Search for pair production of a heavy up-type quark decaying to a W boson and a b quark in the lepton+jets channel with the ATLAS detector

Physical Review Letters, 108(26)

0031-9007

Aad, G
Abbott, B
Abdallah, J
et al.

2012-06-26

10.1103/PhysRevLett.108.261802

CC BY 4.0

Peer reviewed
Search for Pair Production of a Heavy Up-Type Quark Decaying to a W Boson and a b Quark in the lepton + jets Channel with the ATLAS Detector

G. Aad et al.*
(Atlas Collaboration)
(Received 14 February 2012; published 26 June 2012)

A search is presented for production of a heavy up-type quark (t') together with its antiparticle, assuming subsequent decay to a W boson and a b quark, \( t'\rightarrow W^+ bW^-\bar{b} \). The search is based on 1.04 fb\(^{-1}\) of proton-proton collisions at \( \sqrt{s} = 7 \) TeV collected with the ATLAS detector at the CERN Large Hadron Collider. Data are analyzed in the lepton + jets final state, characterized by a high transverse momentum isolated electron or muon, high missing transverse momentum, and at least three jets. No significant excess of events above the background expectation is observed. A 95% C.L. lower limit of 404 GeV is set for the mass of the t' quark.

DOI: 10.1103/PhysRevLett.108.261802
PACS numbers: 14.65.Jk, 12.60.—i, 13.85.Rm, 14.80.—j

The discovery of the top quark [1] completed the third generation of fundamental fermions in the quark sector of the standard model (SM) of particle physics. It is natural to ask whether heavier quarks may exist. These quarks are often present in new physics models aimed at solving the limitations of the SM. For example, models with a fourth generation of heavy chiral fermions could provide new sources of CP violation to explain the matter-antimatter asymmetry in the Universe and allow for a heavier Higgs boson while remaining consistent with precision electroweak data [2]. The latter is accomplished by keeping a small mass splitting between the heavy up-type quark (t') and the heavy down-type quark (b'). Assuming that \( m_t - m_b < m_W \), where \( m_W \) is the W boson mass, results in the t' quark predominantly decaying to a W boson and a down-type quark q (q = d, s, b). Another possibility is the addition of isospin singlets or doublets of vectorlike quarks, which appear in many extensions of the SM such as little Higgs or extra-dimensional models [3]. In both scenarios, the t' quark can decay into Wb with a large branching ratio, provided there is a significant mixing with the third generation of quarks, consistent with the existing mass and mixing patterns of the known quarks.

The high center-of-mass energy and integrated luminosity in pp collisions available at the Large Hadron Collider (LHC) offer a unique opportunity to probe these scenarios. At the LHC, these new heavy quarks would be predominantly produced in pairs via the strong interaction for masses below ~1 TeV, while for larger masses electroweak production of single heavy quarks could become the primary production mechanism, depending on the strength of their interactions with the SM quarks and weak gauge bosons [3].

A search is presented in this Letter for t'\(\bar{t}'\) production using pp collision data at \( \sqrt{s} = 7 \) TeV collected with the ATLAS detector. It is assumed that the t' quark decays exclusively into Wb. The lepton + jets final-state signature is considered, characterized by a high transverse momentum (pT) isolated electron or muon, high missing transverse momentum (Emiss\( T \)), and at least three jets. Similar searches in this channel have been published by the CDF and D0 Collaborations [4,5]; the most stringent limits preclude the existence of a t' quark with a mass below 358 GeV at 95% confidence level (C.L.). A search for t'\(\bar{t}'\) in the dilepton final state has been performed by the ATLAS Collaboration [6], excluding a t' quark with a mass below 350 GeV at 95% C.L. The lepton + jets signature has also been recently exploited by the ATLAS Collaboration to search for b\(\bar{b}\) production with a mass below 2012 CERN, for the ATLAS Collaboration

*Full author list given at the end of the article.

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.
beam conditions and for which all detector subsystems were fully operational. The corresponding integrated luminosity is 1.04 fb$^{-1}$. The event selection criteria closely follow those used in recent ATLAS top quark studies, e.g., Ref. [10]. Electron candidates are required to satisfy $p_T > 25$ GeV and $|\eta| < 2.47$, excluding the transition region 1.37 $< |\eta| < 1.52$ between the barrel and end cap electromagnetic calorimeters. Muon candidates are required to satisfy $p_T > 20$ GeV and $|\eta| < 2.5$. The $p_T$ threshold requirement ensures that the selected leptons are in the efficiency plateau of the single-lepton triggers. Background from multijet production is suppressed by a requirement of $E_T^{\text{miss}} > 35(20)$ GeV [11] in the electron (muon) channel, followed by $E_T^{\text{miss}} + m_T > 60$ GeV, where $m_T$ is the transverse mass of the lepton and $E_T^{\text{miss}}$ [12]. The $E_T^{\text{miss}}$ is constructed from the vector sum of all calorimeter cells contained in topological clusters [13], calibrated at the energy scale of the associated high-$p_T$ object, and including contributions from selected muons. Further requirements are that there be at least three jets with $p_T > 25$ GeV and $|\eta| < 2.5$, with at least one jet satisfying $p_T > 60$ GeV. Jets are reconstructed with the anti-$k_t$ algorithm [14] with radius parameter $R = 0.4$, from topological clusters of energy deposits in the calorimeters calibrated at the electromagnetic scale. These jets are then calibrated to the particle level [15] by using a $p_T$- and $\eta$-dependent correction factor derived from simulated events and validated by using data. Finally, to further reduce the backgrounds, at least one jet is required to be identified as originating from the hadronization of a $b$ quark ($b$ tagging). This is achieved via an algorithm [16] using multivariate techniques to combine information from the impact parameters of displaced tracks as well as topological properties of secondary and tertiary decay vertices reconstructed within the jet; a working point is used with $\sim 70\%$ efficiency for $b$-quark jets and a rejection factor of $\sim 100$ for jets originating from light quarks ($u$, $d$, $s$) or gluons. Events with exactly one electron or one muon and with exactly three jets or with four or more jets are analyzed separately to take advantage of their different signal-to-background ratio and background composition, as discussed below.

After event selection, the main background is $t\bar{t}$ production, followed by the production of a W boson in association with jets ($W + $ jets). Smaller contributions arise from multijet events, single top quark, $Z + $ jets, and diboson production. All of the backgrounds which do not involve top quarks are significantly suppressed by the $b$-tagging requirement. Multijet events contribute to the selected sample via the misidentification of a jet or a photon as an electron or the presence of a nonprompt lepton, e.g., from a semileptonic $b$- or $c$-hadron decay. The normalization and shape of the multijet background kinematic distributions are estimated via data-driven methods [11]. For the $W + $ jets background, the shape is estimated from the simulation, but the normalization is estimated from the asymmetry between $W^+ + $ jets and $W^- + $ jets production [17] in the data. All other backgrounds, as well as the signal, are estimated from the simulation and normalized to their theoretical cross sections. A summary of the background estimates in each of the four channels analyzed and a comparison with the observed yields in data are presented in Table I, showing a good agreement within the uncertainties.

Monte Carlo (MC) samples of $t\bar{t}$ and single top quark background are generated by using MC@NLO v3.41 [18], assuming a top quark mass of 172.5 GeV, using the CT10 set of parton distribution functions (PDFs) [19], and are normalized to the approximate next-to-next-to-leading-order (NNLO) theoretical cross sections [20,21]. Samples of $W/Z + $ jets background are generated by using ALPGEN V2.13 [22] and the CTEQ6L1 PDF set [19]. The $Z + $ jets background is normalized to the NNLO theoretical cross section [23], while the $W + $ jets background normalization is extracted from the data. Both MC@NLO and ALPGEN are interfaced to HERWIG v6.5 [24] to model the parton shower and fragmentation, while JIMMY [25] is used to simulate the underlying event. The diboson backgrounds are modeled by using HERWIG v6.5 and normalized to their NLO theoretical cross sections [26]. The signal is modeled by using PYTHIA 6.421 [27]. Signal samples are generated for

| TABLE I. Number of events observed compared to the background expectation after final event selection in each of the four channels considered. Also shown are the expected signal yields assuming $m_t = 400$ GeV. The quoted uncertainties are prior to the fit to data and include both statistical and systematic contributions, taking into account correlations among processes. |
|-------------------|-------------------|-------------------|-------------------|
|                   | $e + 3$ jets       | $\mu + 3$ jets    | $e + \geq 4$ jets |
| $t\bar{t}$        | $2320 \pm 460$    | $3000 \pm 630$    | $4470 \pm 920$    |
| $W + $ jets       | $1440 \pm 790$    | $2200 \pm 1200$   | $830 \pm 580$     |
| $Z + $ jets       | $92 \pm 53$       | $118 \pm 62$      | $86 \pm 56$       |
| Single top        | $382 \pm 68$      | $554 \pm 94$      | $262 \pm 70$      |
| Dibosons          | $28 \pm 7$        | $37 \pm 11$       | $12 \pm 5$        |
| Multijet          | $520 \pm 520$     | $550 \pm 550$     | $320 \pm 320$     |
| Total prediction  | $4800 \pm 1000$   | $6500 \pm 1500$   | $6000 \pm 1100$   |
| Data              | $4533$            | $6421$            | $6145$            |
| $t\bar{t}P$(400 GeV)| $20.0 \pm 3.3$    | $21.0 \pm 3.6$    | $102.0 \pm 10.5$  |
|                   | $98.1 \pm 11.1$   |
a range of masses $m_\ell$ from 250 to 500 GeV in steps of 50 GeV and are normalized to the approximate NNLO theoretical cross sections [20] using the CTEQ6.6 PDF. The MC samples generated by using HERWIG or PYTHIA use the MRST2007 LO+ PDF set [28]. All MC samples include multiple $p\bar{p}$ interactions and are processed through a full simulation [29] of the detector geometry and response using GEANT4 [30] and the same reconstruction software as the data. Simulated events are corrected to match the object identification efficiencies and resolutions determined in data control samples. The total signal detection efficiency, considering both lepton flavors and determined in data control samples, ranges from 5.2% for $m_\ell = 250$ GeV to 17.3% for $m_\ell = 500$ GeV.

This analysis uses the reconstructed heavy quark mass ($m_{\text{reco}}$) as the primary discriminating variable. In the case of events with $\geq 4$ jets, $m_{\text{reco}}$ is estimated by performing a kinematic likelihood fit [17] to the $t\bar{t}\to W^+bW^-\bar{b}\to \ell\nu bQ\bar{Q}b$ hypothesis, imposing the constraints that $t'$ and $\bar{t}'$ have the same mass and that the mass of the lepton-neutrino system, as well as that of a jet pair, equals the nominal $W$ boson mass. The final-state objects considered are the lepton, $E_T^{\text{miss}}$, and the four jets with highest $p_T$. Among all possible jet-parton permutations, the one yielding the highest likelihood value after maximization over the fit parameters is kept. In the case of events with exactly three jets, $m_{\text{reco}}$ is taken to be the invariant mass of the three-jet system. In order to ensure a robust background prediction in the tail of the $m_{\text{reco}}$ distribution, a dynamic binning scheme is adopted; starting from the high side and low side of the distributions, bins are merged until the statistical uncertainty in the sum of the background predictions in that bin drops below 5%.

Systematic uncertainties affecting the normalization and shape of the $m_{\text{reco}}$ distribution are estimated for both signal and background, taking into account correlations among processes as well as channels. The dominant sources of uncertainty arise from the modeling of the $t\bar{t}$ background. The uncertainties on the $t\bar{t}$ background come from the theoretical uncertainty on the cross section ($+9\%/-6\%$) as well as the effects on both normalization and shape of the $m_{\text{reco}}$ distribution from a number of sources; these are uncertainties on the fragmentation model (based on the comparison of HERWIG and PYTHIA fragmentations), on the NLO event generator (based on the comparison of MC@NLO and POWHEG [31]), and on the top quark mass (taken to be $\pm 1$ GeV).

The uncertainty on the jet energy scale affects the normalization of signal (2%–12%) and backgrounds (5%–30%) modeled through the simulation, as well as the shape of their $m_{\text{reco}}$ distributions.

Uncertainties on the modeling of initial- and final-state QCD radiation, evaluated by varying corresponding generator parameters, are considered as correlated between the $t\bar{t}$ background and the $t\bar{t}^\prime$ signal.

While the normalization is obtained from the asymmetry measurement, the uncertainties on the normalization of the $W +$ jets background are derived from measurements of $W + 2$ jets dominated data samples and take into account the uncertainty on the heavy-flavor content of the samples as well as the extrapolation to higher jet multiplicities. The total uncertainty on the $W +$ jets normalization is 50% and 70% for events with exactly 3 jets and $\geq 4$ jets, respectively. Uncertainties on the shape of the $m_{\text{reco}}$ distribution for the $W +$ jets background are estimated by varying the choices of the matching scale (from 15 to 10 GeV) and the factorization scale (from $\mu_r^2 = m_W^2 + \sum p_{T,jet}^2$ to $\mu_r^2 = m_W^2 + p_{T,W}^2$) in ALPGEN.

Uncertainties on the modeling of the $b$-tagging algorithms affect the identification of $b/c$ jets (6%–8% for signal and backgrounds containing top quarks and 6%–12% for the other backgrounds) as well as the misidentification of light jets (<0.5% for signal and backgrounds containing top quarks and up to 5% for the other backgrounds). The $Z +$ jets, single top, and diboson backgrounds are varied within the uncertainty on their theoretical cross sections. The uncertainty on the multijet background event normalizations is conservatively taken as 100%. Uncertainties on the shapes of the multijet background are derived by varying the lepton identification criteria used to extract this background.

The uncertainties on the lepton identification and trigger efficiencies, as well as their energy scales and resolutions, impact the yields by 3% for electrons and 6% for muons.

Uncertainties on the integrated luminosity (3.7%) [32], jet reconstruction efficiency, jet resolution modeling, effect of multiple $p\bar{p}$ interactions on the modeling of the $E_T^{\text{miss}}$, and treatment of imperfections in the detector description in the MC simulation are also considered and are all found to have a very small effect on the result.

Good agreement between the data and the background prediction is observed in terms of both overall normalization and shape of the $m_{\text{reco}}$ distribution. The $m_{\text{reco}}$ distribution is analyzed by using a log-likelihood ratio $LLR = -2\log(L_{x+y}/L_{b})$ as a test statistic, where $L_{x+y}$ ($L_{b}$) is a Poisson likelihood to observe the data under the signal-plus-background (background-only) hypothesis. The per-bin signal and background predictions are parameterized in terms of 12 nuisance parameters, describing the effect of leading sources of systematic uncertainty such as jet energy scale, initial- and final-state QCD radiation, and $t\bar{t}$, $W +$ jets, and QCD multijet normalizations. The impact of systematic uncertainties on the sensitivity of the search is reduced by maximizing both likelihood functions, $L_{x+y}$ and $L_{b}$, with respect to these nuisance parameters, subject to Gaussian constraints of their prior values. The set of fitted nuisance parameters is chosen based on their overall impact on the search sensitivity, the expected constraining power of the data, and their suitability to be treated as continuous parameters. The simultaneous constraint of
several of these systematic uncertainties is possible because of the inclusion of the 3-jet channel in the analysis. The latter has a higher fraction of $W+$jets background than the $\geq 4$-jets channel and provides sensitivity to event migration to different jet multiplicities when varying uncertainties such as jet energy scale or initial- and final-state QCD radiation. In addition to the jet multiplicity spectrum, the jet energy scale affects the peak position of the $m_{\text{reco}}$ spectrum for $t\bar{t}$ background and can be constrained owing to the small uncertainty on the measured top quark mass [33]. Nuisance parameters associated with smaller systematic uncertainties (e.g., lepton identification or trigger) are only weakly constrained.

Figure 1 shows a comparison of the postfit $m_{\text{reco}}$ distribution between the data and the background prediction for the combined $e/\mu + 3$ jets and $e/\mu + \geq 4$ jets channels. The fitted parameters are typically within 1 standard deviation of their nominal values, and their uncertainties are consistent with expectations based on pseudoeXperiments. Several additional studies were performed to check the integrity of the fitting procedure. The likelihood was verified to be parabolic near the minimum for each of the fitted parameters and to yield reasonable fit uncertainties; the lack of sensitivity to the assumed $p_T$ and $\eta$ correlation of the jet energy scale uncertainty was verified.

In the absence of any significant data excess, either in the $e +$jets or $\mu +$jets channels individually or in their combination, 95% C.L. upper limits on the $t\bar{t}$ production cross section are derived by using the $CL_s$ method [34], which employs the $LLR$ test statistic described above. PseudoeXperiments are generated under both the signal-plus-background ($s+b$) and background-only ($b$) hypotheses, taking into account per-bin statistical fluctuations of the total predictions according to Poisson statistics, as well as Gaussian fluctuations in the signal and background expectations describing the effect of systematic uncertainties. The fraction of $s+b$ and $b$ pseudoeXperiments with $LLR$ larger than the median or observed $LLR$ defines $\text{CL}_{s+b}$ and $\text{CL}_b$ for the expected or observed limits, respectively. Signal cross sections for which $CL_s = CL_{s+b}/CL_b < 0.05$ are deemed excluded at the 95% C.L.

The resulting observed and expected upper limits on the $t\bar{t}$ production cross section are shown in Fig. 2 as a function of the $t'$ mass, compared to the theoretical prediction, assuming a $BR(t'\rightarrow Wb)=1$. As a result, an

![FIG. 1 (color online). $m_{\text{reco}}$ distribution in the combined (a) $e/\mu + 3$ jets and (b) $e/\mu + \geq 4$ jets channels. The data (points) are compared to the SM background predictions using the values of the nuisance parameters obtained from the fit to data under the background-only hypothesis (stacked histograms). In the top panels, the bin contents have been divided by bin width. The bottom panels show the background-subtracted data distribution. The underflow and overflow have been folded into the first and last bins, respectively. Also shown is the expected contribution from a signal with mass $m_t = 400$ GeV (histogram).](image1)

![FIG. 2 (color online). Observed (solid line) and expected (dashed line) 95% C.L. upper limits on the $t\bar{t}$ cross section as a function of the $t'$ mass. The surrounding shaded bands correspond to the 1 and 2 standard deviations (s.d.) around the expected limit. The thin line shows the theoretical prediction including its 1 s.d. uncertainty band. The shaded area is the mass region previously excluded by the CDF experiment [4].](image2)
observed (expected) 95% C.L. lower limit of 404 (394) GeV on the mass of the $t'$ quark is derived.

In summary, a search for $t't$ production has been performed in the lepton + jets final state under the assumption $BR(t' \rightarrow Wb) = 1$. No significant excess of events in the tail of the $m_{WW}$ distribution was found, resulting in an observed lower limit of $m_{t'} > 404$ GeV at 95% C.L. This represents the most stringent limit to date. This limit is also directly applicable to a down-type vectorlike quark with electric charge of $-4/3$ decaying into a W boson and a $b$ quark [3].

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently. We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC, and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST, and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR, and VSC CR, Czech Republic; DNRF, DNSRC, and Lundbeck Foundation, Denmark; ARTEMIS and ERC, European Union; IN2P3-CNRS, CEA-DSM/IRFU, France; INFN, Italy; MEXT and JSPS, Japan; NKP, Korea; LEFST ER, Latvia; MEPhI, Russia; MES of Russia and ROSATOM, Russian Federation; JINR; MSTD, Serbia; MSSR, Slovakia; ARRS and MVZT, Slovenia; DST/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SER, SNSF, and Cantons of Bern and Geneva, Switzerland; NSC, Taiwan; TAEK, Turkey; STFC, the Royal Society, and Leverhulme Trust, United Kingdom; DOE and NSF, United States of America. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular, from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (The Netherlands), PIC (Spain), ASGC (Taiwan), RAL (United Kingdom), and BNL (USA) and in the Tier-2 facilities worldwide.

[9] Pseudorapidity is defined as $\eta = -\ln \tan(\theta/2)$, where $\theta$ is the polar angle relative to the beam direction and $\phi$ is the azimuthal angle in the plane transverse to the beam direction.
[12] The transverse mass is defined by the formula $m_T = \sqrt{p_T^2 E_{T\text{miss}}^2 (1 - \cos \Delta \phi)}$, where $p_T$ is the $p_T$ of the lepton and $E_{T\text{miss}}$ is the azimuthal angle separation between the lepton and $E_{T\text{miss}}$.