PROCEEDINGS OF SCIENCE

Selection of Top-like events in pp collisions at $\sqrt{s}$ = 7 TeV in early CMS data

Pedro Q. Ribeiro* **on behalf of the CMS Collaboration**
LIP - Laboratório de Instrumentação e Física Experimental de Partículas, Lisbon, Portugal
E-mail: ribeiro@lip.pt

Studies are presented of the selection of events consistent with top quark pair production in data recorded by the CMS detector at the LHC, corresponding to an integrated luminosity of 0.84 pb^{-1} and a center-of-mass energy $\sqrt{s}$ = 7 TeV. Results are presented for the lepton+jets and dilepton channels. Events yields in data are compared to those in simulation and several background processes are estimated using data-driven techniques. The observed yields of top-antitop candidate events are roughly consistent with the Standard Model.

The Xth Nicola Cabibbo International Conference on Heavy Quarks and Leptons,
October 11-15, 2010
Frascati (Rome) Italy

*Speaker.

PoS(HQL 2010)064

1. Introduction

The study of top quark physics is an important part of the CMS research program. Its abundant production at the LHC will allow the measurement of fundamental properties such as the production cross section, mass, couplings and rare decays with improved precision with respect to the Tevatron results [1].

Unlike the other known Standard Model quarks, the top decays rapidly before having the chance to form a bound state hadron, therefore it presents an opportunity to measure directly the properties of bare quark production. The large value of the top quark mass [2] has important implications. Its coupling to the Standard Model Higgs is large, therefore a detailed study of top quark properties can provide additional constraints on the undetected Higgs boson. Furthermore, new physics might couple preferentially to top quark. Clear signatures of new physics would be the decay of new massive particles to the top or the observation of top rare decay modes at a rate inconsistent with expectations, indicating non-standard top couplings. Additionally, several signatures of new physics at the LHC suffer from top-quark production as a significant background.

At the LHC, the top quark is expected to be produced primarily via the strong interaction in $\mathrm{t\bar{t}}$ pairs, mostly trough gluon fusion. Within the Standard Model, the top quark decays via the weak process $\mathrm{t} \rightarrow \mathrm{Wb}$ with a branching ratio very close to 100%; subsequently top-quark events are characterized according to the decay of the W bosons. Two channels are the focus of the studies discussed bellow : the dilepton channel, in which both W bosons decay leptonically to either an electron or a muon and its associated neutrino; and the lepton+jets channel, in which one of the W bosons decays leptonically while the other decays to quarks which subsequently hadronize, resulting in jets of charged and neutral particles.

In this note, results on top quark physics obtained with 0.84 pb^{-1} of data of proton-proton collisions at 7 TeV center-of-mass energy are reported [3, 4].

2. Data and Simulated Samples

The results reported on this note are based on the analysis of a selected sample of data corresponding to an integrated luminosity of 0.84 ± 0.09 pb^{-1} [5], recorded by CMS [6] up to August 2010. Data events are required to come from periods in which the CMS detector was fully operational. Additionally, events are vetoed if they are identified as resulting from beam halo interactions or from beam scraping. Finally, events are required to possess at least one well-reconstructed primary vertex within $|z| < 15$ cm.

Simulated samples of top-quark pair production events are made using the MADGRAPH event generator [7], subsequently processed with PYTHIA [8], and then processed with a full CMS detector simulation based on GEANT4 [9]. Events are generated with up to four additional hard partons. Various background samples were produced. MADGRAPH is used for $W/Z/\gamma$+jets production and single top. Leptonic taus decays are included in the Drell-Yan samples. PYTHIA is used to generate QCD events used in the study of the multijet backgrounds.

The top-quark pair production simulation has been normalized using a NLO cross section of $\sigma_{t\bar{t}} = 157.5^{+23.2}_{-24.4}$ pb, obtained using MCFM [10]. The uncertainty in the cross section includes the scale uncertainties, determined by varying the factorization and renormalization scales by a factor

2 and 0.5 around the central scale choice of m_t = 172.5 GeV/c^2, and the uncertainties from the parton distribution functions and the value of the strong coupling α_S. Similarly, the simulations of $W/Z/\gamma$+jets production and single top production have been normalized using available inclusive N(N)LO cross section calculations.

3. Lepton+jets channel

Both the e+jets and the μ+jets modes are considered in the lepton+jets channel. Events passing a single muon or electron trigger which contain exactly one high transverse momentum p_T lepton are selected. The lepton is required to be isolated within a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$, using a relative isolation variable which employs sums of track transverse momenta and calorimeter transverse energy deposits, scaled to the lepton p_T. The lepton must be consistent with originating from the primary hard interaction, both in the transverse plane as well as along the beam direction. For e + jets, electrons [11] passing tight identification criteria based on cluster shape properties and track-cluster matching, inconsistent with originating from photon conversion and fulfilling p_T > 30 GeV/c and pseudorapidity $|\eta| < 2.4$ are selected. For μ + jets, high quality muons [12] with $p_T > 20$ GeV/c and $|\eta| < 2.1$ are selected. Jets [13] and missing transverse energy (MET) [14] are reconstructed using calorimeter information. Jets are clustered using the anti-k_T algorithm [15] with R = 0.5 . Jet energies are corrected to achieve uniform response in η (relative) and p_T (absolute). The jet energy scale uncertainty is estimated as 5%. Jets are required to have $p_T > 30$ GeV/c and $|\eta| < 2.4$ and must not overlap with any electron or muon within $\Delta R < 0.4$. No explicit MET requirement is used in the event selection. At least four jets are expected for the signal.

Event yields as function of the jet multiplicity are shown in Figure 1. Good agreement between data and simulation, scaled to the integrated luminosity of data, is observed in all jet bins. As expected, the signal-to-background ratio increases with the jet multiplicity.

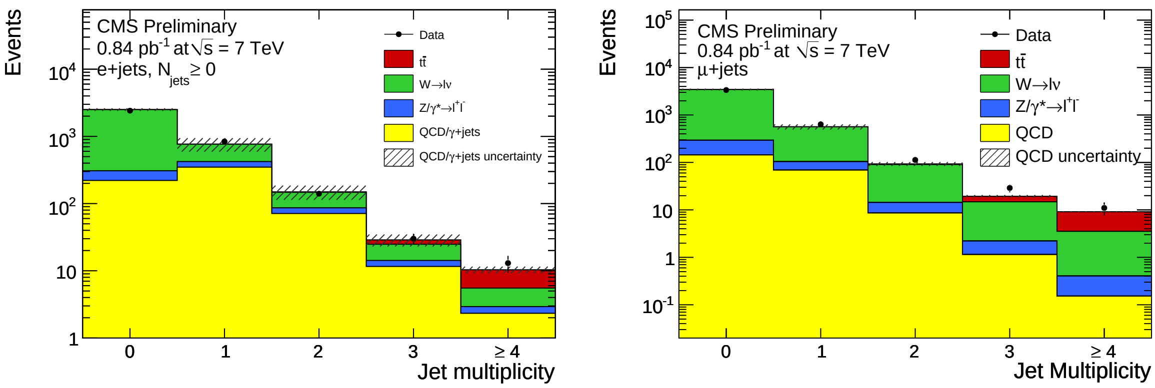

Figure 1: Jet multiplicity for the e+jets (left) and the μ+jets (right) event selections, without b-tagging.

Several methods are studied which allow a data-driven estimate of the amount of QCD background in the selected samples, which is not expected to be realistically modeled by simulation in the low p_T kinematic region. In both e+jets and μ+jets, a method based on the lepton isolation is employed. The isolation distribution is fitted with a suitable function in the non-isolated (QCD dominated) sideband region, which is then extrapolated into the isolated signal region. Another method, often referred as the *ABCD* method, exploits two nearly uncorrelated variables (lepton impact parameter and lepton isolation) which separate signal and QCD background in μ+jets. A third method, applied in e+jets, is based on a template fit of the MET or $H_{T,lep}$ = MET + $p_{T,lep}$ distribution, using a data-driven QCD template. Two models are considered in order to obtain template distributions for QCD multijet events: "background" electrons, in which the electron candidate very nearly satisfies the selection criteria but instead is a marginal failure; and "jet-electrons", positively identified jet objects with large electromagnetic fraction, that closely resemble electron candidates. In e+jets (μ+jets), a 50% (100%) systematic uncertainty is assigned to the QCD background, based on the data-driven estimates.

A signature of the top-quark production are jets from b-quarks. One expects two b-jets to be present in every event in case of signal. For the important backgrounds processes such as QCD and W+jets, which contain a mixture of light and heavy quarks, less b-jets are expected. In Figure 2 is shown the jet multiplicity for the combined e+jets and μ+jets event selection after requesting at least one b-tagged jet, using a secondary vertex tagger [16].

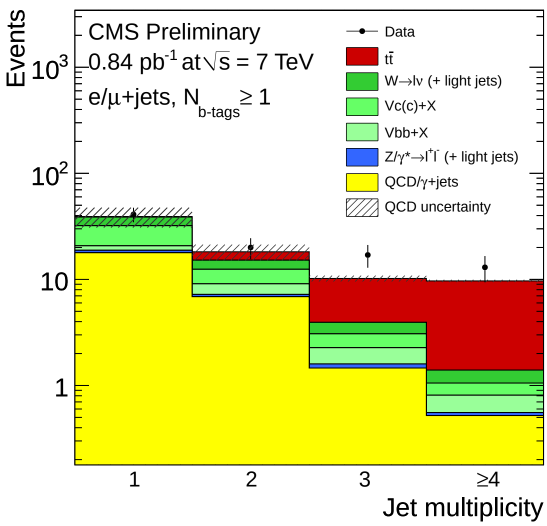

Figure 2: Jet multiplicity for the combined selection of e+jets and μ+jets where at least one the jets is b-tagged using a secondary vertex algorithm.

As expected, the purity of the selection is enhanced by selecting events with a least one b-jet. For $N_{\text{jets}} \geq 3$, 30 events are observed in data, in a region where ~ 5.5 events are expected from non-top background. The observed rates are roughly consistent with expectations within uncertainties.

4. Dilepton channel

Dilepton events in the dielectron (ee), dimuon ($\mu\mu$) and electron-muon ($e\mu$) modes are considered. Events passing a single muon or electron trigger which contain two oppositely charged, isolated high p_T leptons with $p_T > 20$ GeV/c and $|\eta| < 2.5(2.4)$ for muons (electrons) are selected. Reconstructed muons are required to have high quality. Electrons passing tight identification criteria and inconsistent with originating from photon conversion are selected. For ee and

$\mu\mu$ candidates, the dilepton invariant mass is required to satisfy $|M_{ll} - M_Z| > 15$ GeV/c^2, to reject Z events. Jets and MET are calculated from calorimeter signals and are made more accurate by applying track-based correction for the inexact calorimeter response. A cut MET > 30 (20) GeV/c^2 is applied in the ee, $\mu\mu$ ($e\mu$) channels. Jets are required to have $p_T > 30$ GeV/c and $|\eta| < 2.4$ and must not overlap with any electron or muon within $\Delta R < 0.4$. At least two jets are requested for the full event selection.

Distributions for a relaxed event selection (without jets+MET requirements, no Z veto applied) are shown in Figure 3.

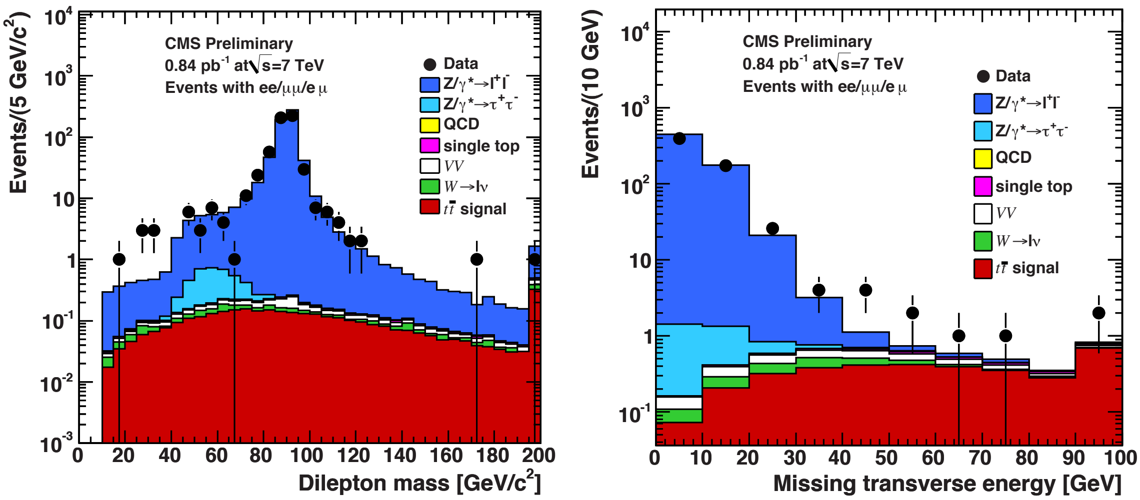

Figure 3: Distributions of the dilepton invariant mass (left) and missing transverse energy (right) for a relaxed event selection, as described in the text.

Good agreement is observed between simulation and data, scaled to the integrated luminosity of data.

For several sources of backgrounds, data-driven estimation techniques are tested. Drell-Yan events passing the Z veto are estimated by counting data events rejected by this veto, scaled by the ratio of events outside and inside the veto region, obtained from simulation. The estimated systematic uncertainty of the method is 50%. Background from events with non-genuine isolated leptons (i.e. not originating from W/Z decays) is estimated by weighting events passing loose lepton identification with a tight-to-loose ratio which is parametrized in p_T and η, measured in a inclusive QCD sample. The method is used to estimate the contributions from QCD multi-jet and W+jets events, containing two and one non-genuine lepton respectively, with a 50% systematic uncertainty per lepton. The data-driven estimates are in reasonable agreement with expectations from simulation.

Applying the full event selection, including Z-veto, MET requirement and requesting at least two jets, there are four events selected in the sample. The expected non-top background from simulation is less than 0.3 events, while 2.1 top signal events are expected. Figure 4 shows the b-jet multiplicity distribution using a b-tag algorithm based on the impact parameter significances of the tracks associated with the jets. A loose working point with 80% b-jet efficiency and 10% mistagging rate in QCD events is used. The observed events are consistent with a top-antitop hypothesis.

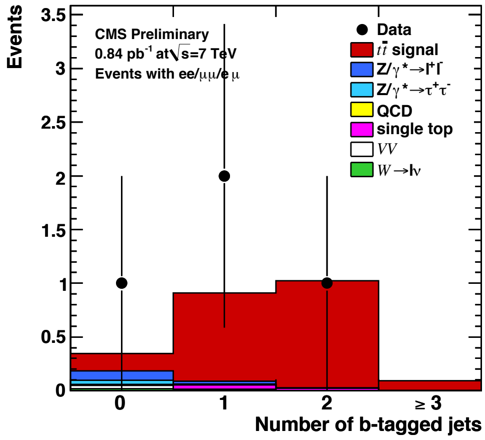

Figure 4: Distributions of the b-jet multiplicity for dilepton candidate events passing the full event selection.

5. Dilepton candidate event

A $\mu\mu$ candidate event which is passing full event selection is shown in Figure 5. This event has two opposite-sign muons (p_T = 57 and 27 GeV/c), with dimuon mass of 26 GeV/c^2, two tagged b-jets (p_T = 56 and 45 GeV/c) and large MET = 57/c^2 GeV. The reconstructed masses found in this event are consistent with a $t\bar{t}$ hypothesis.

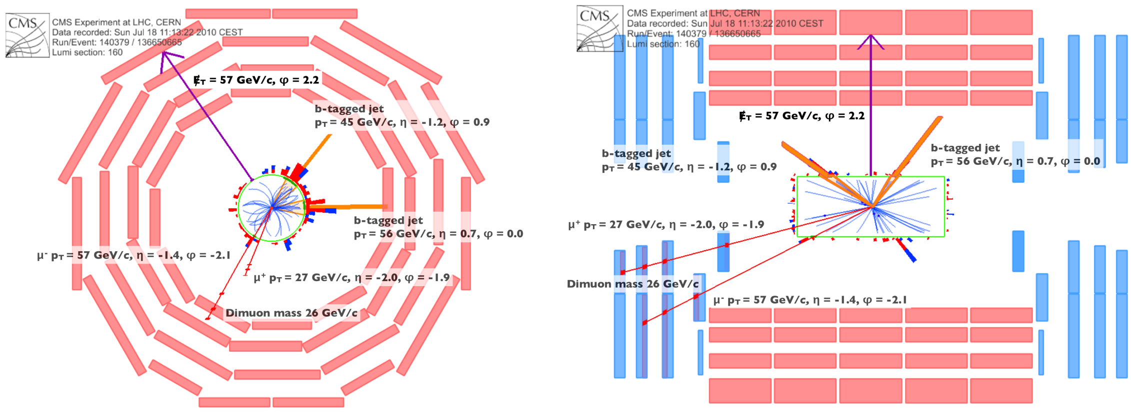

Figure 5: Display of an $\mu\mu$ candidate event in the $r\phi$ view (left) and the ρz view (right).

6. Conclusions

In both the lepton+jets and dilepton channel, events are observed in signal regions expected to be dominated by top quark pair production. The observed rates are roughly consistent with current theory expectations for top quark pair production, taking into account the experimental uncertainties due to e.g. jet energy scale, b-tagging performance, but also the theoretical uncertainties (e.g. scale and parton distributions for top signal, heavy flavor treatment for W/Z+jets backgrounds).

Acknowledgments

The author wishes to thank the members of the CMS Top Quark Physics Analysis Group for their hard work documented in this note.

References

[1] J. R. Incandela, A. Quadt, W. Wagner and D. Wicke, *Status and Prospects of Top-Quark Physics*, Prog. Part. Nucl. Phys. **63** (2009) 239 [arXiv:0904.2499 [hep-ex]].

[2] CDF and D0 Collaboration, *Combination of CDF and D0 Results on the Mass of the Top Quark*, arXiv:1007.3178 [hep-ex].

[3] CMS Collaboration, *Selection of Top-Like Events in the Dilepton and Lepton-plus-Jets Channels in Early 7 TeV Data*, Physics Analysis Summary **PAS TOP-10-004** (2010).

[4] Updated results for $\mathscr{L} = 0.84$ pb^{-1} (August 2010):
https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPAug2010LeptonPlusJets
https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPAug2010Dileptons

[5] CMS Collaboration, *Measurement of CMS Luminosity*, Physics Analysis Summary **PAS EWK-10-004**(2010).

[6] R. Adolphi *et al.* [CMS Collaboration], *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[7] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, JHEP **0302** (2003) 027 [arXiv:hep-ph/0208156].

[8] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026 [arXiv:hep-ph/0603175].

[9] J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53**, 270 (2006).

[10] J. M. Campbell and R. K. Ellis, *MCFM for the Tevatron and the LHC*, arXiv:1007.3492 [hep-ph].

[11] CMS Collaboration, *Electron Reconstruction and Identification at sqrt(s) = 7 TeV*, Detector Performance Note **DP 2010/032**(2010).

[12] CMS Collaboration, *Performance of muon identification in pp collisions at sqrt(s) = 7 TeV*, Physics Analysis Summary **PAS MUO-10-002**(2010).

[13] CMS Collaboration, *Jet Performance in pp Collisions at sqrts=7 TeV*, Physics Analysis Summary **PAS JME-10-003**(2010).

[14] CMS Collaboration, *Missing Transverse Energy Performance in Minimum-Bias and Jet Events from pp Collisions at sqrts=7 TeV*, Physics Analysis Summary **PAS JME-10-004**(2010).

[15] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **0804** (2008) 063 [arXiv:0802.1189 [hep-ph]].

[16] CMS Collaboration, *Commissioning of b-jet identification with pp collisions at sqrt(s) = 7TeV*, Physics Analysis Summary **PAS BTV-10-001**(2010).