CMS Physics Analysis Summary

Contact: cms-pag-conveners-top@cern.ch 2009/08/07

Measurements of the $t\bar{t}$ cross section in the dilepton channels with $\mathcal{L} = 100\ pb^{-1}$ using the CMS detector

The CMS Collaboration

Abstract

We present an analysis for the measurement of the top quark pair ($t\bar{t}$) production cross-section at $\sqrt{s} = 14\ TeV$ using dilepton final states. The event selection, based on simple cuts is described and an almost background free final sample can be derived. For a total integrated luminosity of 100 pb^{-1}, the cross-section is extracted by means of a robust event counting method and the associated statistical uncertainty is found to be 8%.

1 Introduction

The potential for measurements of the $t\bar{t}$ production cross section with the first 100 pb^{-1} of integrated luminosity is presented in this analysis summary. The $t\bar{t}$ production ($\sigma_{NLO}(t\bar{t}; 14\text{ TeV}) =$ 833 pb [1]), dominated by the gluon-gluon fusion mechanism is considered in the dileptonic final states (ee, $\mu\mu$, and $e\mu$ including leptons from τ decays). Realistic detector conditions at the LHC start-up are taken into account.

The analysis presented in this note uses b-tagging to isolate the $t\bar{t}$ dilepton signature. A complementary analysis [2] is being pursued at CMS without b-tagging in the event selection. This complementary analysis is tuned for initial data-taking, before the detector is calibrated well enough to allow the use of b-tagging. Both analyses are based on selecting two isolated leptons (ee, $\mu\mu$, and $e\mu$), missing E_T (MET), and jets. However, as CMS is exploring various options, the details of the lepton and MET selections are different between the two analyses and cannot be directly compared. These selections will be optimized with real data in light of detector performance and measured backgrounds.

The analyses presented here have been carried out using the CMS event data model and the official software framework of CMS for event generation, simulation and reconstruction. Signal and background events come from the CMS Monte Carlo production performed with the full simulation of the detector in which experimental conditions at the beginning of data taking have been taken into account by various miscalibration and misalignment scenarios. For an integrated luminosity of 100 pb^{-1}, the detector is assumed to be calibrated and aligned with an amount of data corresponding to 10 pb^{-1}. In this scenario, the detector is aligned and calibrated using cosmic muons, low mass resonances and minimum bias events.

The generation of signal events ($t\bar{t}$+jets) and major background events (Z+jets, W+jets , for different transverse momentum (p_T) of the boson values ranging from 0 to 300 GeV/c) was performed using the ALPGEN generator [3]. The ALPGEN samples are produced exclusively for numbers of partons up to 3(4) and inclusively for the 4th(5th) partons for the signal (backgrounds respectively). They contain a total number of about 40M generated events. Di-bosons processes (WW, ZZ, WZ representing 1.4M events) and multi-jet QCD background (8.5M events) productions are generated with PYTHIA [4].

Due to the fact that muons and electrons have different reconstruction and selection efficiencies and because backgrounds are final state dependent, exclusive cross section measurements have been studied.

2 Trigger selection

The number of collected $t\bar{t}$ events will depend on the trigger efficiency to select these events. In the CMS design, the real time selection of events is achieved in two physical steps, namely the fast Level-1 Trigger and the High-Level Trigger (HLT) operating on longer timescales. The Level-1 trigger is built of mostly hardware level information of the detectors while the HLT selection is implemented as a sequence of reconstruction and filter steps of increasing complexity. The first step was to "skim" the data (make an offline preselection with loose cuts). Several High Level inclusive triggers have been considered according to the skim path used: For the dielectron studies, the skim path is based on the single electron trigger requiring a non isolated electromagnetic energy deposit at Level-1 followed by a loose HLT isolation (based on HCAL and tracking information [5])and an explicit cut on the calorimeter energy over the tracker momentum of the reconstructed electron $|E/P|$ ($|E/P| < 1.5$ in barrel and $|E/P| < 2.45$

in the forward region) and HLT selection (transverse energy (E_T) threshold of 17 GeV) and the double electron trigger (requiring two loose isolated HLT electrons with $E_T > 12$ GeV for each lepton). For the dimuon and the electron-muon studies, the skim path is based on the single muon trigger (requiring a seeded Level-1 non isolated muon passing the HLT p_T threshold of 16 GeV/c), the double muon trigger (two non isolated HLT muons with $p_T > 3$ GeV/c for each muon) and the double mixed electron-muon trigger (a loose isolated electron with the same $|E/P|$ requirements as in the single electron trigger and a non isolated HLT muon with E_T and p_T thresholds of 10 GeV and 10 GeV/c respectively for both leptons). Additional offline selection cuts on both leptons ($p_T > 20$ GeV/c) for the skims are imposed. More details on the CMS HLT trigger can be found on Ref. [5]

3 Reconstruction tools

The e/μ final states for the signal is characterized by the presence of two highest-p_T isolated leptons (coming from W boson decays) associated with a large missing transverse energy MET and 2 b-jets. The reconstruction of such objects will be briefly described.

Electron candidates are reconstructed starting from electromagnetic calorimeter towers. The corresponding region in the pixel detector is used to define track seeds. Electron trajectories are then reconstructed from the seed to the calorimeters towers using a Gaussian-Sum filter technique [6]. During the data taking, the electron triggers and the reconstruction efficiencies will be estimated with $Z \rightarrow ee$ events using the so-called "tag and probe" method.

Muon candidates are reconstructed by extending the muon trajectories from the Muon chambers to the tracker. Muons tracks are then reconstructed in the region of interest using a Kalman filter technique [6]. As for electrons, muon triggers and muon reconstruction efficiencies will be estimated with $Z \rightarrow \mu\mu$ events using the "tag and probe" method.

Lepton isolation is applied on candidates at the level of tracker and calorimeter reconstruction. The tracker (calorimeter) isolation is defined as the scalar sum of the transverse momenta of all tracks found within a cone of size $R = \sqrt{\eta^2 + \phi^2} = 0.3$ (the sum of the energies of all calorimeter towers found within a cone of size $R = 0.3$) around the lepton direction excluding the lepton track (the lepton energy deposited in the calorimeter respectively). Cuts on tracker and calorimeter isolation are treated as a fixed pair of cuts. For the selection of electron and muon candidates the cut values are set to (3 GeV/c^2; 6 GeV) and (3 GeV/c^2; 1 GeV) respectively.

The tracker isolation variable is illustrated by the Fig. 1 for reconstructed leptons coming from W decays and other reconstructed leptons, in $t\bar{t}$ events.

Jets are reconstructed using an iterative cone jet algorithm with a cone radius of $R = 0.5$, using calorimeter towers as input. A response correction derived from Monte Carlo calibration is applied to the reconstructed jet transverse energies. Similarly, the missing transverse energy is reconstructed using calorimeter towers and is corrected for jet energy rescaling as well as for the presence of muons in the event. Finally, the b-jets can be identified using a track counting algorithm. First, the tracks contained in a cone ($\Delta R = 0.3$) around the jet axis are associated to this jet. Then, a jet is defined as a b-jet if it contains at least two tracks with a 3D impact parameter significance larger than a given cut. Although this b-tagging method appears robust and efficient in its conception, it can be affected by misalignments or inefficiency effects not well described by the simulation. For these reasons, we used the loose working point of the b-jet discriminator, see [7], [8] and [9].

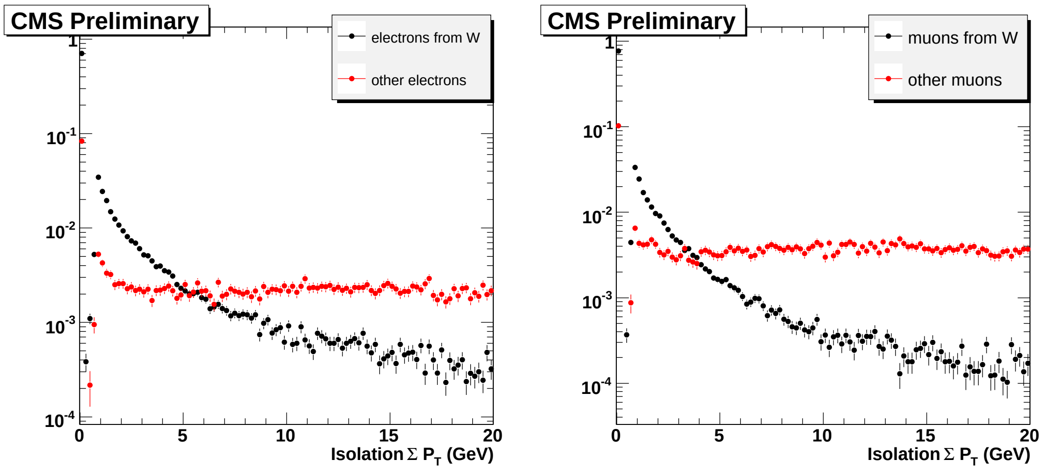

Figure 1: Distribution of variables used in muon (left) and electron (right) tracker isolation, as described in the text. The black distributions correspond to reconstructed leptons matched with leptons originating from the decay of a W boson using an angular matching criterion ($R < 0.15$), while the red distributions correspond to any other lepton. $t\bar{t}$ events were used. The distributions are normalized to 1. The peak for "other leptons" for isolation near to 0 is due to leptons coming from b-jet which appears to be isolated. The first bin peaked at 0 is associated to leptons without any track in the cone. The dip close to 0 is due selection criteria (mainly p_T cut) on the tracks.

4 Event selection

The objects previously defined will now be used for the events selection. The event selection features also a significant acceptance for $t\bar{t} \rightarrow (\text{di})\tau + \text{X}$ final states with leptonic tau decays. The aim of the analysis is to find a good balance between background rejection and signal selection efficiency. The backgrounds mainly originate from three types of processes, according to the nature of the lepton (real or fake) and its isolation status: Physical processes with two real isolated leptons are mainly Z+jets events and to a lower extent dibosons events (with leptonic decays of vector bosons); physical processes with one real lepton associated with a fake lepton (W+jets and $t\bar{t}$ with semi-leptonic decays) can also be selected. The third possibility appears when two jets are reconstructed as lepton candidates: this is the case for multi-jet QCD events (including $b\bar{b}$ and $c\bar{c}$ events) as well as $t\bar{t}$ with fully hadronic decays.

4.1 Pre-selection

Firstly a pre-selection is applied in order to reconstruct $t\bar{t}$ event candidates. Additional cuts are then applied to remove the main backgrounds. During the event reconstruction, only jets with a transverse energy $E_T > 15$ GeV in the detector acceptance $|\eta| < 2.4$ are taken into account. In order to reject leptons reconstructed as jets, the jets which are closer than $R = 0.3$ to an electron candidate are rejected, if the electron has a tracker isolation greater than 2 GeV/c. Then, only leptons with a transverse momentum $p_T > 10$ GeV/c in the detector acceptance $|\eta| < 2.4$ will be used for the event reconstruction. All the selected leptons are also required to pass the isolation criteria. Candidate events are selected when the two highest p_T reconstructed leptons have opposite charge. Since no b-tagging is applied at this stage of the analysis, the two b jets candidates are defined as the ones with the highest E_T.

4.2 Main selection

In order to remove backgrounds related to fake leptons (QCD multi-jets, W+jets, $t\bar{t}$ semi-leptonic), the selected leptons are required to have a $p_T > 20$ GeV/c. Additional quality cuts are applied on electrons in order to keep a good control on electron fake rates: the ratio between the energies deposited in the hadronic and the electromagnetic calorimeters must be below 0.05, while the ratio between the energy in the electromagnetic calorimeter and the track momentum associated to the electron has to be between 0.8 and 3. At this level, about 96(95)% of the reconstructed leptons pair are correctly assigned to the corresponding generated leptons. In addition, the two b-jet candidates are required to have a corrected $E_T > 30$ GeV.

In order to reject the major $Z + jets$ background contribution, events reconstructed with a dileptonic invariant mass close to the Z mass (within the range 75 to 105 GeV/c^2) are rejected. This cut is not applied for the $e\mu$ channel. The presence of two neutrinos in signal events allows suppression of Drell-Yan events as well as multi-jets QCD events by applying a selection on the missing transverse energy ($MET > 50$ GeV).

Finally, the b-tagging is applied on the two b-jets candidates. In order to keep the b-tagger as robust as possible, the loose working point was chosen. The mistag rate will be evaluated using the method described in the note [9]. Since we require 2 b-tagged jets, the mistag rate is expected to be small.

Based on simulated events the residual background is found to be very small. Despite a very large rejection rate of instrumental background such as QCD multi-jet events, the corresponding huge multi-jet cross-section represents a potential spoiling effect that can be estimated using a data-driven background expectation approach. From the Monte Carlo point of view, a very large data sample is needed to properly estimate the QCD multi-jets contamination. For this reason a factorizing approach was used to estimate an order of magnitude of the QCD background rejection. Using this method, QCD contamination was found to be negligible.

5 Results

Because the dileptonic signal is clean and almost background-free (after b-tagging), the dileptonic $t\bar{t}$ cross section can be extracted with a simple counting method as:

$$\sigma \times BR = \frac{N_{sel} - N_{bkg}}{\varepsilon_{t\bar{t}} \times \int \mathcal{L}}$$

where N_{sel} is the number of events passing the selection cuts, N_{bkg} the residual background and $\varepsilon_{t\bar{t}}$ expanded as $\varepsilon_{t\bar{t}} = \varepsilon_{t\bar{t}}^{HLT} \times \varepsilon_{t\bar{t}}^{MC} \times \varepsilon_{t\bar{t}}^{reco/sel}$ ($\varepsilon_{t\bar{t}}^{HLT}$ being the trigger efficiency estimated from data, $\varepsilon_{t\bar{t}}^{MC}$ the selection efficiency estimated from the Monte Carlo and $\varepsilon_{t\bar{t}}^{reco/sel}$ a correction factor which accounts for reconstruction differences between data and Monte Carlo (since we are using only Monte Carlo, this correction factor is set to 1 in our analysis).
The total number of dileptonic events produced for a total integrated luminosity of 100 pb^{-1} is

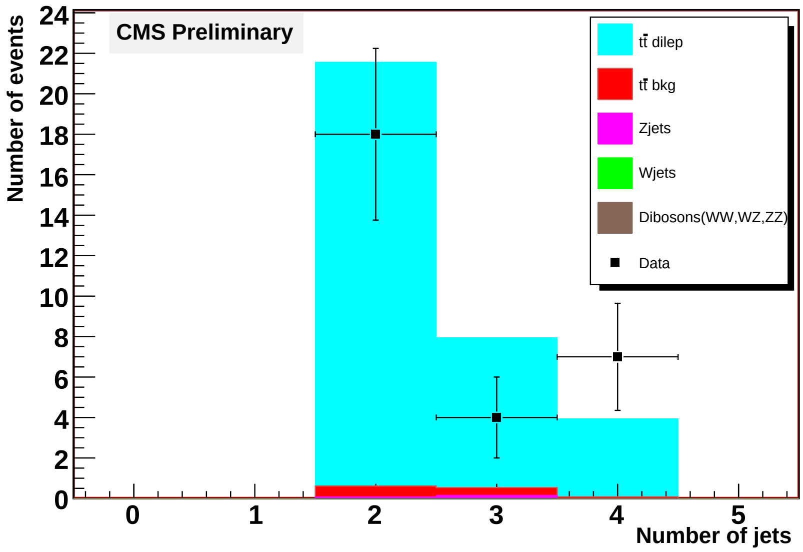

Figure 2: Distribution of the jet multiplicity for the final selected events in *ee* channel compared with pseudo-data for a total integrated luminosity of 100 pb^{-1}. The last bin in the distribution is inclusive (≤ 4).

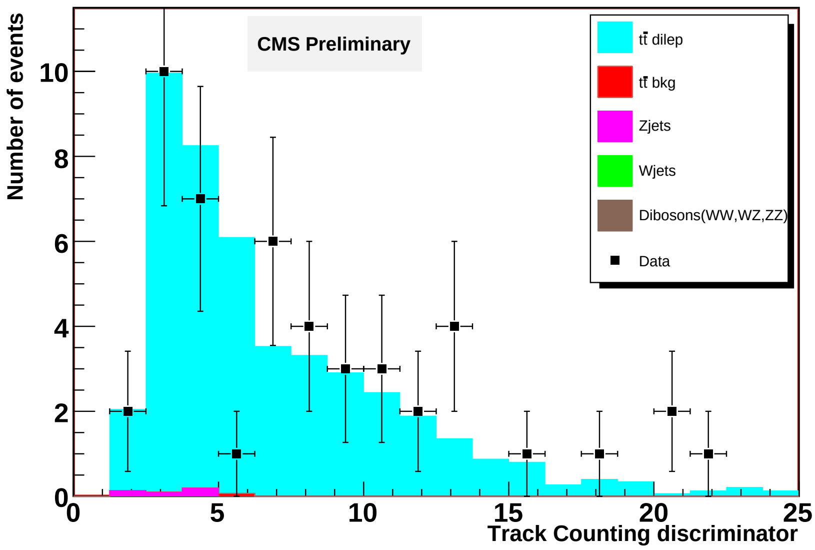

Figure 3: Distribution of the b-tagging discriminator for the final selected events in $\mu\mu$ channel compared with pseudo-data for a total integrated luminosity of 100 pb^{-1}.

5340 at $\sqrt{s} = 14$ TeV. With this luminosity, the expected total number of signal events is about 160 events with a very small residual background (Z+jets, W+jets and dibosons) of about 3 events. Fig. 2 shows the jet multiplicity for the *ee* channel. The main background in this channel come from other $t\bar{t}$ and Z+jets events (around 1 event in total). Fig. 3 shows the distribution of the b-tagging discriminator in the $\mu\mu$ channel, the main contribution coming from Z+jets events (with a total number of events smaller than 1) and finally Fig. 4 shows the MET in the

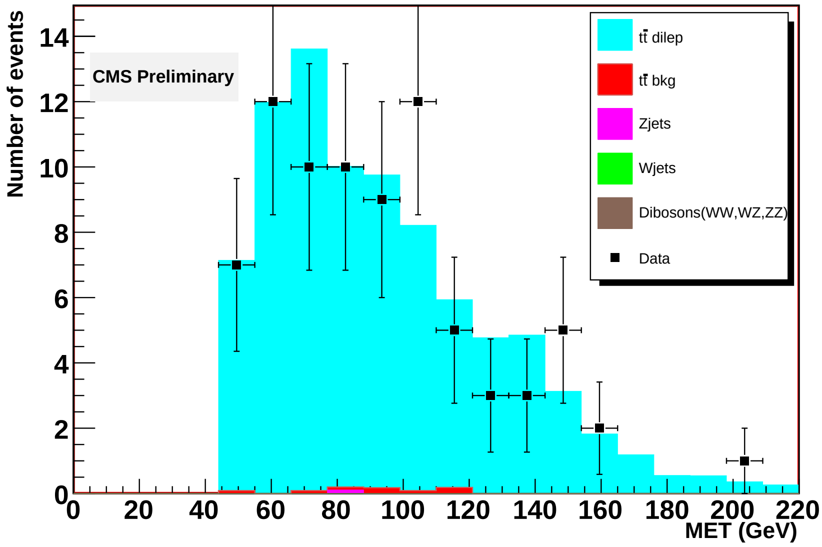

Figure 4: Distribution of the missing transverse energy for the final selected events in $e\mu$ channel compared with pseudo-data for a total integrated luminosity of 100 pb^{-1}.

$e\mu$ channel. The main background contribution comes from other $t\bar{t}$ events and Z+jets events (around 1 event in total). These numbers can change significantly by removing the b-tagging selection (about 600 signal events, 9 $t\bar{t}$ semi-leptonic, 46 Z+jets, 4 W+jets, 8 dibosons events and 15 multi-jet events). For ee, $\mu\mu$ and $e\mu$ final states, the corresponding total efficiencies $\varepsilon_{t\bar{t}}$ are 2.3%, 3.5% and 3.2% leading to a statistical error of 15%, 18% and 11% respectively. For $t\bar{t}$ signal events and for jets with $|\eta| < 2.4$ and $E_T > 30$ GeV, the b-tagging efficiency was found to be around 65% while the mistag rate was found to be around 13%. The fractions of selected signal events containing at least one tau which decays leptonically are also estimated. For the ee channel, the fraction of selected events for $t\bar{t} \to e(\tau \to e)$ is 13% and for $t\bar{t} \to (\tau \to e)(\tau \to e)$ is about 1%. For the $\mu\mu$ channel, the fraction of selected events for $t\bar{t} \to \mu(\tau \to \mu)$ is 11% and for $t\bar{t} \to (\tau \to \mu)(\tau \to \mu)$ is almost negligible. Finally, for the $e\mu$ channel, the fraction of selected events for $t\bar{t} \to e(\tau \to \mu)$, $\mu(\tau \to e)$ is 13% and for $t\bar{t} \to (\tau \to e)(\tau \to \mu)$ is almost negligible.

The statistical uncertainty summed over the three channels is found to be around 8%. The signal-to-background ratio is 4 for ee channel, 6 for $\mu\mu$ channel and more than 20 for $e\mu$ channel before b-tagging. The signal-to-background ratio varies from 26 for ee channel and up to 90 for $\mu\mu$ and $e\mu$ channels after b-tagging.

6 Conclusion

In summary, we have defined a counting method used to measure the top quark pair production in dilepton final states (ee, $\mu\mu$ channels and $e\mu$ including dilepton final states of τ decays) with an event sample collected during the early data taking ($\mathcal{L} = 100$ pb^{-1}) . This method is based on a simple and robust selection completed with a b-jet identification and the residual background is found to be very small. Assuming a $t\bar{t}$ cross section of 833 pb, the $t\bar{t}$ cross section can be measured with a statistical uncertainty of 8%.

References

[1] R. Bonciani, S. Catani, M. L. Mangano, and P. Nason, "NLL resummation of the heavy-quark hadroproduction cross-section," *Nucl. Phys.* **B529** (1998) 424–450, arXiv:hep-ph/9801375.

[2] **CMS** Collaboration, "Expectations for observation of top quark pair production in the dilepton final state with the first 10 pb^{-1} of CMS data," *CMS PAS* **TOP-08-001** (2008).

[3] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions," *JHEP* **07** (2003) 001, arXiv:hep-ph/0206293.

[4] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

[5] **CMS** Collaboration, "CMS High Level Trigger," *CERN/LHCC* **2007-021** (2007).

[6] **CMS** Collaboration, "Physics Technical Design Report, Vol. 1: Detector Performance and Software," *CERN/LHCC* **2006-001** (2006).

[7] A. Rizzi, F. Palla, and G. Segneri, "Track impact parameter based b-tagging with CMS," *CMS Note* **2006/019** (2006).

[8] **CMS** Collaboration, "b tag efficiency from System 8 Ptrel Method," *CMS PAS* **BTV-07-001** (2007).

[9] **CMS** Collaboration, "Measuring uds mistag rate of b tag using negative tags," *CMS PAS* **BTV-07-002** (2007).