CMS Physics Analysis Summary

Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS detector

The CMS Collaboration

Abstract

The top-antitop quark ($t\bar{t}$) production cross section is measured for the first time by the CMS experiment in proton-proton collisions at $\sqrt{s} = 13$ TeV at the CERN LHC, using data corresponding to an integrated luminosity of 42 pb$^{-1}$. The measurement is performed by analyzing events with one electron and one muon and at least two jets. The measured cross section is $\sigma_{t\bar{t}} = 772 \pm 60 \text{ (stat)} \pm 62 \text{ (syst)} \pm 93 \text{ (lumi)}$ pb, in agreement with the expectations from the standard model.
1 Introduction

The first measurement of the $t\bar{t}$ cross section at an increased center-of-mass energy (CME) that was not previously accessible has great discovery potential for physics beyond the standard model (SM). It also provides a test of the production mechanism, dominated by gluon-gluon fusion at the LHC, and can be used to check the validity of the theory of quantum chromodynamics (QCD). Furthermore, top quark production is an important source of background in many searches for physics beyond the SM. Previously, large samples of top quark events were collected at the CERN LHC for proton-proton (pp) collisions at CME of 7 TeV and 8 TeV, and were utilized to study top quark pair production in many different decay channels [1–27], as well as to perform searches for deviations from SM predictions [28–50].

This paper presents the first measurement of the $t\bar{t}$ production cross section, $\sigma_{t\bar{t}}$, utilizing data corresponding to an integrated luminosity of 42 $\text{pb}^{-1}$ recorded by the CMS experiment in pp collisions at a CME of 13 TeV. In the SM, top quarks are predominantly produced in $t\bar{t}$ pairs via the strong interaction, and each top quark decays almost exclusively to a W boson and a bottom quark. Final states that contain at least one electron and one muon of opposite electric charge, and at least two jets of opposite electric charge, and at least two jets of particles are selected.

2 Monte Carlo simulation

Several MC event generators are used to simulate signal and background events. The NLO POWHEG (v2) [51, 52] generator is used for $t\bar{t}$ events, assuming a top quark mass of 172.5 GeV. For the default $t\bar{t}$ MC sample, these events are interfaced to PYTHIA (v8.2) [53, 54] to simulate parton showering, hadronization, and the underlying event. An alternative sample is obtained by showering the events with HERWIG++ [55]. Another sample of $t\bar{t}$ events is generated using MG5_AMC@NLO [56] and MADSPIN [57], as well as PYTHIA for parton showering.

The MG5_AMC@NLO generator is also used to simulate W+jets events and Drell–Yan (DY) quark-antiquark annihilation into lepton-antilepton pairs through virtual photon or Z boson exchange. The normalization is taken from data as described below. Single top quark events are simulated using POWHEG (v1) [58, 59] and PYTHIA, and are normalized to the approximate next-to-next-to-leading order (NNLO) cross sections [60]. The contributions from WW, WZ and ZZ (referred to as "VV") are simulated with PYTHIA, and are normalized to the next-to-leading order (NLO) cross sections [61]. All other backgrounds are estimated from control samples extracted from collision data. The simulated samples include additional interactions per bunch crossing (pileup), with distributions that are corrected to match the observed data.

The SM prediction for the $t\bar{t}$ production cross section is $\sigma_{t\bar{t}}^{\text{NNLO+NNLL}} = 832^{+20}_{-29}$ (scale) $\pm 35$ (PDF + $\alpha_s$) $\text{pb}$, as calculated with the TOP++ program [62] at next-to-next-to-leading-order (NNLO) in perturbative QCD, including soft-gluon resummation at next-to-next-to-leading-log order [63], assuming a top quark mass $m_t = 172.5$ GeV. The first uncertainty reflects uncertainties in the factorization and renormalization scales, $\mu_F$ and $\mu_R$, while the second one is associated with possible choices in parton distributions (PDF) and the strong coupling ($\alpha_s$) that follow the PDF4LHC prescriptions [64]. The expected yields for signal in all figures and tables are normalized to this value.
3 Event selection

At trigger level, events are required to contain one electron and one muon, where the electron has transverse momentum $p_T > 12 \text{ GeV}$ and the muon has $p_T > 17 \text{ GeV}$, or the electron has transverse momentum $p_T > 17 \text{ GeV}$ and the muon has $p_T > 8 \text{ GeV}$. Events are selected offline with one electron and one muon of opposite charge, reconstructed using the CMS particle-flow (PF) algorithm [65]. The efficiency for the $e\mu$ triggers is measured in data using triggers based on $p_T$ imbalance in the event. The trigger efficiency is $0.91 \pm 0.05$ when the selection on the leptons described below is applied. The trigger in simulation is modeled through a multiplicative data-to-simulation scale factor (SF) given by the measured dilepton trigger efficiency in data.

Electrons and muons are required to be isolated and to have $p_T > 20 \text{ GeV}$ and $|\eta| < 2.4$. For each electron (muon) candidate, a cone of $\Delta R < 0.3$ ($\Delta R < 0.4$) is constructed around the track direction at the event vertex. Here, $\Delta R$ is defined as $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$, and $\Delta \eta$ and $\Delta \phi$ are the differences in pseudorapidity and azimuth between any energy reconstructed with the PF algorithm and the lepton track. The scalar sum of the $p_T$ of all such particle contributions consistent with arising from the chosen event primary vertex is calculated, excluding the contribution from the lepton candidate. The relative isolation discriminant, $I_{\text{rel}}$, is defined as the ratio of this sum to the $p_T$ of the lepton candidate. The neutral component is corrected for pileup based on the average energy density deposited by neutral particles in the event, which corresponds to an average transverse energy from pileup determined event by event that is subtracted from the transverse energy in the isolation cone. An electron (muon) candidate is selected if $I_{\text{rel}} < 0.11$ ($I_{\text{rel}} < 0.12$).

In events with more than one pair of leptons passing these selections, the two opposite-sign different-flavor leptons with largest transverse momenta are selected for further study. Events with $\tau$ leptons contribute to the measurement only if they decay to electrons or muons that satisfy the selection requirements.

The efficiency of the lepton selection is measured using a tag-and-probe method in same-flavor dilepton events enriched in Z boson candidates, as described in Ref. [22]. The measured values for the combined identification and isolation efficiencies are typically 0.92 for muons and 0.77 for electrons. Based on a comparison of lepton selection efficiencies in data and simulation, the event yield in simulation is corrected using $p_T$- and $\eta$-dependent SF to provide consistency with data. They have an average value of 1.00 for muons and 0.96 for electrons.

Candidate events with a dilepton pair of invariant mass $M_{e^+\mu^-} < 20 \text{ GeV}$ are removed to suppress backgrounds from heavy-flavor resonances, as well as contributions from DY processes with low invariant dilepton mass.

Jets are reconstructed from the PF particle candidates using the anti-$k_T$ clustering algorithm [66] with a distance parameter of 0.4. The jet energy is corrected for pileup in a manner similar to the correction of the energy inside the lepton isolation cone. Jet energy corrections are also applied as a function of the jet $p_T$ and $\eta$ [67] to both data and simulations. Events are required to have at least two reconstructed jets with $p_T > 30 \text{ GeV}$ and $|\eta| < 2.4$.

4 Background determination

Backgrounds in this analysis arise primarily from single top quark, DY, and VV events in which at least two prompt leptons are produced by Z or W decays. Background yields from single top quark and VV events are estimated from simulation. The DY background normalization is estimated from data using the “$R_{\text{out/in}}$” method [22, 26, 27]. Events with $e^+e^-$ and $\mu^+\mu^-$ final
states are explored to normalize $e^\pm \mu^\mp$ DY production. A data-to-simulation normalization factor is estimated from the number of events in data within the Z boson mass window and extrapolated to the number of events outside the Z mass window with corrections applied using control regions enriched in DY events in data. This factor is found to be $1.06 \pm 0.17$.

Other background sources, such as $t\bar{t}$ or $W+$jets events with decays into lepton+jets, can contaminate the signal sample if a jet is incorrectly reconstructed as a lepton (which mainly happens for electrons), or contains a lepton from the decay of bottom or charm hadrons (which mainly happens for muons). These are grouped into the non-prompt leptons category ("non-$W/Z$ leptons", since prompt leptons originate from decays of $W$ or $Z$ bosons), together with contributions that can arise, for example, from the decays of mesons, photon conversions, or finite-resolution detector effects.

The non-$W/Z$-lepton background is estimated from a control region of same-sign (SS) events to the final region of opposite-sign (OS) signal. The same-sign control region is defined using the same criteria as the nominal signal region, except for requiring $e\mu$ pairs of same electric charge. The same-sign dilepton events are predominantly events containing false leptons, as processes that produce non-prompt leptons produce OS and SS events at similar rates. Other SM processes have significantly smaller rates to produce prompt SS or charge-misidentified dileptons, and are estimated using simulation and subtracted from the observed number of events in data.

The scaling from the SS control region in data to the signal region is performed through the ratio of the number of OS events with false leptons to the number of SS events with false leptons in the MC simulation. This ratio, calculated using $t\bar{t}$ and single top quark samples, is measured to be $1.37 \pm 0.06$ (stat). With eight same-sign events observed in data and an expected contamination of $1.8 \pm 0.4$ (stat), $8.5 \pm 4.3$ events with fake leptons contaminating the signal region are predicted including statistical and systematic uncertainties. This agrees with predictions from MC simulations of semileptonic $t\bar{t}$ and $W+$jets events within uncertainties.

Figure 1 (a) shows the multiplicity of jets, and (b) the the scalar sum of the transverse momenta of all jets ($H_T$) for events passing the dilepton criteria. After requiring at least two jets, Figs. 2 (a,b,c,d) show the $p_T$ and $\eta$ distributions of the highest-$p_T$ muons and electrons, respectively, and Figs. 3 (a,b,c,d) show the $p_T$ and $\eta$ distributions of the highest- and second highest-$p_T$ jets, respectively. The ratios of the data to the sum of simulations and data-based predictions for the signal and backgrounds are shown in the bottom panels of all figures. Agreement between data and the predictions for signal and background is observed.

5 Sources of systematic uncertainty

Table 1 summarizes the magnitude of the statistical and systematic uncertainties on the $t\bar{t}$ production cross section from different sources. All sources of uncertainties are added in quadrature to obtain the total uncertainty.

The uncertainties on the SF applied to simulation to correct for differences to data for trigger efficiencies are 5%. The uncertainties on the SF applied to correct electron (muon) identification efficiencies are found to be 4% (1–3%).

The modeling of lepton energy scales was studied using $Z \rightarrow ee/\mu\mu$ events in data and simulation, resulting to an uncertainty of the electron energy scale of 1% and the muon energy scale of 0.5%.
5 Sources of systematic uncertainty

Figure 1: The distributions for (a) the jet multiplicity and (b) $H_T$ in events passing the dilepton criteria. The expected distributions for $t\bar{t}$ signal and individual backgrounds are shown after data-based corrections are applied; the last bin contains the overflow events. The ratios of data to the sum of the expected yields are given at the bottom of each panel.

The impact of uncertainty in jet energy scale (JES) and jet energy resolution (JER) are estimated from the change observed in the number of selected MC $t\bar{t}$ events after changing the jet momenta within the JES uncertainties of 4%, and for JER by an $\eta$-dependent variation of the JER scale factors up and down by one sigma.

The uncertainty assigned to the number of pileup events in simulation is obtained by changing the inelastic cross section by $\pm 5\%$.

The systematic bias related to the missing higher-order diagrams in POWHEG are estimated as follows: The uncertainty in the signal acceptance is determined by changing the renormalization and factorization scales in POWHEG simultaneously up and down by a factor of two, with the uncertainty taken as the maximum observed difference.

In addition, the predictions of the NLO generators POWHEG and MC5_AMC@NLO for $t\bar{t}$ production are compared, where both use PYTHIA for hadronization and extra radiation.

The uncertainty arising from the hadronization model mainly affects the JES and the fragmentation of $b$ quark jets. The uncertainty in the JES already contains a contribution from the uncertainty in the hadronization. The hadronization uncertainty is also determined by comparing samples of events generated with POWHEG, where the hadronization is modeled with PYTHIA or HERWIG++ [55].

The uncertainty from the choice of PDF is determined by reweighting the sample of simulated $t\bar{t}$ events according to the 100 NNPDF3.0 error PDF sets [68].

An uncertainty in the cross sections for single top quark and VV backgrounds, taken from measurements and approximately 30% [69–77] is added in quadrature. For DY production, an uncertainty of 15% is assumed.

The systematic uncertainty on the estimate of the non-W/Z lepton background is given mainly by the systematic uncertainty on the ratio of the number of OS events with false leptons to the number of SS events with false leptons in the MC simulation. It depends on how closely the simulation models the production of false leptons. This is checked by examining additional control regions, and the observed discrepancy is used to assign an uncertainty of 30% to the
Figure 2: The distributions for (a) $p_T$ and (b) $\eta$ for the highest-$p_T$ muon, and for (c) $p_T$ and (d) $\eta$ for the leading electron after all selections. The expected distributions for $t\bar{t}$ signal and individual backgrounds are shown after data-based corrections are applied; for the left plots (a,c) the last bin contains the overflow events. The ratios of data to the sum of the expected yields are given at the bottom of each panel.
Figure 3: The distributions for (a) $p_T$ and (b) $\eta$ for the highest-$p_T$ jet, and for (c) $p_T$ and (d) $\eta$ for the second highest-$p_T$ jet after all selections. The expected distributions for $t\bar{t}$ signal and individual backgrounds are shown after data-based corrections are applied; for the left plots (a,c) the last bin contains the overflow events. The ratios of data to the sum of the expected yields are given at the bottom of each panel.
A preliminary luminosity scale and its uncertainty is estimated from \(x-y\) beam-beam scans performed in July 2015 utilizing the methods from Ref. [78]. The uncertainty in the integrated luminosity is 12%.

Table 1: Summary of the individual contributions to the systematic uncertainty on the \(\sigma_{tt}\) measurement. The uncertainties are given in pb and as relative uncertainties. The statistical uncertainty and the total uncertainty on the result are also given.

<table>
<thead>
<tr>
<th>Source</th>
<th>(\Delta\sigma_{tt}) (pb)</th>
<th>(\Delta\sigma_{tt}/\sigma_{tt}) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data statistics</td>
<td>60</td>
<td>7.7</td>
</tr>
<tr>
<td>Trigger efficiencies</td>
<td>39</td>
<td>5.0</td>
</tr>
<tr>
<td>Lepton efficiencies</td>
<td>33</td>
<td>4.3</td>
</tr>
<tr>
<td>Lepton energy scale</td>
<td>(&lt; 1)</td>
<td>(\leq 0.1)</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>20</td>
<td>2.6</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>(&lt; 1)</td>
<td>(\leq 0.1)</td>
</tr>
<tr>
<td>Pileup</td>
<td>2.8</td>
<td>0.4</td>
</tr>
<tr>
<td>Scale ((\mu_F \text{ and } \mu_R))</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>(t\bar{t}) NLO generator</td>
<td>15</td>
<td>1.9</td>
</tr>
<tr>
<td>(t\bar{t}) hadronization</td>
<td>14</td>
<td>1.8</td>
</tr>
<tr>
<td>PDF</td>
<td>12</td>
<td>1.5</td>
</tr>
<tr>
<td>Single top quark</td>
<td>14</td>
<td>1.8</td>
</tr>
<tr>
<td>VV</td>
<td>3.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Drell–Yan</td>
<td>3.9</td>
<td>0.5</td>
</tr>
<tr>
<td>Non-W/Z leptons</td>
<td>8</td>
<td>1.0</td>
</tr>
<tr>
<td>Total systematic (no integrated luminosity)</td>
<td>62</td>
<td>8.0</td>
</tr>
<tr>
<td>Integrated luminosity</td>
<td>93</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>126</td>
<td>16.4</td>
</tr>
</tbody>
</table>

6 Results

Table 2 shows the total number of events observed in data together with the total number of signal and background events expected from simulation or estimated from data. The mean acceptance multiplied by the selection efficiency and the branching fraction, as estimated from simulation for a top quark mass of 172.5 GeV, is \(\epsilon = (0.60 \pm 0.04)\%\), including statistical and systematic uncertainties. The extracted cross section using the event yields is \(\sigma_{tt} = 772 \pm 60 \text{ (stat)} \pm 62 \text{ (syst)} \pm 93 \text{ (lumi)}\) pb for a top quark mass of 172.5 GeV. The measured fiducial cross-section for \(t\bar{t}\) production with two leptons (one electron and one muon) in the range \(p_T > 20\) GeV and \(|\eta| < 2.4\), is \(\sigma_{tt} = 12.9 \pm 1.0 \text{ (stat)} \pm 1.1 \text{ (syst)} \pm 1.5 \text{ (lumi)}\) pb.

The dependence of the acceptance on the top quark mass has been studied in the range 169.5–175.5 GeV and is parameterized as a linear dependence. Assuming a top quark mass value of 173.34 GeV [79], the cross section decreases by \(\approx 0.7\%\).

Figure 4 (a) shows the distribution of the invariant dilepton mass \(M_{e^+\mu^-}\) which can probe, for example, the existence of a new heavy object decaying into a top quark pair. Figure 4 (b) shows the difference in azimuthal angle between the two selected leptons \(\Delta\Phi(e^\pm, \mu^\mp)\) and explores the correlation between the top and antitop quark spins [30, 80–84]. For both distributions data are in agreement with the SM predictions.

Figure 5 shows the result for \(\sigma_{tt}\) together with previous CMS results at \(\sqrt{s} = 7\) and at 8
Table 2: Number of dilepton events obtained after applying the full selection. The results are given for the individual sources of background, $t\bar{t}$ signal with a top quark mass of 172.5 GeV and $\sigma_{t\bar{t}}^{\text{NNLO}+\text{NNLL}} = 832^{+40}_{-46}$ pb, and data. The uncertainties correspond to the statistical and systematic components added in quadrature.

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of events $e^\pm\mu^\mp$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drell–Yan</td>
<td>$6.4 \pm 1.2$</td>
</tr>
<tr>
<td>Non-W/Z leptons</td>
<td>$8.5 \pm 4.3$</td>
</tr>
<tr>
<td>Single top quark</td>
<td>$10.6 \pm 3.4$</td>
</tr>
<tr>
<td>VV</td>
<td>$2.6 \pm 0.9$</td>
</tr>
<tr>
<td>Total background</td>
<td>$28.1 \pm 5.7$</td>
</tr>
<tr>
<td>$t\bar{t}$ dilepton signal</td>
<td>$207 \pm 16$</td>
</tr>
<tr>
<td>Data</td>
<td>$220$</td>
</tr>
</tbody>
</table>

Figure 4: Distributions in (a) the dilepton invariant mass, and (b) the difference in the azimuthal angle between the two selected leptons, after all selections. For the first plot the last bin contains the overflow events. The ratios of data to the sum of the expected yields are given at the bottom.

7 Summary

A measurement of the $t\bar{t}$ production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV is presented for events containing an electron-muon pair and at least two jets. The measurement is obtained through an event-counting analysis based on a data sample corresponding to an integrated luminosity of 42 pb$^{-1}$. The result obtained is $\sigma_{t\bar{t}} = 772 \pm 60$ (stat) $\pm 62$ (syst) $\pm 93$ (lumi) pb, resulting in a total relative uncertainty of 13.5%. This measurement is consistent with the SM prediction of $\sigma_{t\bar{t}}^{\text{NNLO}+\text{NNLL}} = 832^{+40}_{-46}$ pb for a top quark mass of 172.5 GeV and with a recent measurement from the ATLAS Collaboration [88].

References

Figure 5: Top quark pair production cross section in p¯p and pp collisions as a function of CME. The Tevatron combination at $\sqrt{s} = 1.96$ TeV are displayed, as well as CMS results at 7 and 8 TeV in the dilepton and semileptonic (“l+jets”) channels. For the $e\mu$ channel, the LHC combination at 8 TeV and the CMS result at 13 TeV is shown. The measurements are compared to the NNLO+NNLL theory predictions.


[40] CMS Collaboration, “Search for Resonant $t\bar{t}$ Production in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV”, arXiv:1506.03062.


