Journal of Physics: Conference Series **452** (2013) 012053 doi:10.1088/1742-6596/452/1/012053

Combined measurement of the $t\bar{t}$ cross section and the b-tagging efficiency using l+jet events with CMS

M. Maes on behalf of the CMS Collaboration

Vrije Universiteit Brussel, Inter-university Institute for High Energies
Brussels, Belgium

E-mail: michael.maes@cern.ch

Abstract. The Large Hadron Collider, build at CERN, is the first collider in the world able to act as a top quark factory. The large amounts of top quark events produced allow us to study the properties of this heavy quark in great detail. We propose a combined measurement of the $t\bar{t}$ cross section and the b-tagging efficiency using the lepton+jets event topology observed by the CMS experiment in the 8 TeV proton collisions of the LHC. The data collected in 2012 used for this measurement corresponds to an integrated luminosity of 2.8 fb^{-1} for muon+jets events and 2.7 fb^{-1} for electron+jets events. The obtained $t\bar{t}$ cross section is 228.4 $\pm$ 9.0 (stat.) $^{+29.0}_{-26.0}$ (syst.) $\pm$ 10.0 (lumi) pb.

1. Introduction

The measurement of the top quark pair cross section at the LHC [1] is important for multiple reasons. The measurement on its own is a test of the Standard Model as well as a benchmark for the theoretical calculations. The result is also relevant for those searches beyond the Standard Model for which the top quark pair processes are the dominant background. In the decay of top quarks bottom quarks are present as the CKM element V_{tb} is constrained to be close to unity. Therefore the top quark pair signal, two b quarks and two W bosons in the final state, can be isolated from the background in the proton collisions at the LHC by use of b-quark identification algorithms or so-called b-tagging [2]. Although the application of b-tagging will purify the selected sample, it will also introduce important systematic uncertainties due to our understanding of the performance of these b-tagging algorithms. With a combined measurement of the top quark pair cross section with the b-tagging efficiency we will be able to remove this additional uncertainty. We present the potential of performing this combined measurement in a maximally data-driven way at the CMS [3] experiment. The b-tagging efficiency will be estimated from a specific sample which is enhanced in b-quark jets. In a data-driven way the non b-quark jets will be subtracted, and the measured distribution of the b-tag variable will be obtained for true b-quark jets. The b-tagging efficiency at different working points or thresholds can be determined from this distribution. Using the same event sample, but after applying a b-tag cut in the event selection, the top quark pair cross section will be obtained from a template fit on a kinematic variable which differentiates between the top quark process and the main background.

The measurement will be applied on the semi-leptonic decaying top quark pairs, namely $t\bar{t} \to bWbW \to bqqb\ell\nu_\ell$. The lepton is either a muon or an electron. Both studied decay modes


Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

account for about 15% of the total branching ratio of top quark pairs.

2. Event selection and reconstruction

The results are obtained on a dataset of about $2.8 fb^{-1}$ proton-proton collisions recorded at the CMS experiment at a centre-of-mass energy of 8 TeV. A trigger is used requiring an isolated muon with a transverse momentum (p_T) exceeding 20GeV/c or an electron with a p_T exceeding 25 GeV/c. Moreover, the trigger requires the events to contain at least three jets with a p_T above 30 GeV/c.

Then the events are reconstructed using the ParticleFlow algorithm [4] and required to contain an isolated lepton. This lepton is either a muon with p_T exceeding 26GeV/c within a pseudorapidity of $|\eta| < 2.1$ or an electron with p_T exceeding 30GeV/c within a pseudorapidity of $|\eta| < 2.5$, both passing a series of tight identification criteria. Furthermore, we require at least four jets in the event with $p_T >$ 45,45,35 and 35 GeV/c respectively within the acceptance of $|\eta| < 2.5$. Finally events are rejected if they contain one or more looser leptons. No missing transverse energy ($\not\!\!E_T$) is used.

3. Measuring the $t\bar{t}$ cross section

The $t\bar{t}$ production cross section is measured using a template fit on a the invariant mass of the lepton and leptonic b-jet candidate ($M_{\ell b}$), shown in Figure 1. To reconstruct $M_{\ell b}$ distribution, the b quark jet in the $t \to b\mu\nu_\mu$ decay should be identified among all selected jets. A χ^2 sorting algorithm is applied using as input the top quark and W boson masses and widths from simulation to points out the b-jet candidate in each event. Since the $M_{\ell b}$ variable is sternly dependent on the $t\bar{t}$ decay kinematics it has proven to have good discriminate between $t\bar{t}$ events and the various backgrounds. The templates for $t\bar{t}$ and background distributions are obtained from simulation, after applying the event selection outlined in the previous section together with a requirement of medium b-tag on the jet associated to the leptonically decaying top quark, and are fitted on the distribution from data. A b-tag algorithm deriving a probability based on the 2D track impact parameter significance of the tracks inside the jet [2] is applied with a cut corresponding to a miss tag rate of 1%. The template fit results in a number of observed $t\bar{t}$ events. From this number, using Equation (1), the event selection efficiency and b-tagging efficiency, the cross section is obtained. The b-tagging efficiency entering the cross section calculation is measured simultaneously with the $t\bar{t}$ cross section.

$$\sigma_{t\bar{t}} = \frac{N_{t\bar{t}}}{L} = \frac{N_{t\bar{t}^{obs}}}{L} \times \frac{1}{\epsilon_{sel}\epsilon_{\chi^2_{cut}}\epsilon_{b-tag}} \tag{1}$$

The efficiencies ϵ_{sel} and $\epsilon_{\chi^2_{cut}}$ used in Equation (1) are determined on simulation and a scale factor is applied to account for differences with respect to data. The b-tagging efficiency is determined by reconstructing the b tag discriminant distribution for true b jets ($\hat{\Delta}_b$) in a fully data-driven way.

Before applying the b-tagging criteria on the b-jet associated to the leptonically decaying top quark in the events, these candidates for b-flavored jets are used to define a b candidate jet sample. This jet sample is divided into a b-flavor enriched and b-flavor depleted subsample. As shown in Figure 2 the subdivision is performed according to the $M_{\ell b}$ variable of the event, which yields the b enriched region as 50 GeV/c^2 $< M_{\ell b} <$ 140 GeV/c^2 and the b depleted region as 140 GeV/c^2 $< M_{\ell b} <$ 240 GeV/c^2. The subdivision is defined such that the $M_{\ell b}$ variable remains uncorrelated with the b-tagging discriminators in the selected enriched and depleted regions.

To obtain the true b-jet discriminator ($\hat{\Delta}_b$), the non-b jet contamination in the b flavor enriched subsample is removed by subtracting the b-tag discriminant distribution of the b-flavor depleted subsample (Δ_b^{depl}) from the b-flavor enriched subsample (Δ_b^{enr}).

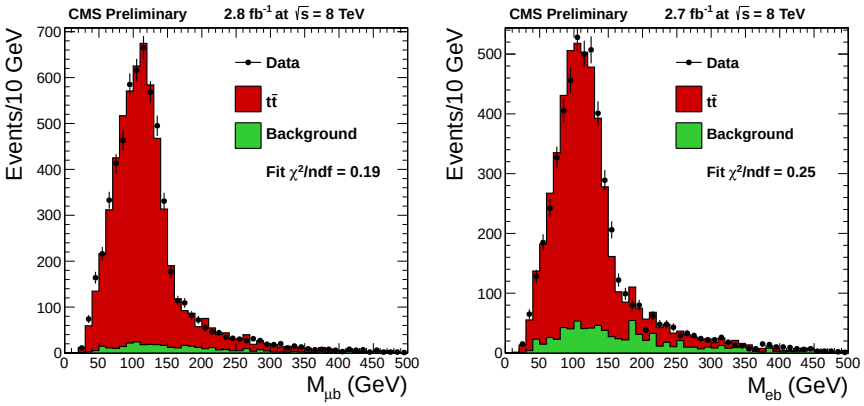


Figure 1. (left) Invariant mass distribution of the b jet candidate and the muon system and (right) Invariant mass distribution of the b jet candidate and the electron system. Signal and background distributions are normalised to the template fit results

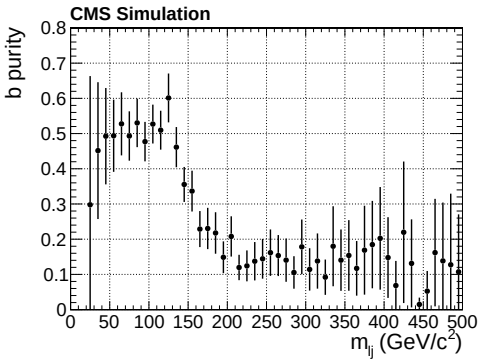


Figure 2. b-jet purity as a function of the invariant mass distribution of the b jet candidate and the muon system

$$\hat{\Delta}_b = \Delta_b^{enr} - F \times \Delta_b^{depl} \tag{2}$$

The factor F in Equation (2) represents the ratio between the number of non-b jets in the b-flavor enriched and b-flavor depleted subsamples which therefore rescales the number of non-b jets in the depleted region to the number of non-b jets in the enriched subsample. This factor F is obtained by constructing a control sample with a high purity in non-b jets. The control sample of jets is defined as those jets in the events which are associated with the hadronically decaying W boson. Additionally these jets in the control sample are requested to be anti-b-tagged to further reduce the b jet contamination. In the control sample, the same subsamples are defined as within the signal b candidate jet sample by use of the $M_{\ell b}$ variable which allows direct access to the factor F from data. The kinematics of the jets in the control sample are matched to those in the signal b candidate sample using reweighing techniques based on the observed data itself . Also in the data driven determination of F and in the application of Equation (2) the participating jets in the b-flavor depleted region are reweighed to match the kinematics of the jets in the b-flavor enriched region. From the estimated b-tag discriminator distribution, $\hat{\Delta}_b$, the b-tagging efficiency for diverse working points is estimated.

4. Results

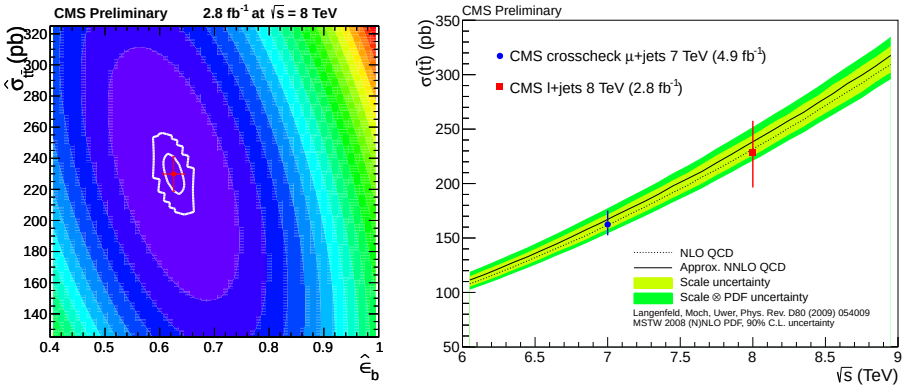


Figure 3. $\Delta\chi^2$ distribution in the ($\hat{\epsilon}_b$,$\hat{\sigma}_{t\bar{t}}$) plane showing the correlation between the two estimators within muon+jets events (left) and the cross section obtained with this analysis technique at 7 TeV and 8 TeV centre-of-mass energy [5, 6] (right)

On a dataset of 2.8 (2.7) fb^{-1} in the muon (electron) channel at $\sqrt{s} = 8$TeV recorded by the CMS experiment at the Large Hardron Collider during 2012, the $t\bar{t}$ cross-section was measured [5].

$$\sigma_{t\bar{t}}(\mu + jets) = 229.9 \pm 11.1(stat.) \pm^{27.6}_{29.0} (syst.) \pm 10.1(lumi.)pb$$

$$\sigma_{t\bar{t}}(e + jets) = 227.3 \pm 12.2(stat.) \pm^{35.5}_{30.0} (syst.) \pm 10.0(lumi.)pb$$

A combined fit to both channels yields the following cross section value.

$$\sigma_{t\bar{t}}(combined) = 228.4 \pm 9.0(stat.) \pm^{29.0}_{26.0} (syst.) \pm 10.0(lumi.)pb$$

Figure 3 shows the correlation between the b-tagging efficiency and the $t\bar{t}$ cross section from data which arises from the combined measurement of the two properties. The dominating systematic uncertainty in the measurements of the $t\bar{t}$ cross section arises from the uncertainty in the determination of the jet energy scale and the uncertainties on the b-tagging efficiency.

The combined fit result is also shown on the right side of Figure 3 together with the cross check result at $\sqrt{s} = 7$TeV provided by the same method in the muon+jets channel [6].

References

[1] N. Kidonakis, R. Vogt, *The theoretical top quark cross section at the Tevatron and the LHC*, **arXiv:0805.3844v1 [hep-ph]** (2008)
[2] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, **arXiv:1211.4462 [hep-ex]** (2012)
[3] CMS Collaboration **JINST, 3 S08004** (2008), doi=10.1088/1748-0221/3/08/S08004
[4] CMS Collaboration, *Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp collisions at $\sqrt{s}$ = 7 TeV*, **CMS-PAS-PFT-10-002** (2010)
[5] CMS Collaboration, *Top pair cross section in e/mu+jets at 8 TeV*, **CMS-PAS-TOP-12-006** (2012)
[6] CMS Collaboration, *Measurement of the t t-bar production cross section in pp collisions at sqrt(s) = 7 TeV with lepton + jets final states*, **arXiv:1212.6682 [hep-ex]** (2012)