Observation of the Associated Production of a Single Top Quark and a W Boson in pp Collisions at √s = 8 TeV

S. Chatrchyan et al.*
(CMS Collaboration)

(Received 13 January 2014; published 9 June 2014)

The first observation of the associated production of a single top quark and a W boson is presented. The analysis is based on a data set corresponding to an integrated luminosity of 12.2 fb⁻¹ of proton-proton collisions at √s = 8 TeV recorded by the CMS experiment at the LHC. Events with two leptons and a jet originating from a b quark are selected. A multivariate analysis based on kinematic and topological properties is used to separate the signal from the dominant t¯t background. An excess consistent with the signal hypothesis is observed, with a significance which corresponds to 6.1 standard deviations above a background-only hypothesis. The measured production cross section is 23.4 ± 5.4 pb, in agreement with the standard model prediction.

The central feature of the CMS apparatus [23] is a superconducting solenoid with an internal diameter of 6 m, providing a magnetic field of 3.8 T. Within the bore of the solenoid are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter. Muons are measured in gas-ionization detectors embedded in the steel flux return yoke outside the magnet. In addition, CMS has extensive electromagnetic calorimeter (including energy deposits in the tracker). Muons, electrons, and hadrons are measured in the向外磁场.

Jets are reconstructed by clustering PF candidates using the anti-kt algorithm [27] with a distance parameter of 0.5. Selected jets must be within |η| < 2.4 and have p_T > 30 GeV. Corrections are made to the jet energies for detector response as a function of η and p_T [28]. Additional corrections are made to subtract energy in the jet from multiple pp collisions (pileup) [29]. Jets originating from the decay of a b quark are tagged based on the presence of a secondary vertex, identified using a multivariate algorithm combining tracking information in a discriminant. A working point is chosen, corresponding to b-tagging efficiencies of approximately 70% and with a misidentification rate of 1%–2%. Loose jets, whose discrimination power against tt̄ background is discussed later, are defined as jets failing the requirements on p_T and η, but passing the less restrictive selection requirement of p_T > 20 GeV and |η| < 4.9, while still passing all other selection criteria. In particular, loose jets that fall within |η| < 2.4 are classified as central loose jets.

For events passing the dilepton and E_T^{miss} criteria described above, a region in which the tt̄ signal is enhanced (signal region) and two regions dominated by background (control regions) are defined. The signal region contains events with exactly one jet passing the selection requirements, which is b tagged (1j1l region). Two control regions enriched in tt̄ background are defined as having exactly two jets with either one or both being b tagged (2j1l and 2j2l regions, respectively).

Events from Monte Carlo simulation are used to estimate the contributions and kinematics of signal and background processes. Single-top-quark events are simulated at NLO with the Powheg 1.0 event generator [31–34]; MadGraph 5.1.3 is used for simulating tt̄ and single-boson events (V + jets, where V = W, Z) [35]. Samples are produced using a top-quark mass m_t = 172.5 GeV, consistent with its current best measurement [36]. Diboson backgrounds are simulated using Pythia 6.426 [37]. In all samples, fragmentation and hadronization are modeled with Pythia, and Tauola v27.121.5 is used to simulate τ decays [38]. The CTQ6L1 and CTQ6.6M PDF sets [39] are used for samples simulated at leading-order and NLO, respectively. A full simulation of the response of the CMS detector is performed for all generated events using a Geant4-based model [40]. The simulation includes modeling of pileup, with the distribution of the number of interactions in simulation matching that in data. Simulated samples are normalized to the NNLO cross sections for tt̄ [σ_R = 245.8 ± 6.2 (scale) ± 6.4 (PDF) pb] [41], Z/γ* , and W + jets processes, with approximate NNLO cross sections used for single top quark [18] and NLO for diboson processes. The Z/γ* simulation is reweighted to reproduce the E_T^{miss} distribution observed in data, using events with m_ℓℓ in the vicinity of the Z-boson mass (81 to 101 GeV) to derive scale factors.

After the selection, the simulated samples in the 1j1l signal region contain predominantly tt̄W and tt̄ events (comprising 16% and 76% of the events, respectively), with a smaller contribution from Z/γ* events (6%). The two control regions are dominated by tt̄ production. Event yields in simulation and data in the signal and control regions are shown in Table I.

In order to separate the tt̄W signal from the tt̄ background, a multivariate analysis based on boosted decision
trees (BDT) [42] is used, implemented with the toolkit for multivariate data analysis [43]. The BDT analyzer is trained using 13 variables, chosen for their separation power in distinguishing \( tW \) and \( \bar{t}t \), as well as being well modeled in simulation when checked in control regions. The most powerful variables are those involving loose jets in the event: the number of loose jets, number of central loose jets, and the number of loose jets that are \( b \) tagged. Other variables with significant separation power are related to the kinematics of the system comprised of the leptons, jets and \( E_{\text{T}}^{\text{miss}} \): the scalar sum of their transverse momenta \( (H_T) \), the magnitude of the vector sum of their transverse momenta \( (p_T^{\text{miss}}) \), and invariant mass of the system. A complete list of the variables used can be found in the Supplemental Material [44]). The distributions of the number of loose jets and the \( p_T \) of the system in the \( 1j1t \) signal region are shown in Fig. 1 for all three final states \((ee, e\mu, \mu\mu)\) combined.

The BDT analyzer provides a single discriminant value for each event. The distributions of the BDT discriminant in data and simulation are shown in Fig. 2 for the \( 1j1t \), \( 2j1t \), and \( 2j2t \) regions, combining all three final states together.

The uncertainty from all systematic sources is determined by estimating their effect on the normalization and shape of the BDT discriminant for all regions and final states. The dominant systematic uncertainties come from the choice of thresholds for the matrix element and parton showering (ME/PS) matching in simulation of \( \bar{t}t \) production and the renormalization and factorization scale. The effect of these uncertainties was estimated by producing simulated samples with the value of the ME/PS matching thresholds and renormalization and factorization scale doubled and halved from their respective initial values of 20 GeV and \( m_t^2 + \sum p_T^2 \) (where the sum is over all additional final state partons), contributing a 14% and 12% uncertainty, respectively, to the measured cross section. The uncertainty due to the value of the top-quark mass used in simulation is estimated by simulating \( tW \) and \( \bar{t}t \) processes with a varied value for \( m_t \), resulting in a 9% effect on the cross section. The complete list of systematic uncertainties and corresponding effects on the cross section can be found in the Supplemental Material [44].

A simultaneous binned likelihood fit to the rate and shape of the BDT distributions of the three final states in the three regions is performed. The two control regions are included in the fit to allow for better determination of the \( \bar{t}t \) contribution. The distributions for signal and background are taken from simulation. In the likelihood function, for each source of systematic uncertainty \( u \), a nuisance parameter \( \theta_u \) is introduced. The rates of signal and background are allowed to vary in the fit, constrained in the likelihood function by the systematic uncertainties. The excess of events is quantified based on the score statistic \( q \), chosen to enhance numerical stability, defined as

<table>
<thead>
<tr>
<th>( rW )</th>
<th>( \bar{t}t )</th>
<th>( Z/\gamma^* ) , other</th>
<th>Total simulation</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1500 ± 20 ± 130</td>
<td>790 ± 20 ± 80</td>
<td>12910 ± 80 ± 1320</td>
<td>9260 ± 70 ± 1040</td>
<td>9353</td>
</tr>
<tr>
<td>7090 ± 60 ± 900</td>
<td>12910 ± 80 ± 1320</td>
<td>14070 ± 90 ± 1410</td>
<td>1410 14070</td>
<td>13479</td>
</tr>
<tr>
<td>670 ± 30 ± 90</td>
<td>370 ± 30 ± 60</td>
<td>36 ± 7 ± 12</td>
<td>7910 ± 70 ± 1020</td>
<td>7615</td>
</tr>
</tbody>
</table>

TABLE I. Event yields in the signal and control regions. Yields from simulation are shown with statistical (first) and systematic (second) uncertainties.

FIG. 1 (color online). The number of loose jets in the event and the \( p_T \) of the system \((p_T^{miss})\) composed of the jet, leptons, and \( E_{\text{T}}^{\text{miss}} \), in the signal region \((1j1t)\) for all final states combined. Shown are data (points) and simulation (histogram). The hatched band represents the combined effect of all sources of systematic uncertainty.
The signal strength parameter (defined as the cross section divided by the SM prediction) is determined by a profile likelihood method that is used to find the central 68% confidence level (C.L.) interval of the score statistic values obtained in pseudoexperiments using a background-only hypothesis. The significance is determined based on the probability of producing a score statistic value in the background-only hypothesis as high or higher than that observed in data. The expected significance is evaluated using the median and central 68% interval of the score statistic values obtained in pseudo-experiments generated under a signal-plus-background hypothesis. A profile likelihood method is used to determine the signal cross section and 68% confidence level (C.L.) interval.

We observe an excess of events above the expected background with a $p$ value of $5 \times 10^{-10}$ corresponding to a significance of $6.1\sigma$, compared to an expected significance from simulation of $5.4 \pm 1.4\sigma$. The measured cross section is found to be $23.4 \pm 5.4\text{ pb}$, where the uncertainty is mainly systematic, in agreement with the predicted SM value of $22.2 \pm 0.6\text{ (scale)} \pm 1.4\text{ (PDF)}\text{ pb}$.

The cross section measurement is used to determine the absolute value of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{tb}|$, assuming $|V_{tb}| \gg |V_{td}|$ and $|V_{ts}|$

$$|V_{tb}| = \sqrt{\frac{\sigma_{tw}}{\sigma_{tw}^{th}}} = 1.03 \pm 0.12(\text{exp}) \pm 0.04(\text{th.}),$$

where $\sigma_{tw}^{th}$ is the theoretical prediction of the $tW$ cross section assuming $|V_{tb}| = 1$, and the uncertainties are separated into experimental and theoretical values. Using the SM assumption $0 \leq |V_{tb}|^2 \leq 1$, a lower bound $|V_{tb}| > 0.78$ at 95% C.L. is found using the approach of Feldman and Cousins [45].

Using the same selection as in the BDT analysis, two cross-check analyses are performed. Events containing any
The production of a single top quark in association with a W boson is observed for the first time. The analysis uses data collected by the CMS experiment in pp collisions at $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 12.2 fb$^{-1}$. An excess of events above background is found with a significance of 6.1σ, and a $t\bar{t}$ production cross section of 23.4±5.4 pb is measured, in agreement with the standard model prediction.

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing centers and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMWF and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES and CSF (Croatia); RPF (Cyprus); MoER, Contract No. SF069003s09, and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and NIH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); NRF and WCU (Republic of Korea); LAS (Lithuania); CINVESTAV, CONACYT, SEP, and UASLP-FAI (Mexico); MBIE (New Zealand); PAEC (Pakistan); MSHE and NSC (Poland); FCT (Portugal); JINR (Dubna); MON, RosAtom, RAS, and RFBR (Russia); MESTD (Serbia); SEIDI and CPAN (Spain); Swiss Funding Agencies (Switzerland); NSC (Taipei); ThEPCenter, IPST, STAR, and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU (Ukraine); STFC (United Kingdom); DOE and NSF (USA).
[44] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.112.231802 for a list and distributions of the variables used in the BDT, BDT distributions in additional control regions, the event yields split by channel, and a list of the systematic uncertainties.

(CMS Collaboration)
Institute of Nuclear and Particle Physics (INPP), NCSR Demokritos, Aghia Paraskevi, Greece
University of Athens, Athens, Greece
University of Ioannina, Ioannina, Greece
Wigner Research Centre for Physics, Budapest, Hungary
Institute of Nuclear Research ATOMKI, Debrecen, Hungary
University of Debrecen, Debrecen, Hungary
National Institute of Science Education and Research, Bhubaneswar, India
Panjab University, Chandigarh, India
University of Delhi, Delhi, India
Saha Institute of Nuclear Physics, Kolkata, India
Bhabha Atomic Research Centre, Mumbai, India
Tata Institute of Fundamental Research - EHEP, Mumbai, India
Tata Institute of Fundamental Research - HECR, Mumbai, India
Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
University College Dublin, Dublin, Ireland
INFN Sezione di Bari, Università di Bari, Politecnico di Bari, Bari, Italy
INFN Sezione di Bologna, Università di Bologna, Bologna, Italy
INFN Sezione di Catania, Università di Catania, CSFNSM, Catania, Italy
INFN Sezione di Firenze, Università di Firenze, Firenze, Italy
INFN Laboratori Nazionali di Frascati, Frascati, Italy
INFN Sezione di Genova, Università di Genova, Genova, Italy
INFN Sezione di Milano-Bicocca, Università di Milano-Bicocca, Milano, Italy
INFN Sezione di Napoli, Università di Napoli 'Federico II', Università della Basilicata (Potenza), University G. Marconi (Roma), Napoli, Italy
INFN Sezione di Padova, Università di Padova, Università di Trento (Trento), Padova, Italy
INFN Sezione di Pavia, Università di Pavia, Pavia, Italy
INFN Sezione di Pisa, University of Pisa, Scuola Normale Superiore di Pisa, Pisa, Italy
INFN Sezione di Torino, Università di Torino, Università del Piemonte Orientale (Novara), Torino, Italy