Study of heavy-flavor quarks produced in association with top-quark pairs at $\sqrt{s} = 7$ TeV using the ATLAS detector


DOI
10.1103/PhysRevD.89.072012

Publication date
2014

Document Version
Final published version

Published in
Physical Review D. Particles, Fields, Gravitation, and Cosmology

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
I. INTRODUCTION

In order to characterize the recently observed Higgs-like particle (H) [1,2], quantities such as the Yukawa coupling of the top quark and the Higgs boson need to be measured with precision. For a Standard Model (SM) Higgs boson with a mass of 125 GeV, the decay mode with the largest branching ratio is $H \rightarrow b\bar{b}$. Thus, the channel with the largest yields for studying $t\bar{t} + H$ production is $t\bar{t} + H$, $H \rightarrow b\bar{b}$. Production of top-quark pair ($t\bar{t}$) events featuring additional heavy-flavor (HF) b and c quarks, $t\bar{t} + b + X$ and $t\bar{t} + c + X$, referred to as $t\bar{t} + $HF, is the main irreducible background to $t\bar{t} + H$, $H \rightarrow b\bar{b}$. A study of $t\bar{t} + $HF production is useful to constrain models of heavy-flavor quark production at the scale of the top-quark mass. This analysis is also of interest because of the many potential phenomena beyond the SM, such as composite Higgs models [3] and processes leading to final states with four top quarks [4–9], that could produce additional heavy-flavor quarks in the $t\bar{t}$ candidate sample.

This paper describes a study of $t\bar{t} + $HF production. Within the SM, heavy-flavor quark pairs, $c\bar{c}$ and $b\bar{b}$, are expected to be produced in association with $t\bar{t}$ mainly via gluon splitting from initial- and final-state radiation [10]. In addition, the heavy-flavor content of the proton could lead to $t\bar{t}$ final states with at least one additional HF quark, $t\bar{t} + c$ and $t\bar{t} + b$. The data analyzed correspond to an integrated luminosity of 4.7 fb$^{-1}$ at a center-of-mass energy of $\sqrt{s} = 7$ TeV produced at the Large Hadron Collider (LHC) and recorded in 2011 with the ATLAS detector.

This analysis is performed on $t\bar{t}$ dilepton candidate events in which each top quark decays to a b quark and a W boson, which subsequently decays to a neutrino and an isolated, charged lepton. The dilepton signature is selected for this measurement because it is relatively background free and precludes an additional b-tagged jet from a hadronically decaying W boson, predominantly via $W \rightarrow s\bar{c}$. The $t\bar{t} + $HF signal region is the subset of these events with three or more jets identified as containing HF quarks (b-tagged jets, or b tags). However, jets without HF quarks may also be b-tagged, so that care must be taken to properly identify the flavor composition of the b-tagged jets in the sample. Two b-tagged jets from each event are presumed to originate from the b quarks from top-quark decays, $t\bar{t} \rightarrow W^+bW^-\bar{b}$. Therefore, all events in the signal region have at least one additional b tag either from a b- or c-quark jet, or from a light-quark or gluon jet that was misidentified. The latter two are referred to as light-flavor or LF jets.

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 articles title, journal citation, and DOI.
Because of limited data statistics and discrimination between $b$ and $c$ jets, the sum of $b$-quark and $c$-quark jet rates is measured. Information about the composition of $t\bar{t} + b + X$ and $t\bar{t} + c + X$ in $t\bar{t} + \text{HF}$ is nevertheless required for the total correction due to acceptance, which is different for $b$- and $c$-quark jets. The composition is estimated with Monte Carlo simulation.

From the measurement of the fraction of jets with heavy-flavor content, the cross section for $t\bar{t}$ production with at least one additional HF jet can be extracted. To reduce some systematic uncertainties, the result is quoted as a ratio, termed $R_{\text{HF}}$, of the cross section for $t\bar{t}$ production with at least one additional HF jet to the cross section for $t\bar{t}$ production with at least one additional jet ($t\bar{t} + j$), regardless of flavor. The measurement of $t\bar{t} + j$ production is performed in dilepton $t\bar{t}$ candidate events with at least three jets, at least two of which are $b$-tagged and assumed to come from top-quark decays.

The paper is organized as follows. The ATLAS detector is briefly described in Sec. II. The data and Monte Carlo samples used in the analysis are described in Sec. III, followed by a description of the event selection in Sec. IV. The definition of the fiducial phase space used in the measurement of $R_{\text{HF}}$, and the calculation of acceptances and efficiencies are presented in Sec. V. In Sec. VI, observed and expected numbers of events with $\geq 3$ $b$-tagged jets are shown. Section VII describes a two-dimensional fit to the vertex mass and jet transverse momentum distributions of $b$-tagged jets in these data events to extract the fraction of HF jets produced in association with $t\bar{t}$. A discussion of the systematic uncertainties of the measurement is presented in Sec. VIII. Section IX shows the result of the measurements, followed by conclusions in Sec. X.

II. THE ATLAS DETECTOR

A detailed description of the ATLAS detector can be found elsewhere [12]. The innermost part of the detector is a tracking system that is immersed in a 2 T axial magnetic field and that measures the momentum of charged particles. The inner detector comprises a silicon pixel detector, a silicon microstrip detector, and a transition radiation tracker, providing tracking capability within the pseudorapidity range $|\eta| < 2.5$ [11]. The tracking system is also used to identify the displaced secondary vertex that is formed by hadrons containing a $b$ or $c$ quark. Calorimeter systems, which measure the electron, photon, and hadron energies, reside outside the inner detector and cover the region $|\eta| < 4.9$. Outside the calorimeters there is a muon spectrometer that is used to identify and measure the momentum of muons in an azimuthal magnetic field in the region $|\eta| < 2.7$. To reduce the data rate, a three-level trigger system selects the potentially interesting events that are recorded for off-line analysis.

III. DATA AND MONTE CARLO SAMPLES

The total integrated luminosity for the analyzed data sample is 4.7 fb$^{-1}$ at a center-of-mass energy of $\sqrt{s} = 7$ TeV. During the 2011 data-taking period the instantaneous luminosity of the LHC increased, causing the average number of simultaneous inelastic $pp$ interactions per beam crossing (pileup) at the beginning of a $pp$ fill to increase from about 6 to 17. Multiple $pp$ interactions can occur either in the same bunch crossing as the primary vertex (termed “in-time pileup”) or in an adjacent bunch crossing (termed “out-of-time pileup”). To account for these effects, all Monte Carlo simulated events are overlaid with additional inelastic events generated with the PYTHIA AMBT1 tune [13], and the distribution of the number of vertices in the simulation is reweighted to match the distribution of the number of additional interactions per bunch crossing measured in the data.

Monte Carlo simulation is used to study signal and background processes. Inclusive $t\bar{t}$ production and dedicated $t\bar{t} + \text{HF}$ samples are simulated using the multileg matrix-element generator ALPGEN v2.13 [14] with the CTEQ6L1 [15] parton distribution function (PDF) set. Parton showering and hadronization are performed by HERWIG v6.520 [16]. Effects due to the mass of the heavy-flavor quarks are included by default in ALPGEN. In these samples, additional jets (including heavy-flavor) can also be produced in the parton shower. The parton-jet matching scheme described in [14] is applied to avoid double counting of configurations generated by both the parton shower and the leading-order (LO) matrix-element (ME) calculation. In addition, overlap between $t\bar{t}$ events with HF quarks that originate from ME production and those that originate from the parton shower is removed. This heavy-flavor overlap removal is based on the $\Delta R_{qq}$ [17] between simulated HF quarks. Events generated by the ME calculation containing two HF quarks separated by $\Delta R_{qq} < 0.4$ are vetoed, as are events containing two HF quarks from the parton shower with $\Delta R_{qq} > 0.4$.

To study the effect of different fixed-order calculations and matching schemes, samples of top-quark pair events are also generated using POWHEG v1.01 [18] and showered with HERWIG. In this sample the $t\bar{t}$ process is described at next-to-leading order (NLO), while the extra jets are described at LO. For each sample showered with HERWIG, JIMMY v4.31 [19] and the AUET1 tune [20] are used to simulate the underlying event and to model various soft interactions. To assess the effect of different parton shower models, a sample is generated using ALPGEN v2.14 with the PYTHIA v6.425 [21] parton shower and hadronization, using the CTEQ5L PDF set [22]. The uncertainty associated with the CTEQ6L1 PDF set is evaluated with an envelope calculated using the uncertainty set from the NLO PDF MSTW2008nlo68cl [23], and an additional term to account for the difference between the central values of the LO and NLO calculations.
STUDY OF HEAVY-FLAVOR QUARKS PRODUCED IN ...

Initial- and final-state radiation (ISR/FSR) variations are studied using samples generated with AcerMC v2.0 [24] interfaced with Pythia v6.2. In these samples the parameters that control the amount of ISR/FSR are set to points consistent with the PERUGIA Hard/Soft tune [25] in a range constrained by current experimental data [26].

In all samples the top-quark mass is set to $m_t = 172.5$ GeV. The cross section for Standard Model $t\bar{t}$ production at this mass is calculated using the approximate next-to-next-to-leading-order QCD calculation described in [27].

Background samples from the production of $W$ and $Z$ bosons are generated using the CTEQ6L1 PDFs with Alpgen, which is interfaced to Herwig for parton showering and hadronization; the Alpgen matrix elements include diagrams with up to five additional partons. Separate samples of $W + b\bar{b}$ and $Z + b\bar{b}$ events are generated. The overlap between jets from the parton shower and the matrix element in the $n + 1$ jet multiplicity samples is removed for the $W +$ jets and $Z +$ jets samples in the same manner as for the $t\bar{t}$ samples. Single top-quark production is modeled using AcerMC in the $t$ channel and MC@NLO v3.41 [28] for the $Wt$ and $s$ channels. Diboson ($WW, WZ,$ and $ZZ$) production is modeled using Alpgen interfaced with Herwig. Less than 0.5% of the expected yield in the $t\bar{t} +$ HF sample comes from the associated production of $t\bar{t} + W/Z$ and $t\bar{t} + H$, and these processes are thus neglected in this analysis.

The resulting generated samples are passed through a Geant4 simulation [29] of the ATLAS detector [30]. Events are then reconstructed in the same manner as the data.

IV. EVENT SELECTION

Events for the analysis are selected by at least one of the high-$p_T$ [11] single-electron or single-muon triggers, as described in Refs. [31,32]. The single-electron triggers are based on calorimeter energy deposits, shower shape, and matching track quality constraints, while the single-muon triggers are based on a reconstructed track in the muon spectrometer that matches a track found in the inner detector. To ensure a final trigger rate that is compatible with the ATLAS data acquisition system, a minimum $p_T$ threshold for the electron and muon triggers is used. The $p_T$ threshold for the muon trigger is 18 GeV. For the electron trigger, the threshold is 20 or 22 GeV, depending on the data-taking period due to varying LHC luminosity conditions.

The selected events are required to contain a reconstructed primary vertex with at least five associated tracks with $p_T > 0.4$ GeV. Event reconstruction makes use of electrons ($e$), muons ($\mu$), jets, and missing transverse momentum ($E_T$). Electrons are reconstructed by matching energy deposits in the electromagnetic calorimeter with tracks in the inner detector and are required to have $p_T > 25$ GeV and $|\eta| < 2.47$, excluding the transition region between the barrel and end cap calorimeters at $1.37 < |\eta| < 1.52$ [33]. Muons are reconstructed by matching tracks in the inner detector with tracks measured in the muon spectrometer and are required to have $p_T > 20$ GeV and $|\eta| < 2.5$.

Tight isolation cuts are applied to both the electron and muon candidates to reduce the number of identified leptons ($e, \mu$) that come from nonprompt (non-$W/Z$) sources and from misidentified hadrons. For electrons, the $E_T$ deposited in the calorimeter cells in a cone in $\eta\phi$ space of radius $\Delta R = 0.2$ around the electron position is summed, and the $E_T$ due to the electron is subtracted. The scalar sum of track transverse momenta in a cone of $\Delta R = 0.3$, excluding the electron, is also measured. Cuts parametrized by the electron $\eta$ and $E_T$ are made on these two isolation variables to ensure a constant efficiency over the entire ($\eta, E_T$) range. For muons, the corresponding calorimeter isolation energy in a cone of $\Delta R = 0.2$ is required to be less than 4 GeV, and the scalar sum of track transverse momenta in a cone of $\Delta R = 0.3$ is required to be less than 2.5 GeV after subtraction of the muon $p_T$.

Jets are reconstructed from clustered energy deposits in the calorimeters with the anti-$k_t$ [34] algorithm with a radius parameter $R = 0.4$ [35]. Jets selected for the analysis are required to have $p_T > 25$ GeV and $|\eta| < 2.5$. In order to reduce the background from jets originating from pileup interactions, the jet vertex fraction, defined as the sum of the $p_T$ of tracks associated to the jet and originating from the primary vertex divided by the sum of the $p_T$ from all tracks associated with the jet, is required to be greater than 0.75.

The transverse momentum of neutrons produced in the top-quark decays, measured as $E_T$, is inferred by balancing the vector sum of all visible transverse momenta. Specifically, the $E_T$ is constructed from the vector sum of all calorimeter cell energies contained in topological clusters [35] with $|\eta| < 4.9$, projected onto the transverse plane. Contributions to the $E_T$ from the calorimeter cells associated with jets are taken at the corrected energy scale that is used for jets, while the contribution from cells associated with electrons is substituted by the calibrated transverse momentum of the electron. The contribution to the $E_T$ from the $p_T$ of muons passing the selection requirements is also included.

The $b$-tagging algorithm [36,37] employed for this analysis uses track impact parameters (the distance of closest approach of each track to the primary vertex) and the properties of any reconstructed secondary vertex associated to the jet as inputs to a neural network. The $b$-tagging efficiency was calibrated in a multijet data sample where at least one jet contains a muon [37]. The $c$-tagging efficiency was calibrated in a data sample with reconstructed $D^*$ mesons [38]. For this analysis, $b$-tagged jets are required to satisfy a selection that is 75% efficient for $b$-quark jets, approximately 30% efficient for $c$-quark jets,
and rejects light-flavor jets by a factor of approximately 35 in simulated $\bar{t}t$ events. In this paper, a $b$ tag (or a $b$-tagged jet) refers to any jet passing this selection, regardless of flavor. A $b$ jet, by contrast, refers to a jet (which may or may not be $b$-tagged) which contains a $b$ quark. Similarly, $c$ jet and HF jet are statements of the flavor composition of the jet, not whether the jet is $b$-tagged. Three distinct subsets of the selected $b$-tagged jets with different $b$-jet purity are used in the measurement of $\sigma_{\mathrm{fid}}(\bar{t}t + \text{HF})$, as described in Sec. VII.

Dilepton $\bar{t}t$ candidate events are selected by requiring exactly two opposite-sign leptons and at least two jets. To reduce the background from $Z/\gamma^{*}$ processes, events with like-flavor leptons are required to have $E_T$ above 60 GeV and a dilepton invariant mass satisfying $|m_{\ell^{+}\ell^{-}} - m_Z| > 10$ GeV. For events with one electron and one muon, the scalar sum of the lepton and jet transverse momenta is required to be above 130 GeV to reduce the backgrounds from $Z/\gamma^{*} \to \tau^{+}\tau^{-}$, as well as $WW$, $WZ$, and $ZZ$ processes. This set of selection criteria is termed the “nominal” $\bar{t}t$ selection criteria. The measurement of $\bar{t}t + \text{HF}$ production is carried out in the subset of these events that contain three or more $b$-tagged jets, whereas the measurement of $\bar{t}t$ production with at least one additional jet is performed in the subset with at least three jets, at least two of which are $b$-tagged.

Using the nominal selection criteria described above, data and Monte Carlo events are compared in three control regions: dilepton $\bar{t}t$ candidate events with zero, one, or two $b$-tagged jets. Data-to-simulation normalization corrections are applied to Monte Carlo simulation samples when calculating acceptances to account for observed differences in predicted and observed trigger and lepton reconstruction efficiencies, jet flavor tagging efficiencies and mistag rates, as well as jet and lepton energy scales and resolutions. In Fig. 1, the jet multiplicity distributions in the three regions are compared to Monte Carlo predictions. Agreement is observed within uncertainties.

V. DEFINITION OF THE FIDUCIAL PHASE SPACE AND CALCULATION OF CORRECTION FACTORS

To allow comparison of the analysis results to theoretical predictions, the measurement is made within a fiducial phase space. The fiducial volume is defined in Monte Carlo simulation by requiring two leptons ($e, \mu$) from the $t \to Wb$ decays (including electrons and muons coming from $t \to \ell\nu\tau$) with $p_T > 25(20)$ GeV for $e(\mu)$, and $|\eta| < 2.5$ as well as three or more jets with $p_T > 25$ GeV and $|\eta| < 2.5$.

In the simulation, jets are formed by considering all particles with a lifetime longer than 10 ps, excluding muons and neutrinos. Particles arising from pileup interactions are not considered. For the determination of the $\bar{t}t + \text{HF}$ fiducial cross section, $\sigma_{\mathrm{fid}}(\bar{t}t + \text{HF})$, three or more jets are required to match a $b$ or $c$ quark, two of which must match a $b$ quark from top-quark decay. All simulated $b$ and $c$ quarks that were generated with $p_T > 5$ GeV are considered for the matching and are required to satisfy $\Delta R(\text{quark, jet}) < 0.25$. Jets that match both a $b$ or a $c$ quark are considered as $b$ jets. For the calculation of $\sigma_{\mathrm{fid}}(\bar{t}t + j)$ three or more jets are required, two of which must contain a $b$ quark from top-quark decay.

Each fiducial cross section is determined using measured quantities from the data, and a correction factor derived from the Monte Carlo simulation. The ratio of cross sections is defined as
The fiducial cross section for $\bar{t}t +$ HF production is determined from

$$\sigma_{\text{fid}}(\bar{t}t + \text{HF}) = \frac{N_{\text{HF}}}{\int \mathcal{L} \cdot \epsilon_{\text{HF}}},$$

(1)

where $N_{\text{HF}}$ is the number, after background subtraction, of $b$ tags from HF jets observed in the data, in addition to the two $b$ jets from top-quark decays. The integrated luminosity of the sample is denoted as $\int \mathcal{L} dt$, and $\epsilon_{\text{HF}}$ is a correction factor taken from Monte Carlo simulation that converts the number of observed $b$ tags from additional HF jets to the number of events in the signal fiducial volume. This correction factor includes the acceptance within the fiducial region, the reconstruction efficiency, and a factor to account for the multiplicity of extra $b$-tagged HF jets per $\bar{t}t +$ HF event in the signal region. This correction factor is different for $\bar{t}t + b + X (\epsilon_b)$ and $\bar{t}t + c + X (\epsilon_c)$, and thus $\epsilon_{\text{HF}}$ is determined as a weighted sum of these two contributions. The weight used to form the sum is the fraction of $\bar{t}t +$ HF events in the fiducial volume which contain additional $b$ jets as opposed to $c$ jets. This fraction is termed $F_{b/\text{HF}}$. The total correction factor ($\epsilon_{\text{HF}}$) is calculated as

$$\epsilon_{\text{HF}} = F_{b/\text{HF}} \cdot \epsilon_b + (1 - F_{b/\text{HF}}) \cdot \epsilon_c.$$

The denominator for $R_{\text{HF}}$, $\sigma_{\text{fid}}(\bar{t}t + j)$, is computed using a similar prescription:

$$\sigma_{\text{fid}}(\bar{t}t + j) = \frac{N_j}{\int \mathcal{L} dt \cdot \epsilon_j},$$

(2)

where $N_j$ is the yield of dilepton events in data with at least three jets, at least two of which are $b$-tagged, and $\epsilon_j$ is the $\bar{t}t + j$ acceptance factor calculated from the Monte Carlo simulation. The acceptance calculation for each fiducial cross section assumes that all $b$-tagged jets are from real HF quarks. Events with $b$-tagged jets from LF quarks are treated as a background and subtracted when computing both $N_{\text{HF}}$ and $N_j$.

The ALPGEN+HERWIG Monte Carlo sample predicts $\epsilon_b = 0.19$, $\epsilon_c = 0.06$, and $F_{b/\text{HF}} = 0.31$. The total correction factor is thus predicted to be $\epsilon_{\text{HF}} = 0.106 \pm 0.005$ (stat) for $\sigma_{\text{fid}}(\bar{t}t + \text{HF})$. For $\sigma_{\text{fid}}(\bar{t}t + j)$ the acceptance factor is calculated to be $\epsilon_j = 0.129 \pm 0.001$ (stat).

The prediction for $R_{\text{HF}}$ from the ALPGEN+HERWIG Monte Carlo sample is 3.4%. The value obtained from the POWHEG v1.01 [18] generator showered with HERWIG [16] is $R_{\text{HF}} = 5.2\%$, with $F_{b/\text{HF}} = 0.34$. While this $R_{\text{HF}}$ value is different to that from ALPGEN+HERWIG, the predicted $F_{b/\text{HF}}$ values are similar. Furthermore, a parton-level study using MADGRAPH5 v1.47 [39] gives $F_{b/\text{HF}} = 0.29$. The value of $F_{b/\text{HF}}$ is also stable when different showering algorithms are used: the ALPGEN+PYTHIA Monte Carlo sample predicts a value of $F_{b/\text{HF}} = 0.32$, in good agreement with the prediction when HERWIG is used. Based on comparison of these predictions for $F_{b/\text{HF}}$, a symmetric 10% Monte Carlo systematic uncertainty is assigned, $F_{b/\text{HF}} = 0.31 \pm 0.03$.

### VI. Expected Signal and Background Yields

Table I shows the number of events with $\geq 3$ $b$-tagged jets expected in the Monte Carlo simulation from dilepton $\bar{t}t$ production and from various background sources. At this point, no distinction is made between events with a true additional HF jet and those containing a mistagged LF jet. The number of observed events is also shown. While Monte Carlo simulation is used to estimate $\bar{t}t +$ HF event rates and kinematic features, data-driven methods and Monte Carlo simulation are both used to estimate background processes, as detailed below.

Background processes containing real $b$ jets and leptons, such as single top-quark, $Z/\gamma^* +$ jets, and diboson ($WW$, $WZ$, and $ZZ$) production, are estimated using Monte Carlo simulation. Contributions from diboson production are found to be negligible.

A major source of background comes from $\bar{t}t$ events in which one or more of the $b$-tagged jets is from a mistagged LF jet. This background is estimated using Monte Carlo simulation for the measurement of $\sigma_{\text{fid}}(\bar{t}t + \text{HF})$. However, in the measurement of $\sigma_{\text{fid}}(\bar{t}t + \text{HF})$, the final $\bar{t}t +$ LF background is determined by a fit to the vertex mass distribution of $b$-tagged jets in data, as explained in Sec. VII.

Background from events in which at least one of the leptons is either nonprompt (originating from e.g. a photon conversion or $b$-quark decay) or is a misidentified hadron is estimated using data and Monte Carlo simulation. For instance, $W +$ jets, multijet, and $\bar{t}t$ events with one hadronically decaying $W$ boson can contribute in this way. This contribution is determined by scaling the yield of events in the data with a pair of same-sign leptons by the ratio of opposite-sign to same-sign yields ($R_{\text{OS/SS}}$) obtained in Monte Carlo simulation. The opposite-sign to same-sign

<table>
<thead>
<tr>
<th>Process</th>
<th>Number of events</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{t}t$</td>
<td>106.7 ± 3.4</td>
</tr>
<tr>
<td>Single top</td>
<td>2.2 ± 0.5</td>
</tr>
<tr>
<td>$Z +$ jets</td>
<td>0.2 ± 0.1</td>
</tr>
<tr>
<td>Fake leptons</td>
<td>0−05</td>
</tr>
<tr>
<td>Total expectation</td>
<td>109±29 (stat) ± 35 (syst)</td>
</tr>
<tr>
<td>Data</td>
<td>106</td>
</tr>
</tbody>
</table>
ratio is determined separately for the three dilepton channels and found to be $1.3 \pm 0.1 \text{(stat)} \pm 0.2 \text{(syst)}$ for $e^+e^-$ events, $1.2 \pm 0.1 \text{(stat)} \pm 0.7 \text{(syst)}$ for $\mu^+\mu^-$ events, and $1.2 \pm 0.1 \text{(stat)} \pm 0.5 \text{(syst)}$ for events with one electron and one muon. The systematic uncertainty takes into account the unknown relative mixture of fake-lepton sources (photon conversions, $b$- and $c$-hadron decays, or misidentified hadrons) in the $R_{\text{OS/SS}}$ calculation. As there are no same-sign events observed in the data, the central value of the prediction for this background is zero events. Therefore only upward fluctuations of this data count, and positive variations of $R_{\text{OS/SS}}$, are considered in the uncertainty calculation. This method for estimating the background due to events with fake leptons is validated in a control sample of dilepton events with less restrictive lepton identification requirements and no isolation criteria.

The dominant uncertainties on the total yield in Table I come from the jet energy scale, $b$-tagging efficiency, parton showering model, and initial- and final-state radiation.

### VII. TEMPLATE FIT

For the measurement of $\sigma_{\text{id}}(t\bar{t}+\text{HF})$, the fraction of heavy-flavor jets produced in association with $t\bar{t}$ is extracted by performing a binned maximum-likelihood fit on the displaced-vertex mass distribution using all $b$-tagged jets in the events with $\geq 3$ $b$-tagged jets. Although the final result is for both flavors combined, the fit includes separate $b$- and $c$-quark components to improve the determination of the LF fraction. This displaced-vertex mass is constructed from the inner detector tracks associated with the secondary vertex using the multivariate algorithm described in Ref. [40]. While the presence of a displaced vertex is an indication that a jet contains a $b$ quark, a jet can be $b$-tagged even without a secondary vertex by impact parameter information alone. In this case, the vertex mass is undefined. These jets are assigned a mass value of $-1$ GeV and they are included in the fit to the displaced-vertex mass distribution. Keeping the events without a reconstructed vertex improves the discrimination between heavy-flavor and light-flavor jets.

While the vertex mass is a powerful discriminant, Monte Carlo studies indicate that the sensitivity to the fitted fraction of LF jets increases when the jet $p_T$ is used as an additional discriminant. Considering only the statistical uncertainty, it is seen that a fit with both jet $p_T$ and vertex mass is approximately half a standard deviation more sensitive than a fit with only the vertex mass. It was thus decided to define a two-dimensional probability density function, termed a “template,” for the fit using the vertex mass and jet $p_T$.

The fit is performed simultaneously in three mutually exclusive bins of $b$-jet purity, defined by different ranges of the $b$-tagging neural network output value. Certain values of the neural network output, termed “operating points,” are defined by the average $b$-jet selection efficiency resulting from the applied selection. In this analysis, operating points of 60%, 70% and 75% efficiency are used to define the boundaries of the $b$-jet purity bins.

The first bin uses only the tightest calibrated operating point (60%), contains the highest-purity sample of $b$ jets (referred to as “high purity”), and has a $b$-tagging efficiency of 60% for $b$ jets. The second bin (referred to as “medium purity”) requires a $b$-tag selection between the tightest and second tightest (70%) operating points and contains a larger fraction of LF jets and $c$ jets. The efficiency for this bin is 10% for $b$ jets, i.e. the difference between the 70% and 60% operating points. The final bin ("low purity") requires a $b$-tag selection between the second (70%) and third operating points (75%) and contains the largest fraction of LF jets. The efficiency for this bin is 5% for $b$ jets. The $b$-tagging efficiencies for $b$ jets, $c$ jets, and light-flavor jets for each selection are given in Table II.

All three classes of $b$-tag purity are used in the analysis so that a jet is considered $b$-tagged if it satisfies any of these criteria. The discrimination power between LF and $c$ jets is greatly improved by using three (as opposed to one) classes of $b$ purity. The vertex mass distributions for all $b$-tagged jets in events passing the nominal $t\bar{t}$ selection criteria are shown in Fig. 2 to confirm that (a) the data are well described by the Monte Carlo simulation (b) and the $b$-jet, $c$-jet and LF-jet fractions differ in the three purity selections. For the purpose of illustration, the normalization of the $b$-jet, $c$-jet, and LF-jet components is taken from Monte Carlo simulation.

The template fit has five components: $b$ jets from top-quark decays, non-$t\bar{t}$ background, extra $b$-tagged jets from $b$ quarks, extra $b$-tagged jets from $c$ quarks, and light-flavor $b$-tagged jets. The template for $b$ jets from top-quark decays is obtained from the data in $t\bar{t}$ dilepton events with exactly two $b$ tags. Monte Carlo simulation indicates that $97\%$ of $b$-tagged jets in $t\bar{t}$ dilepton events with exactly two $b$ tags come from the decay of the top quark. To account for this in the shape of the data template, a template for $b$ tags not from the top-quark decays is derived from the $t\bar{t}$ Monte Carlo simulation and subtracted with a $3\%$ relative normalization from the data template. In the fit, the normalization for the template for $b$ jets from the top-quark decays is fixed assuming it contributes two of the three or more $b$ tags per observed event.

### TABLE II. Summary of the $b$-tagging efficiencies for $b$ jets, $c$ jets, and light-flavor jets for the three mutually exclusive $b$-tagging selections used in the vertex mass template fit.

<table>
<thead>
<tr>
<th>$b$ purity</th>
<th>$b$-jet efficiency</th>
<th>$c$-jet efficiency</th>
<th>Light-flavor efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>60%</td>
<td>17%</td>
<td>0.43%</td>
</tr>
<tr>
<td>Medium</td>
<td>10%</td>
<td>7%</td>
<td>1.00%</td>
</tr>
<tr>
<td>Low</td>
<td>5%</td>
<td>6%</td>
<td>1.33%</td>
</tr>
<tr>
<td>Total</td>
<td>75%</td>
<td>30%</td>
<td>2.76%</td>
</tr>
</tbody>
</table>
Background events from nondilepton $t\bar{t}$ processes are included using Monte Carlo simulation and enter the fit with a fixed normalization. Monte Carlo simulation is used to obtain templates for additional (non-$t \rightarrow Wb$) b jets, c jets, and LF jets.

In the fit to determine the number of b tags from HF jets in addition to the two b jets from top-quark decay, $N_{\text{HF}}$, separate templates for each category of jet in each of the three purity classes (high, medium, and low) are used. The $b$-tagging efficiencies (Table II) for each flavor of jet are used to relate the number of jets in each purity bin. After the application of all constraints, the fit has two floating parameters: the fraction of LF jets and the fraction of additional $b$ jets. The fraction of additional $c$ jets makes up the remainder.

Monte Carlo pseudoexperiments show that the fitting method is unbiased in both best-fit values and estimated uncertainties. The fit strategy (including estimates of statistical and systematic uncertainties) was verified using 10% of the full data sample as well as with Monte Carlo pseudoexperiments before the fit was performed on the full data sample. These studies indicated that the fit could achieve only a 1σ separation of $b$ vs c jets based on the expected statistical uncertainty alone. Inclusion of the systematic uncertainty would further reduce the sensitivity. However, the LF-jet fraction is expected to be measured with sufficient precision to give a statistically significant measurement of the total HF content, defined as the fraction of additional $b$-tagged jets not coming from LF jets. In the fit, the individual fractions are not constrained to be positive or below unity.

VIII. SYSTEMATIC UNCERTAINITIES

Systematic uncertainties may affect the shape of the vertex mass and $p_T$ templates as well as the acceptance calculations. For the systematic uncertainties on the template shapes, the fit to the data is reevaluated using new templates, derived by varying the relevant parameters by their systematic uncertainties, and a new fit to the data is performed. Major uncertainties that affect the fit are the jet
energy scale and resolution, the tagging efficiencies for $b$, $c$ and LF jets, the parton-shower and hadronization models, and the Monte Carlo event generators.

The template for $b$ jets from top-quark decays is nominally taken from the data with exactly two $b$ tags. To account for kinematic biases due to additional heavy-flavor jets in the event, a systematic uncertainty on the shape of this template is assessed using $b$ jets from top-quark decays from Monte Carlo inclusive $t\bar{t}$ events with three or more $b$-tagged jets.

The vertex mass of additional $b$ and $c$ jets is sensitive to the number of HF quarks contained in a jet (for instance, for $b\bar{b}$ or $c\bar{c}$ produced via gluon splitting). The dominant uncertainty from this effect would manifest itself as a difference in the shape of the template for additional $b$ jets. To assess this uncertainty, the template for additional $b$ jets is replaced by the template for $b$ jets from top-quark decays.

By default, the normalization of the template for $b$ jets from top-quark decays is fixed to two per event. A systematic uncertainty on this normalization is assessed by using the predicted normalization from Monte Carlo simulation, which includes events with less than two $b$ tags from top-quark decays, due to $b$-tagging inefficiency. The total uncertainty due to specific template shape variations is referred to as “additional fit uncertainties” for the rest of this paper.

Systematic uncertainties also affect the overall event reconstruction efficiency. Dominant sources of uncertainty for this category are the tagging efficiencies for $b$, $c$ and LF jets, the jet energy scale and resolution, and the Monte Carlo event generator. Uncertainties on the lepton identification efficiency, $E_T$ reconstruction, and fragmentation modeling are negligible. In general, systematic uncertainties are evaluated on the full data sample, with each uncertainty being taken as the difference between the nominal and the varied resulting values of $R_{HF}$.

An important uncertainty in this analysis comes from the flavor composition in the fiducial volume, namely in the value of $F_{b/\text{HF}}$, the fraction of $t\bar{t}+\text{HF}$ events in the fiducial volume which contain $b$ jets, used to calculate the correction factor $\epsilon_{\text{HF}}$. As described in Sec. VII, an uncertainty of 10% on $F_{b/\text{HF}}$ is estimated using different Monte Carlo generators. While the template fit method described in Sec. VII in principle evaluate $F_{b/\text{HF}}$ using the data, the present data sample does not allow a robust determination of this quantity.

IX. RESULTS

In the 106 events in the signal sample (with $\geq 3$ $b$-tagged jets), there are 325 $b$-tagged jets. After subtracting the non-$t\bar{t}$ background component, and the contribution from the tagged jets from the $t \rightarrow Wb$ decay, the number of additional $b$ tags is found to be 105. As described in Sec. VII, a template fit to all $b$-tagged jets is performed to determine the flavor composition of these additional $b$-tagged jets.

The result of the fit to all 325 $b$-tagged jets is shown in Fig. 3. The weighted sums of all fit templates are shown, with contributions for extra HF and mistagged LF jets shown separately. The fitted fractions of $b$ tags from LF jets and $b$ jets are given in Table III. Using these fractions,
TABLE IV. Summary of the relative systematic uncertainties (in percent) on the measurement of the ratio of fiducial cross sections, $R_{HF}$. Uncertainties are quoted separately for the number of HF jets measured in the fit ($N_{HF}$), the portion of the calculation affecting only the correction factors ($\epsilon_{HF}$), and the full calculation. As the fit prefers 100% charm for additional heavy-flavor jets, it is sensitive to differences in the extra $b$-tagged jets from the $c$-quark template shape.

<table>
<thead>
<tr>
<th>Source</th>
<th>% ($N_{HF}$)</th>
<th>% ($\epsilon_{HF}$)</th>
<th>% (full)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lepton reconstruction</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Jet reconstruction and calibration</td>
<td>3.5</td>
<td>1.6</td>
<td>6.9</td>
</tr>
<tr>
<td>$E_T$ reconstruction</td>
<td>0.5</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Fake-lepton estimate</td>
<td>3.4</td>
<td>0.0</td>
<td>3.4</td>
</tr>
<tr>
<td>Tagging efficiency for $b$ jets</td>
<td>1.1</td>
<td>2.4</td>
<td>3.1</td>
</tr>
<tr>
<td>Tagging efficiency for $c$ jets</td>
<td>25.0</td>
<td>5.9</td>
<td>21.2</td>
</tr>
<tr>
<td>Tagging efficiency for light jets</td>
<td>8.4</td>
<td>0.2</td>
<td>8.4</td>
</tr>
<tr>
<td>Fragmentation modeling</td>
<td>6.5</td>
<td>15.7</td>
<td>10.2</td>
</tr>
<tr>
<td>Generator variation</td>
<td>0.7</td>
<td>1.0</td>
<td>1.8</td>
</tr>
<tr>
<td>Initial- and final-state radiation</td>
<td>0.1</td>
<td>1.7</td>
<td>1.9</td>
</tr>
<tr>
<td>PDF uncertainties</td>
<td>1.6</td>
<td>1.0</td>
<td>2.8</td>
</tr>
<tr>
<td>Additional fit uncertainties</td>
<td>6.6</td>
<td>$\cdots$</td>
<td>6.6</td>
</tr>
<tr>
<td>Flavor composition</td>
<td>0.0</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Total systematic</td>
<td>29.0</td>
<td>18.0</td>
<td>28.0</td>
</tr>
</tbody>
</table>

$79 \pm 14^{(\text{stat})} \pm 22^{(\text{syst})}$ of the 105 additional $b$ tags are attributed to HF jets. A detailed breakdown of the systematic uncertainties on the total number of HF jets is shown in Table IV.

The uncertainty on the fitted fraction of light-flavor jets is significantly smaller than the uncertainty on the fitted fraction of additional $b$ jets. This is understood as an effect of fitting in multiple $b$-purity bins: the low-purity bin is dominated by light-flavor jets and thus gives improved discrimination. The data resolve the total observed HF production rate with a significance of about 3$\sigma$.

Using Eq. (1), the number of HF jets observed in data, and the quoted correction factor $\epsilon_{HF}$ derived from the Monte Carlo simulation for $\bar{t}t +$ HF production as discussed in Sec. V, $\sigma_{\text{fid}}(\bar{t}t + \text{HF})$ is found to be $0.16 \pm 0.03^{(\text{stat})}$ pb. This calculation assumes the value and uncertainty of $F_{b/\text{HF}}$ as determined from Monte Carlo simulation. ALPGEN interfaced with HERWIG predicts $\sigma_{\text{fid}}(\bar{t}t + \text{HF}) = 0.10$ pb.

In the data, 1656 $\bar{t}t$ dilepton candidate events are observed with at least three jets, at least two of which are $b$-tagged. The total background estimate, which is dominated by LF jets misidentified as $b$ jets from top-quark decay, is found to be $112 \pm 4^{(\text{stat})}$, leading to a background subtracted yield of $1544 \pm 41^{(\text{stat})}$. Using Eq. (2), and the quoted acceptance factor for $\bar{t} + j$ production, $\sigma_{\text{fid}}(\bar{t} + j)$ is found to be $2.55 \pm 0.07^{(\text{stat})}$ pb, compared to 2.83 pb predicted by ALPGEN and HERWIG. Taking into account the total uncertainty, it is found that $R_{HF} = [6.2 \pm 1.1^{(\text{stat})} \pm 1.8^{(\text{syst})}]\%$.

A full breakdown of the systematic uncertainties contributing to $R_{HF}$ is given in Table IV.

The extracted value of $\sigma_{\text{fid}}(\bar{t}t + \text{HF})$ is very sensitive to the value of $F_{b/\text{HF}}$. As indicated in Sec. V, the efficiency for $\bar{t} + b + X$ events is approximately a factor of 3 higher than the corresponding efficiency for $\bar{t}t + c + X$ events, implying a potential change in $\sigma_{\text{fid}}(\bar{t}t + \text{HF})$ by a factor of 3 if $F_{b/\text{HF}}$ is allowed to vary over the full range $[0, 1]$.

Ideally the fitted fraction of additional $b$ jets, $F_{b/\text{HF}}$, would be extracted from data and used in the determination of $R_{HF}$. As stated previously, the limited resolving power of this data sample yields a determination of $F_{b/\text{HF}}$ that is consistent with the entire physical range of $[0, 1]$ at the one sigma level when statistical and systematic uncertainties are included. Figure 4 shows $R_{HF}$ as a function of the assumed value of $F_{b/\text{HF}}$. The value obtained in Monte Carlo simulation, $F_{b/\text{HF}} = 0.31 \pm 0.03$, is also shown. Under the most extreme assumptions $F_{b/\text{HF}} = 0(1)$, the central value of $R_{HF}$ is 10.3 (3.4)$\%$. However, assuming the additional HF jets are either entirely $c$ jets or entirely $b$ jets is inconsistent with expectations from QCD and with the predictions of the Monte Carlo generators studied.

X. CONCLUSIONS

A 4.7 fb$^{-1}$ sample of 7 TeV proton-proton collisions recorded by the ATLAS detector at the LHC was used to measure the ratio $R_{HF}$ of the fiducial cross section for the production of $\bar{t}t$ events with at least one additional HF quark jet ($\bar{t}t + b + X$ or $\bar{t}t + c + X$) to that for the production of $\bar{t}t$ events with at least one additional jet, regardless of flavor, each with $p_T > 25$ GeV and $|\eta| < 2.5$.

A fit to the vertex mass distribution for $b$-tagged jets in $\bar{t}t$ candidate events with three or more $b$-tagged jets is performed to determine the heavy- and light-flavor content of the additional $b$-tagged jets. The result of the fit shows...
that 79 ± 14(stat) ± 22(syst) of the 105 selected $b$-tagged jets originate from HF quarks, 3 standard deviations away from the hypothesis of zero $tH +$ HF production. A value of $R_{HF} = [6.2 ± 1.1 (stat) ± 1.8 (syst)]\%$ is extracted. This value of $R_{HF}$ is consistent with the leading-order predictions of 3.4% obtained from the ALPGEN Monte Carlo generator interfaced with HERWIG and 5.2% from a calculation using POWHEG interfaced with HERWIG.

Acknowledgments

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 CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF, DNRSC and Lundbeck Foundation, Denmark; EPLANET and ERC, European Union; IN2P3-CNRS, CEA-DSM/IRFU, France; GNSF, Georgia; BMBF, DFG, HGF, MPG and AvH Foundation, Germany; GSRT, Greece; ISF, MINERVA, GIF, DIP and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; RČN, Norway; MNiSW, Poland; GRICES and FCT, Portugal; MERSYS (MECTS), Romania; 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 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (United Kingdom) and in the Tier-2 facilities worldwide.

\[ \Delta R = \left(\Delta \phi^2 + \Delta \eta^2\right)^{1/2}, \]

where $\Delta \eta$ is the separation in $\eta$ between two objects and $\Delta \phi$ is the separation in $\phi$.

11. ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the $z$ axis along the beam pipe. The $x$ axis points from the IP to the center of the LHC ring, and the $y$ axis points upward. Cylindrical coordinates ($r, \phi$) are used in the transverse plane, $\phi$ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Transverse momentum and energy are defined as $p_T = p \sin \theta$ and $E_T = E \sin \theta$, respectively.
17. $\Delta R = \sqrt{\left(\Delta \phi\right)^2 + \left(\Delta \eta\right)^2}$, where $\Delta \eta$ is the separation in $\eta$ between two objects and $\Delta \phi$ is the separation in $\phi$.

STUDY OF HEAVY-FLAVOR QUARKS PRODUCED IN...

ATLAS Collaboration

1 School of Chemistry and Physics, University of Adelaide, Adelaide, Australia
2 Physics Department, SUNY Albany, Albany, New York, USA
3 Department of Physics, University of Alberta, Edmonton, Alberta, Canada
4 Department of Physics, Ankara University, Ankara, Turkey
4a Department of Physics, Gazi University, Ankara, Turkey
4b Department of Physics, Gazi University, Ankara, Turkey
4c Division of Physics, TOBB University of Economics and Technology, Ankara, Turkey
4d Turkish Atomic Energy Authority, Ankara, Turkey
5 LAPP, CNRS/IN2P3 and Université de Savoie, Annecy-le-Vieux, France
6 High Energy Physics Division, Argonne National Laboratory, Argonne, Illinois, USA
7 Department of Physics, University of Arizona, Tucson, Arizona, USA
8 Department of Physics, The University of Texas at Arlington, Arlington, Texas, USA
9 Physics Department, University of Athens, Athens, Greece
10 Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
10a Institut de Física d’Altes Energies and Departament de Física de la Universitat Autònoma de Barcelona and ICREA, Barcelona, Spain
10b Institute of Physics, University of Belgrade, Belgrade, Serbia
10c Vinca Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia
11 Department for Physics and Technology, University of Bergen, Bergen, Norway
11a Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley, California, USA
11b Department of Physics, Humboldt University, Berlin, Germany
12 Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern, Switzerland
13 School of Physics and Astronomy, University of Birmingham, Birmingham, United Kingdom
13a Department of Physics, Bogazici University, Istanbul, Turkey
13b Division of Physics, Dogus University, Istanbul, Turkey
13c Department of Physics Engineering, Gaziantep University, Gaziantep, Turkey
13d INFN Sezione di Bologna, Italy
14 Department of Physics, Brandeis University, Waltham, Massachusetts, USA
15 Universidad Federal do Rio De Janeiro COPPE/EE/IF, Rio de Janeiro, Brazil
15a Federal University of Juiz de Fora (UFJF), Juiz de Fora, Brazil
15b Instituto de Física, Universidade de Sao Paulo, Sao Paulo, Brazil
15c Physics Department, Brookhaven National Laboratory, Upton, New York, USA
15d National Institute of Physics and Nuclear Engineering, Bucharest, Romania
15e University Politehnica Bucharest, Bucharest, Romania
15f West University in Timisoara, Timisoara, Romania
16 Departamento de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
16a Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
16b Department of Physics, Carleton University, Ottawa, Ontario, Canada
17 Enrico Fermi Institute, University of Chicago, Chicago, Illinois, USA

072012-19