this parametrization, referred to as the ratio R_{LL}^{MM} , the expected amount of background pairs of medium-tagged jets is obtained by weighting each event with

$$w = \sum R_{LL}^{MM}(\langle p_T \rangle, \langle |\eta| \rangle, \Delta R) \quad (4)$$

where the sum runs over all loose-tagged jet pairs in the event with at least two loose b -tags. The $t\bar{t}$ signal is extracted from the reconstructed top quark mass. For the event reconstruction a kinematic fit is used, similar to the one used for the main analysis, with the requirements of $\chi^2 < 40$ and a minimum opening angle between the b -tagged jets of $\Delta R > 1.2$. After the full event selection 937 events remain from which 1125 top quark masses can be reconstructed. Three templates are used to describe the reconstructed top quark mass distribution.

The background template is obtained from data by weighting it with the double-tag probability. The signal template is acquired directly from simulation. A third template representing signal events behaving as background is also derived from simulation, but weighted with the double-tag probability. This is done in order to correct for contamination from signal events in the control region. Finally, the cross section is extracted from a binned maximum likelihood fit of these three templates to the distribution measured from data. Figure 3 shows the comparison between the reconstructed top mass from data and the expected sig-

nal and background distributions normalized to the yields found in the fit. The measured cross section is $\sigma_{t\bar{t}} = 157 \pm 30(stat.) \pm 47(sys.) \pm 8(lumi.)$ pb with a signal fraction of 40%.

The value is in agreement with the Standard Model predictions and with the value obtained by the main analysis. This provides an important and independent cross check of the measurement.

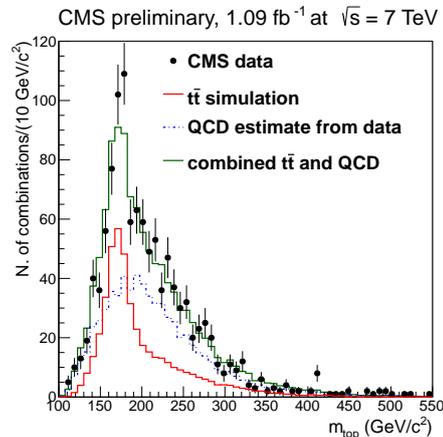


Fig. 3. Reconstructed top quark mass m_{top} for b -tagged jet combinations observed in data (circles) passing the full event selection. For comparison, the expected background (dashed line) and $t\bar{t}$ signal (solid line) are shown, both normalized to the yields from the fit.

7 Conclusions

A first measurement of the top quark pair production cross section in the fully hadronic decay channel at a center-of-mass energy of 7 TeV has been presented. The measurement results in a cross section of

$$\sigma_{t\bar{t}} = 136 \pm 20(stat.) \pm 40(sys.) \pm 8(lumi.) \text{ pb} \quad (5)$$

A cross check analysis, using a neural network based event selection and a different QCD multijet background estimate, gives a result compatible within the errors. Both results are consistent with earlier CMS measurements in the dilepton and lepton+jets decay channels and also with the Standard Model approximate NNLO theoretical calculations [7–9].

References

1. CMS Collaboration, J. Inst. **3**, (2008) 361
2. CMS Collaboration, **CMS-PAS-TOP-11-005**, (2011)
3. CMS Collaboration, **CMS-PAS-TOP-11-006**, (2011)
4. CMS Collaboration, **CMS-PAS-TOP-11-003**, (2011)
5. CMS Collaboration, **CMS-PAS-TOP-11-007**, (2011)
6. CMS Collaboration, **CMS-PAS-BTV-11-002**, (2011)
7. N. Kidonakis, Phys. Rev. D **82**, (2010)
8. Ahrens et al., JHEP **1009**, (2011) 97
9. M. Aliev et al., Comp. Phys. Comm., **182**, (2011)