Search for four-top-quark production in the single-lepton and opposite-sign dilepton final states in \( pp \) collisions at \( \sqrt{s} = 13 \) TeV with the ATLAS detector

M. Aaboud et al.*
(ATLAS Collaboration)

(Received 7 November 2018; published 26 March 2019)

A search for four-top-quark production, \( t\bar{t}tt\bar{t} \), is presented. It is based on proton-proton collision data with a center-of-mass energy \( \sqrt{s} = 13 \) TeV collected by the ATLAS detector at the Large Hadron Collider during the years 2015 and 2016, corresponding to an integrated luminosity of 36.1 fb\(^{-1}\). Data are analyzed in both the single-lepton and opposite-sign dilepton channels, characterized by the presence of one or two isolated electrons or muons with high-transverse momentum and multiple jets. A data-driven method is used to estimate the dominant background from top-quark pair production in association with jets. No significant excess above the Standard Model expectation is observed. The result is combined with the previous same-sign dilepton and multilepton searches carried out by the ATLAS Collaboration and an observed (expected) upper limit of 5.3 (2.1) times the four-top-quark Standard Model cross section is obtained at 95% confidence level. Additionally, an upper limit on the anomalous four-top-quark production cross section is set in the context of an effective field theory model.

DOI: 10.1103/PhysRevD.99.052009

I. INTRODUCTION

With a mass close to the scale of electroweak symmetry breaking, the top quark, besides having a large coupling to the Standard Model (SM) Higgs boson [1,2], is predicted to have large couplings to new particles hypothesized in many models beyond the Standard Model (BSM) [3–5]. Possible new phenomena may enhance the cross sections over SM predictions for various processes involving top quarks, and in particular for the production of four top quarks [6–15]. This paper focuses on a search for four-top-quark (\( t\bar{t}tt\bar{t} \)) production via the SM processes in proton-proton (\( pp \)) collisions at the Large Hadron Collider (LHC) and the results are interpreted in the context of an effective field theory (EFT) approach where the BSM contribution is represented via a four-top-quarks contact interaction [10].

The SM four-top-quark production cross section (\( \sigma_{\text{SM}}^{t\bar{t}tt\bar{t}} \)) in \( pp \) collisions at a center-of-mass energy \( \sqrt{s} = 13 \) TeV is predicted to be \( \sigma_{\text{SM}}^{t\bar{t}tt\bar{t}} = 9.2 \) fb at next-to-leading-order (NLO) accuracy in QCD, with scale and parton distribution function (PDF) uncertainties of the order of 30% and 6%, respectively [16,17]. Previous searches for four-top-quark production using LHC Run 2 data at \( \sqrt{s} = 13 \) TeV were performed by both the ATLAS [18–20] and CMS [21–24] Collaborations. Among them, the most sensitive one is a CMS search [24] obtaining an observed (expected) 95% confidence level (C.L.) upper limit of 4.5 (2.3) times the SM expectation. Searches for anomalous \( t\bar{t}tt\bar{t} \) production via an EFT model were recently performed by the ATLAS Collaboration [19,20], which set an observed (expected) upper limit of 16 fb (31 fb) on the production cross section at 95% C.L.

The four-top-quark production events can give rise to different final states depending on the hadronic or semileptonic decay mode of each of the top quarks. The four-top-quark decay topology considered in this search corresponds to either single-lepton events with one isolated charged lepton (electron or muon)\(^1\) or dilepton events with two opposite-sign charged leptons (electrons or muons). The event topology also features high jet multiplicity and high multiplicity of jets containing \( b \)-hadrons. Signal events are characterized by high jet multiplicity and high scalar sum of the jet transverse momenta (\( H_T^{\text{jet}} \)), which provides good discrimination against the dominant background, i.e., top-quark

---

\(^1\)The \( \tau \)-leptons are not reconstructed; however, isolated electron and muon by-products of the leptonically decaying \( \tau \)-leptons are considered in the analysis.

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

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP³.
pair production in association with jets ($t\bar{t} +$ jets). Given that the four-top-quark production events are expected to contain hadronically decaying top-quark candidates with collimated or partially collimated topologies, the analysis makes use of “mass-tagged reclustered large-$R$ (RCLR) jets” [25,26], which will be described in Sec. III.

Selected events in each of the two channels are classified into several categories according to the number of jets, $b$-tagged jets and mass-tagged RCLR jets. A data-driven method is developed to estimate the dominant $t\bar{t} +$ jets background. The Monte Carlo (MC) simulation is used in order to estimate correction factors and evaluate the systematic uncertainties of the data-driven estimate.

The paper is organized as follows: the ATLAS detector is described in Sec. II. Section III summarizes the selection criteria applied to events and reconstructed objects. The simulation-based signal and background modeling, together with the data-driven estimation of nonprompt and fake lepton backgrounds are discussed in Sec. IV. Section V is devoted to the search strategy and classification of event topologies, while the $t\bar{t} +$ jets background estimation technique using data is described in Sec. VI. The systematic uncertainties are summarized in Sec. VII. Section VIII presents the results and the combination with the same-sign dilepton and multilepton final-states search [20] carried out by ATLAS.

II. ATLAS DETECTOR

The ATLAS detector [27] at the LHC is a multipurpose particle detector with a forward-backward symmetric cylindrical geometry and nearly $4\pi$ coverage in solid angle.² It consists of an inner tracking detector (ID), electromagnetic and hadronic calorimeters, and a muon spectrometer. The inner detector, including the insertable B-layer [28,29], provides charged-particle tracking from silicon pixel and microstrip detectors in the pseudorapidity region $|\eta| < 2.5$, surrounded by a transition radiation tracker that enhances electron identification in the region $|\eta| < 2.0$. The ID is surrounded by a thin superconducting solenoid providing an axial 2 T magnetic field, and by a fine-granularity lead/liquid-argon electromagnetic calorimeter covering $|\eta| < 3.2$, which provides energy measurements of electromagnetic showers.

Hadron calorimetry is also based on the sampling technique and covers $|\eta| < 4.9$, with either scintillator tiles or liquid argon as the active medium and with steel, copper or tungsten as the absorber material. An extensive muon spectrometer with an air-core toroid magnet system surrounds the calorimeters. It includes three layers of high-precision tracking chambers, which provide coverage in the range $|\eta| < 2.7$. The field integral of the toroid magnets ranges from 2.0 to 6.0 Tm across most of the detector. A two-level trigger system [30], the first level using custom hardware and followed by a software-based level, is used to reduce the event rate to a maximum of around one kHz for offline storage.

III. OBJECT AND EVENT SELECTION

Events are selected from $pp$ collisions with $\sqrt{s} = 13$ TeV recorded by the ATLAS detector in 2015 and 2016. Only events for which all relevant subsystems were operational are considered. The data set corresponds to an integrated luminosity of $36.1 \pm 0.8$ fb⁻¹ [31]. The event reconstruction is affected by multiple inelastic $pp$ collisions in a single bunch crossing and by collisions in neighboring bunch crossings, referred to as “pileup.” The number of interactions per bunch crossing in this data set ranges from about 8 to 45 interactions. Events are required to have at least one reconstructed vertex with two or more associated tracks with transverse momentum $p_T > 0.4$ GeV. If multiple vertices are reconstructed, the vertex with the largest sum of the squares of the transverse momenta of associated tracks is taken as the primary vertex [32].

Events in both the single-lepton and dilepton channels were recorded using single-lepton triggers. Events were selected using triggers with either low $p_T$ thresholds and a lepton-isolation requirement, or with higher thresholds but with a looser identification criterion and without any isolation requirement. The lowest $p_T$ threshold used for muons is 20 (26) GeV in 2015 (2016), while the higher $p_T$ threshold is 50 GeV in both years. For electrons, triggers with a $p_T$ threshold of 24 (26) GeV in 2015 (2016) and isolation requirements are used along with triggers with a 60 GeV threshold and no isolation requirement, and with a 120 (140) GeV threshold with looser identification criteria.

Electron candidates are reconstructed [33,34] from an isolated electromagnetic calorimeter energy deposit, matched to a track in the ID, within the fiducial region of $|\eta_{\text{cluster}}| < 2.47$, where $\eta_{\text{cluster}}$ is the pseudorapidity of the calorimeter energy deposit associated with the electron candidate. Candidates within the transition region between the barrel and end cap electromagnetic calorimeters, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are excluded. The electron candidates are required to have $p_T > 30$ GeV and to satisfy “tight” likelihood-based identification criteria [34] based on calorimeter, tracking and combined variables that provide good separation between electrons and jets.

Muon candidates are reconstructed [35] by combining tracks reconstructed in both the ID and the muon...
spectrum. Candidates are required to pass the “medium” identification criteria [35] and to have $p_T > 30$ GeV and $|\eta| < 2.5$. To reduce the contribution from nonprompt leptons (e.g., from semileptonic $b$- or $c$-hadron decays), photon conversions and hadrons, lepton candidates are also required to be isolated. The lepton isolation is estimated using the scalar sum of all tracks excluding the lepton candidate itself ($I_R = \sum p_T^k$) within a cone defined by $\Delta R < R_{\text{cut}}$ along the direction of the lepton. The value of $R_{\text{cut}}$ is the smaller of $r_{\text{min}}$ and 10 GeV/$p_T^l$, where $r_{\text{min}}$ is set to 0.2 (0.3) for electron (muon) candidates, and $p_T^l$ is the lepton $p_T$. All lepton candidates are required to satisfy $I_R/p_T^l < 0.06$. Finally, lepton tracks must match the primary vertex of the event: the longitudinal impact parameter $z_0$ is required to satisfy $|z_0| < 0.5$ mm, where $\theta$ is the polar angle of the track. The transverse impact parameter significance $|d_0|/\sigma(d_0)$ must be less than 5 for electrons and 3 for muons, where $d_0$ is the transverse impact parameter and $\sigma(d_0)$ is its uncertainty.

Jet candidates are reconstructed from three-dimensional topological energy clusters [36] in the calorimeter using the anti-$k_t$ jet algorithm [37–39] with a radius parameter of 0.4, and these are referred to as “small-$R$ jets”. Each topological cluster is calibrated to the electromagnetic energy scale prior to jet reconstruction [40]. The reconstructed jets are then calibrated to the particle level by the application of a calibration, jets are required to satisfy the requirement is made on the jet vertex tagger (JVT) and the overlap removal with leptons are used as inputs for further jet re-clustering [25] using the anti-$k_t$ algorithm with a radius parameter of $R = 1.0$. These reclustered large-$R$ jets are referred to as RCLR jets. The calibration corrections and uncertainties in the RCLR jets are automatically inherited from the small-$R$ jets [26]. In order to further suppress contributions from pileup and other soft radiation, the RCLR jets are trimmed [49] by removing all small-$R$ jets within a reclustered jet that have $p_T$ below 5% of the $p_T$ of the reclustered jet. Due to the pileup suppression and $p_T > 25$ GeV requirements made on the small-$R$ jets, the probability for a small-$R$ jet to be removed from the corresponding reclustered jet by the trimming requirement is less than 1%. The resulting RCLR jets are used to identify hadronically decaying top-quark candidates. RCLR jets that have $p_T > 200$ GeV, $|\eta| < 2.0$, mass larger than 100 GeV and at least one constituent small-$R$ jet are referred to as “mass-tagged RCLR jets.”

The missing transverse momentum in the event, whose magnitude will be denoted in the following by $E_T^{\text{miss}}$, is defined as the negative vector sum of the $p_T$ of reconstructed and calibrated objects in the event, where only primary objects enter the sum (e.g., RCLR jets are not used). This sum includes a term to account for energy from low-momentum particles in the event that are not associated with any of the selected objects, which is calculated from inner detector tracks matched to the reconstructed primary vertex in the event [50].

Events of interest are required to have at least one reconstructed lepton that matches, within $\Delta R < 0.15$, the lepton with the same flavor reconstructed by the trigger algorithm. Events in the opposite-sign dilepton channel are retained if they contain exactly two opposite-sign charged leptons (electrons or muons) and at least four jets satisfying the quality and kinematic criteria discussed

---

3The rapidity is defined as $y = \frac{1}{2} \ln \frac{E + p_T}{E - p_T}$, where $E$ is the energy and $p_T$ is the longitudinal component of the momentum along the beam pipe.

4The reclustered jet mass is computed from the sum of the four-momenta of the associated small-$R$ jets [26].
TABLE I. Summary of preselection requirements for the single-lepton and dilepton channels. Here $m_T^{\ell\ell}$ is the transverse mass of the lepton and the $E_T^{miss}$ vector, and $m_{T\ell\ell}$ denotes the dilepton invariant mass in the ee and $\mu\mu$ channels.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Single-lepton</th>
<th>Dilepton</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger</td>
<td>Single-lepton triggers</td>
<td></td>
</tr>
<tr>
<td>Leptons</td>
<td>1 isolated 2 isolated, opposite-sign</td>
<td></td>
</tr>
<tr>
<td>Jets</td>
<td>$\geq 5$ jets $\geq 4$ jets</td>
<td></td>
</tr>
<tr>
<td>$b$-tagged jets</td>
<td>$\geq 2$ $b$-tagged jets</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>$E_T^{miss} &gt; 20$ GeV $m_{\ell\ell} &gt; 50$ GeV</td>
<td>$E_T^{miss} + m_T^{W} &gt; 60$ GeV $</td>
</tr>
</tbody>
</table>

above, of which at least two must be $b$-tagged. In both the ee and $\mu\mu$ channels, the dilepton invariant mass ($m_{\ell\ell}$) must be above 50 GeV and outside the Z mass window 83–99 GeV. Events not in the opposite-sign dilepton channel may enter the single-lepton channel if they contain exactly one lepton and at least five jets, of which at least two are $b$-tagged. The above selection criteria imply that events containing two leptons with the same charge, or three or more leptons of any charge are excluded from the selection. This is done in order to maintain orthogonality with the complementary search with same-sign dilepton and multilepton final states [20] carried out by ATLAS, as these results are combined with the results presented here (Sec. VIII). Additional requirements are made to suppress the background from multijet production in the single-lepton channel. Requirements are made on $E_T^{miss}$ as well as on the transverse mass of the combined lepton and $E_T^{\ell\ell}$ system$^5$ ($m_T^{W}$): $E_T^{miss} > 20$ GeV and $E_T^{miss} + m_T^{W} > 60$ GeV.

The above requirements are referred to as “preselection” and are summarized in Table I. Events satisfying either the single-electron or single-muon selections are combined and treated as a “single-lepton” analysis channel, and events satisfying any of the opposite-sign lepton selections (ee, $\mu\mu$ or $e\mu$) are combined and treated as a “dilepton” analysis channel.

IV. SIGNAL AND BACKGROUND MODELING

After the event preselection, the main background processes arise from the SM production of $t\bar{t}$ + jets and single top-quarks, as well as $W$ - or Z-boson production in association with jets. Small contributions arise from the associated production of a vector boson $V$ ($V = W, Z$) or a Higgs boson and a $t\bar{t}$ pair ($t\bar{t} + V$ and $t\bar{t} + H$) and from diboson ($WW, WZ, ZZ$) production. Multijet events contribute to the selected sample via the misidentification of hadronic objects (jets, hadrons) as leptons or the presence of a nonprompt electron or muon. These events are referred to as the “fake and nonprompt lepton” background in the remainder of this paper.

MC simulation samples are used to model the expected distributions of the signal and most of the background processes. The fake and nonprompt lepton background in the single-lepton channel is estimated with a fully data-driven method. The $t\bar{t}$ + jets background, which is dominant in regions with very high jet and $b$-jet multiplicities, is estimated via a dedicated data-driven method, with some correction factors taken from the MC simulation, as described in Sec. VI. The MC samples were processed either through the full ATLAS detector simulation [51] based on GEANT4 [52], or through a faster simulation making use of parametrized showers in the calorimeters [53]. To model the effects of pileup, events from minimum-bias interactions were generated using the PYTHIA 8.186 [54] event generator and overlaid on the simulated hard-scatter events according to the luminosity profile of the recorded data. All simulated samples were processed through the same reconstruction algorithms and analysis chain as the data. In the simulation, the top-quark mass was assumed to be $m_{top} = 172.5$ GeV. The heavy-flavor decays were modeled using the EVTGEN 1.2.0 [55] program, except for processes modeled using the SHERPA generator [56].

A. Signal modeling

Simulated events for the main signal process, i.e., the four-top-quark production with SM kinematics, were generated at leading order (LO) with the MADGRAPH5_aMC@NLO 2.2.2 [17] generator and the NNPDF2.3 LO PDF set [57], interfaced to PYTHIA 8.186 using the A14 set of tuned parameters [58], which will be denoted in the following by A14 tune. The SM $t\bar{t}t\bar{t}$ sample is normalized to a cross section of 9.2 fb, computed at NLO in QCD [17].

This search also probes a BSM model with kinematic characteristics similar to those of the SM $t\bar{t}t\bar{t}$ events: the $t\bar{t}t\bar{t}$ production via an effective field theory involving a four-fermion contact interaction [10]. The EFT $t\bar{t}t\bar{t}$ sample was generated at LO with the MADGRAPH5_aMC@NLO 2.2.2 generator and the NNPDF2.3 LO PDF set, interfaced to PYTHIA 8.186 with the A14 tune. It is normalized assuming $|C_4|/\Lambda^2 = 4 \pi$ TeV$^{-2}$, where $C_4$ denotes the coupling constant and $\Lambda$ the energy scale of new physics, which yields a cross section of 926.3 fb computed using MADGRAPH5_aMC@NLO. Details of this BSM scenario can be found in Refs. [19,20].

B. Background modeling

The dominant $t\bar{t}$ + jets background estimation relies on the data-driven technique described in Sec. VI. The
validation of this technique and the extraction of the corresponding correction factors were performed with simulated MC \( \bar{t}t \) + jets events, generated with POWHEG-Box v2 [59–62], which provides NLO accuracy in QCD for the \( \bar{t}t \) process and uses the CT10 PDF set [63]. Showering was performed using PYTHIA 6.428 [64] with the CTEQ6L PDF set [65] and the PERUGIA2012 tune [66]. The hard-process factorization scale \( \mu_F \) and renormalization scale \( \mu_R \) were set to the default POWHEG value: 

\[
\mu = \sqrt{m_{\text{top}}^2 + p_{T,\text{Top}}^2}, \quad \text{where } p_{T,\text{Top}} \text{ is the transverse momentum of the top quark in the } \bar{t}t \text{ center-of-mass reference frame}.
\]

The POWHEG model resummation damping parameter, \( \alpha_{\text{damp}} \), which controls the matching of matrix elements to parton showers and regulates the high-\( p_T \) parton radiation, was set to \( m_{\text{top}} \) [67]. The sample is normalized to the theoretical cross-section value for the inclusive \( \bar{t}t \) process of 832 ± 40 pb obtained with Top++ [68], calculated at next-to-next-to-leading order (NNLO) in QCD, and including resummation of next-to-next-to-leading logarithmic soft gluon terms [69–73].

Samples of \( W/Z + \text{jets} \) events were generated with the SHERPA 2.2 [56] generator. The matrix element calculation was performed with up to two partons at NLO in QCD and up to four partons at LO using matrix elements from COMIX [74] and OPENLOOPS [75]. The matrix element calculation was merged with the SHERPA [76] parton shower (PS) using the ME + PS@NLO prescription [77]. The PDF set used for the matrix element calculation is NNPDF3.0nnlo with a dedicated PS tuning developed by the SHERPA authors. The \( W + \text{jets} \) and \( Z + \text{jets} \) samples are normalized to their inclusive production cross section estimates at NNLO in QCD, calculated with FEWZ [78,79].

Samples of single-top-quark backgrounds, corresponding to the \( Wt \) and \( s \)-channel production mechanisms, were generated with POWHEG-Box v1 [80] at NLO accuracy using the CT10 PDF set. Overlaps between the \( \bar{t}t \) and \( Wt \) final states were removed using the “diagram removal” scheme [81]. Samples of \( t \)-channel single-top-quark events were generated using the POWHEG-Box v1 [82,83] NLO generator that uses the four-flavor scheme. The fixed four-flavor PDF set CT10f4 [63] was used for the matrix element calculations. Showering was performed using PYTHIA 6.428 with the PERUGIA2012 tune. The single-top-quark samples are normalized to the approximate NNLO cross sections [84–86].

Diboson processes with one of the bosons decaying hadronically and the other leptonically were simulated using the SHERPA 2.1.1 generator. They were calculated for up to one \( ZZ \) or zero \( WW, WZ \) additional partons at NLO, and up to three additional partons at LO, using the same procedure as for \( W/Z + \text{jets} \). The CT10 PDF set was used together with a dedicated PS tuning of the SHERPA fragmentation model. All diboson samples are normalized to their NLO cross sections provided by SHERPA.

Samples of \( \bar{t}t + V \) (with \( V = W \) or \( Z \), including nonresonant \( Z/\gamma^* \) contributions) were generated with MADGRAPH5_aMC@NLO 2.3.2, using NLO in QCD matrix elements and the NNPDF3.0NLO [87] PDF set. Showering was performed using PYTHIA 8.210 and the A14 tune. The \( \bar{t}t + V \) events are normalized to their NLO cross section [17]. A sample of \( \bar{t}t + H \) events was generated using MADGRAPH5_aMC@NLO 2.3.2 generator and the NNPDF3.0NLO PDF set. Showering was performed using PYTHIA 8.210 and the A14 tune. Inclusive decays of the Higgs boson are assumed in the generation of the \( \bar{t}t + H \) sample, which is normalized to the corresponding cross section calculated at NLO [88,89]. Rare backgrounds, such as \( \bar{t}t + WW \) and triple-top-quark production (\( \bar{t}t + t \), \( \bar{t}t + tW \)), were generated at LO with MADGRAPH5_aMC@NLO 2.2.2 with no additional partons and interfaced with PYTHIA 8.186. They are normalized using cross sections computed at NLO in QCD [17,90].

C. Estimation of nonprompt and fake lepton backgrounds

In the single-lepton channel, the background from events with a fake or nonprompt lepton is estimated from data using a “matrix method” technique [91,92]. Events are selected using looser isolation or identification requirements for the lepton and are then weighted according to the efficiencies for both prompt and background (fake and nonprompt) leptons to pass the tighter default selection. These efficiencies are measured in data using dedicated control regions. The contribution from events with a fake or nonprompt lepton is found to be consistent with zero in regions defined by the presence of two or more mass-tagged RCLR jets, as well as in the regions requiring the presence of at least one mass-tagged RCLR jet and at least four \( b \)-tagged jets. The contribution is at most 6% in the rest of the signal regions (described in Sec. V).

In the dilepton channel, the majority (90%) of events containing one prompt lepton and one background lepton, arising from either a heavy-flavor hadron decay, photon conversion, jet misidentification or light-meson decay, originate from the single-lepton \( \bar{t}t + \text{jets} \) background. This contribution is included in the estimation via the data-driven technique described in Sec. VI, while the small fraction of fake and nonprompt leptons events arising from \( W + \text{jets} \) and \( \bar{t}t + V \) events is estimated from MC simulation.\(^6\) The total contribution is found to be less than 8% of the total background in the signal regions.

\(^6\) No data-driven estimation was attempted due to the statistical overlap with the same-sign dilepton and multilepton final-states search [20].
V. SEARCH STRATEGY

Signal events from SM four-top-quark production in the single lepton (opposite-sign dilepton) decay channel are characterized by the presence of one charged lepton (two opposite-sign charged leptons), missing transverse momentum from the escaping neutrino(s) and a high number of high-$p_T$ jets. At LO the single-lepton (opposite-sign dilepton) decay will potentially have an event topology with ten (eight) jets, when each parton from a top-quark decay gives rise to a separate jet: six (four) jets are light-jets and four are $b$-quark jets. However, the topology of a reconstructed event could differ due to the limited detector acceptance, the $b$-tagging efficiency, and the possible presence of jets arising either from additional radiation and multiple parton interactions (MPI) or from collimated partons not resolved as separate objects. Events are classified in several regions to optimize the sensitivity of the search, to perform a data-driven estimate of the $t\bar{t} +$ jets background (described in Sec. VI) and to validate the background prediction.

Preselected events in each of the two channels are classified according to their event topology, defined by the number of jets with $p_T > 25$ GeV and the number of $b$-jets. Several regions are split according to the mass-tagged RCLR jet multiplicity in addition to the jet and $b$-tagged jet multiplicities. In the following, a region with $m$ jets (j), of which $n$ are $b$-tagged jets (b) and from which $p$ separate mass-tagged RCLR jets (J) are reconstructed is referred to as “$m_j, n_b, p_J$.” When no mass-tagged RCLR jet multiplicity is specified, no selection on these objects is performed.

The following regions are defined to be orthogonal using the classification described above: 20 “signal regions,” 16 “validation regions,” 18 “source regions” and 2 “efficiency extraction regions,” as shown in Fig. 1.

Twelve regions in the single-lepton channel and eight regions in the dilepton channel with the largest signal-to-background ratios (up to 5.7% in the single-lepton channel and 7.0% in the dilepton channel), assuming SM $t\bar{t}t\bar{t}$ production cross section and kinematics, are referred to as signal regions. These regions are included in the simultaneous fit to extract the signal cross section and have high jet multiplicities ($\geq 9$) and $\geq 7$ for single-lepton and dilepton respectively) and high $b$-tagged jet multiplicities ($\geq 3b$). Since events from the main $t\bar{t} +$ jets background are characterized by at most one hadronically decaying top quark in the single-lepton channel and no hadronically decaying top quarks in the dilepton channel, the signal regions are split into 0, 1 and $\geq 2J$ in the single-lepton case, and into 0 and $\geq 1J$ in the dilepton case.

Twelve validation regions in the single-lepton channel and four validation regions in the dilepton channel are defined. These regions do not overlap with the signal region selections and feature low expected signal-to-background ratios (less than 1%). They are not included in the fit nor used to extract information from the data. These regions are designed primarily to validate the data-driven estimate of the $t\bar{t} +$ jets background (introduced in Sec. VI) and to confirm the validity of the assumption that the $t\bar{t} +$ jets data-driven estimate can be extrapolated to the signal regions. The validation regions in the single-lepton channel contain exactly seven or exactly eight jets of which three or at least four are $b$-tagged. In the dilepton channel, the validation regions have exactly six jets of which three or at least four are $b$-tagged. In each of the two channels these validation regions are split according to the mass-tagged RCLR jet multiplicity in the same way as the corresponding signal regions.

With the goal of estimating the $t\bar{t} +$ jets background in the signal regions, data events with lower jet and/or $b$-jet backgrounds are used.
multiplicities are used in the data-driven method described in Sec. VI. The 18 source regions are built using events with high jet multiplicity: 7, 8, 9, \( \geq 10 \) for the single-lepton channel and 6, 7, \( \geq 8 \) for the dilepton channel, out of which exactly 2 jets are \( b \)-tagged. They are used to build pseudodata event samples in the signal and validation regions with same jet multiplicities but higher number of \( b \)-tagged jets. Efficiency extraction regions are characterized by lower jet multiplicities: five or six jets for the single-lepton channel and four or five for the dilepton channel, out of which 2, 3 or \( \geq 4 \) are \( b \)-tagged. They are used to extract the \( b \)-tagging probabilities, since they provide a sample depleted of signal and dominated by \( t \bar{t} + \) jets. Neither the efficiency extraction regions nor the source regions are included in the final fit to data.

Figure 2 shows the expected shapes of the jet and \( b \)-jet multiplicity distributions after preselection in the single-lepton and dilepton channels. The distributions shown are for the total predicted background, with the \( t \bar{t} + \) jets background estimated via MC simulation, and for the considered four-top-quark signal scenarios. Figures 3(a) and 3(b) show the same distributions but for the mass-tagged RCLR jet multiplicity.

Figures 3(c) and 3(d) compare the expected shapes of the scalar sum of the jet transverse momenta, considering all selected jets (\( H_T^{\text{had}} \)), between the different four-top-quark signal scenarios and the total predicted background. Given the different kinematic features, the \( H_T^{\text{had}} \) distribution provides a suitable discrimination between events from the signal hypotheses and the background, and is used as the main discriminating variable in each of the regions. The signal-to-background discrimination is therefore provided by the combination of the event categorization and the \( H_T^{\text{had}} \) distribution in each category.
VI. $\bar{t}t$ + jets BACKGROUND ESTIMATION USING DATA: THE TRF$_{\bar{t}t}$ METHOD

The MC simulation-based approach at NLO accuracy in QCD for the prediction of the inclusive $\bar{t}t$ background is not expected to model well the very high jet and $b$-jet multiplicity regions exploited in this search. Given the lack of multileg calculations, the MC simulation-based approach relies on the description of such large multiplicities through the parton-shower formalism with consequently large uncertainties. Therefore, a data-driven method is used to estimate the dominant background from $\bar{t}t +$ jets in regions with very high jet and $b$-jet multiplicities. This method provides a more accurate prediction of this background than a purely simulation-based approach and avoids the need to estimate modeling uncertainties (documented in Sec. VII) by extrapolation from kinematic regimes with different numbers of jets and $b$-tagged jets.

The estimate is based on a method introduced in Ref. [93] and is referred to as “tag rate function for $\bar{t}t +$ jets events,” which will be denoted in the following by TRF$_{\bar{t}t}$. The method assumes that the probability of $b$-tagging an additional$^7$ jet in a $\bar{t}t +$ jets event, where the additional jets can include $c$- and $b$-jets, is essentially independent of the number of additional jets. With this assumption, the tagging probability, as a function of the kinematic properties of the jet, can be estimated in lower jet-multiplicity events and then applied to data events with the same jet multiplicity as signal-region events, but lower $b$-tagged jet multiplicity, where the signal contamination is negligible. These $b$-tagging probabilities are measured and applied as a function of some of the jet and event properties. Simulation-based corrections are then applied in order to correct for the fact that the assumptions stated above may not be completely valid. Systematic uncertainties in these corrections are propagated through the final estimate.

$^7$Additional refers to all jets in addition to the $b(\bar{b})$-jets originating from the $\bar{t}t$ decay. This includes the jets possibly originating from hadronically decaying $W$ bosons.
Figure 4 shows the measured values of the $b$-tagging probability in the single-lepton and dilepton channels. It can be seen how the $e_b^{≥3b}$ are systematically lower than the $e_b^{≥2b}$ in the case of the single-lepton channel, while they are systematically higher in the case of the dilepton channel. This effect is due to the presence of hadronically decaying $W$ bosons only in the single-lepton channel, which can give rise only to light-jets or $c$-jets. In the dilepton case, when $e_b^{≥3b}$ is computed in the dominant four jet multiplicity, this leaves only one jet where this $b$-tagging probability can be sampled, and this jet is likely to be a $b$-jet or $c$-jet, neglecting the mistag probability and considering the relative contributions of $t\bar{t}$ + single and double $c/b$ through gluon splitting. This is not the case in the single-lepton channel, where, instead, three tagged jets out of five can easily be the consequence of tagging a $c$-jet from the $W$ boson, hence reducing the probability of tagging an additional jet. In the dilepton case, the dependence on $\Delta R_{\text{jet,jet}}^{\text{jet,jet}} \times N_{\text{jet}}^{\text{jet}}$ for $e_b^{≥3b}$ selection was found to be compatible with a constant value within statistics.

These $b$-tagging probabilities are then used to build “pseudodata samples” in validation and signal regions: this is done by applying the information derived from the measured $e_b$ to the data in the source regions containing the same number of jets and mass-tagged RCLR jets, accounting for the fact that this starting sample contains two

---

The per-jet $b$-tagging probabilities $e_b$ are measured in the efficiency extraction regions (described in Sec. V), after subtracting the contribution from all non-$t\bar{t}$ processes modeled with MC simulation, amounting to $8\%$–$14\%$ of the total background, depending on the channel and on the signal region considered. In order to take into account the correlation of $e_b$ with the $b$-tagged jet multiplicities, two sets of probabilities $e_b^{≥2b}$ and $e_b^{≥3b}$ are extracted separately for each of the two analysis channels. The measurement of $e_b^{≥2b}$ ($e_b^{≥3b}$) is done from events with $≥2$ ($≥3$) $b$-tagged jets. The two (three) $b$-tagged jets with the highest values of the multivariate $b$-tagging discriminant in the event are excluded from the computation. All probabilities $e_b$ are measured both as a function of jet $p_T$ and as a function of the quantity $\Delta R_{\text{min,jet}}^{\text{jet,jet}} \times N_{\text{jet}}^{\text{jet}}$: the minimum distance in the $\eta$-$\phi$ plane between the given jet and all the other jets in the event, multiplied by the jet multiplicity $N_{\text{jet}}$, chosen in order to take into account the correlation between the $b$-tagging probability and the presence of nearby jets (see Ref. [93]).

---

Assuming a uniform random distribution of jets across the $\eta$-$\phi$ plane, $\Delta R_{\text{min,jet}}^{\text{jet,jet}}$ is inversely proportional to $N_{\text{jet}}$. Variables parametrizing the $b$-tagging probability should be chosen to be mostly independent of $N_{\text{jet}}$, to allow the extrapolation of the $b$-tagging probabilities from low to high multiplicity regions.
The small non-$b$-tagged jets background contribution is subtracted, analogously to the procedure described in Ref. [94]. In this way, jets that were not $b$-tagged in the original data sample can be promoted to $b$-tagged jets in a given pseudodata events sample, with a weight determined by $\epsilon_b$, which accounts for the corresponding probability. For the estimate in the $3b$ categories, the procedure above is applied using only $b$-tagging probabilities extracted from events in the $\geq 2b$ region ($e_b^{\geq 2b}$). For the estimate in the $\geq 4b$ categories, a two-step procedure is applied: the estimates in the corresponding $3b$ categories are used as the starting point to

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Category & Events & Weight \\
\hline
$3b$ & 120 & $\epsilon_b^{\geq 2b}$ \\
$4b$ & 40 & $\epsilon_b^{\geq 3b}$ \\
$5b$ & 10 & $\epsilon_b^{\geq 4b}$ \\
\hline
\end{tabular}
\caption{Example table with event weights for different $b$-tagging categories.}
\end{table}

FIG. 5. Comparison between data and prediction of the $H_{\text{had}}$ distributions in the single-lepton validation regions prior to the combined fit to data in the single-lepton and dilepton channels. The $t\bar{t} + \text{jets}$ background is estimated with the data-driven method. The $t\bar{t} + V$ and $t\bar{t} + H$ processes are denoted $t\bar{t} + H/V$. Contributions from $W/Z + \text{jets}$, single-top, diboson and multijet backgrounds are combined into a single background source referred to as “Non-$t\bar{t}$”. The hashed area represents the combined statistical and systematic uncertainties of the prediction. The last bin in all figures contains the overflow. The lower panel shows the ratio between the data and the total prediction, including the SM $t\bar{t}t\bar{t}$ signal. An arrow indicates that the point is off scale.
apply again the same procedure, now using $b$-tagging probabilities extracted from events in the $\geq 3b$ region ($e_b^{\geq 3b}$).

The last step of the method relies on the MC simulation to correct the estimate in each of the considered bins and to assign a set of systematic uncertainties. In order to achieve this, all the steps described above are applied to MC simulated $t\bar{t} +$ jets events: the $b$-tagging probability $e_b$ is extracted from simulated events in the efficiency extraction regions and is then used to reweight simulated events in the source regions, obtaining an estimate in the signal and validation regions. The resulting estimate in bin $i$ of $H_T^{\text{had}}$ distributions, $B_{i}^{\text{TRF,MC}}$, is then compared with the prediction from simulated events selected in the signal and validation regions, $B_{i}^{\text{MC}}$, extracting a correction factor for each considered bin, defined as $C_i = B_{i}^{\text{MC}} / B_{i}^{\text{TRF,MC}}$. The correction is then applied bin-by-bin to the purely data-driven estimate, $B_{i}^{\text{TRF,Data}}$, to obtain a corrected estimate in each bin, $B_{i}^{\text{TRF}} = C_i \times B_{i}^{\text{TRF,Data}}$. These corrections reweight $B_{i}^{\text{TRF,Data}}$ by less than 20% on average, varying in magnitude region by region, and are primarily aimed to account for effects such as the dependence of the $b$-tagging probability on other jet or event properties than the ones used in the parametrization.

A full set of systematic uncertainties is then derived for the estimate $B_{i}^{\text{TRF}}$ by repeating the described procedure on MC simulated events with systematic variations applied. For each considered source of systematic uncertainty affecting the $t\bar{t} +$ jets MC prediction (see Sec. VII), a new set of correction factors $C'_i$ is derived. In this ratio, systematic variations $\Delta B_i$ partially cancel out since $C'_i = \frac{(B + \Delta B_i)^{\text{MC}}}{(B + \Delta B_i)^{\text{TRF,MC}}} \approx C_i \times [1 + (\Delta B_i/B_i)^{\text{MC}}]$. The cancellation is exact for some uncertainties, e.g., overall normalization. Besides the systematic uncertainties, two sources of statistical uncertainties are considered. The first is the statistical uncertainty affecting the purely data-driven estimate, due to the limited numbers of data events in the source regions. The second source comes from the MC correction factor, given the limited number of simulated events both in the source regions and in the signal and validation regions.

Validation regions are designed primarily to validate the TRF $t\bar{t}$ data-driven estimate of the dominant $t\bar{t} +$ jets background and confirm the validity of the assumption that the estimate can be extrapolated to the signal regions. Comparisons of the $H_T^{\text{had}}$ distributions between data and the total SM prediction (including the SM four-top-quark signal) in the validation regions prior to the fit to data are presented in Fig. 5 for the single-lepton channel and in Fig. 6 for the dilepton channel. The $t\bar{t} +$ jets background is estimated with the data-driven method, including the MC correction factors and the systematic uncertainties. Data agree well with the SM expectation within the uncertainties, validating the overall data-driven procedure and the assumptions made.

VII. SYSTEMATIC UNCERTAINTIES

Several sources of systematic uncertainties that can affect the normalization of signal and background and the shape of the $H_T^{\text{had}}$ distributions are considered. The systematic uncertainties of the data-driven estimate for the $t\bar{t} +$ jets background are propagated as described in Sec. VI. For each considered source of systematic uncertainty affecting the $t\bar{t} +$ jets MC prediction, a new set of correction factors $C'_i$ is derived, by coherently replacing the nominal MC prediction with the systematic variation in all regions. The usage of this data-driven technique to estimate the $t\bar{t} +$ jets background, as
opposed to a purely simulation-based approach, allows to reduce significantly the uncertainty on its prediction in the high jet and $b$-tagged jet multiplicity topologies exploited by this search.

A. Experimental uncertainties

The uncertainty in the combined 2015+2016 integrated luminosity affecting the overall normalisation of all processes estimated from the simulation is 2.1%. It is derived, following a methodology similar to that detailed in Ref. [31], and using the LUCID-2 detector for the baseline luminosity measurements [95], from calibration of the luminosity scale using $x$-$y$ beam-separation scans. This systematic uncertainty is applied to all processes modeled using MC simulations.

Uncertainties associated with jets primarily arise from the jet energy scale. The jet energy scale (JES) and its uncertainty are derived by combining information from test-beam data, LHC collision data and simulation [41]. The JES uncertainty is split into 21 uncorrelated sources, which have different dependencies on jet $p_T$ and $\eta$. In particular, three uncertainties account for differences in the jet response and simulated jet composition of light-quark, $b$-quark, and gluon-initiated jets. The flavor response uncertainties are derived by comparing the average jet response for each jet flavor using PYTHIA and HERWIG++. The flavor composition uncertainty is assumed to be a 50% quark and 50% gluon composition with a conservative 100% uncertainty. Uncertainties in the jet mass scale, the jet energy resolution and the efficiency to pass the JVT requirement are also considered.

The efficiency of the $b$-tagging algorithm is measured for each jet flavor using control samples in data and in simulation. From these measurements, correction factors are derived to match the tagging rates in the simulation [44,47,48]. Uncertainties in these corrections include a total of six independent sources affecting $b$-jets and four independent sources affecting $c$-jets. Each uncertainty has a different dependence on jet $p_T$. Seventeen uncertainties are considered for the light-jet tagging, which depend on the jet $p_T$ and $\eta$. These systematic uncertainties are taken as uncorrelated between $b$-jets, $c$-jets, and light-flavor jets. An additional uncertainty is included due to the extrapolation of these corrections to jets with $p_T$ beyond the kinematic reach of the data calibration samples used ($p_T > 300$ GeV for $b$- and $c$-jets and $p_T > 750$ GeV for light-jets) and is taken to be correlated among the three jet flavors.

Uncertainties associated with leptons arise from the trigger, reconstruction, identification, and isolation efficiencies, as well as the lepton momentum scale and resolution. These are measured in data using leptons in $Z \rightarrow \ell^+\ell^-$ and $J/\psi \rightarrow \ell^+\ell^-$ events at $\sqrt{s} = 13$ TeV [34,35].

All uncertainties in energy scales and resolutions are propagated to the missing transverse momentum. Additional small uncertainties associated with the modeling of the underlying event affecting the reconstruction of the missing transverse momentum are also taken into account.

B. Modeling uncertainties

As mentioned in Sec. VI, common normalization uncertainties for $t\bar{t}$ + jets that equally affect $B_{\text{MC}}$ and $P_{\text{TRF}}$ have no impact on their ratios $C_i$, and consequently on the total TRF prediction. Instead, uncertainties in the $t\bar{t}$ + jets heavy-flavor content or kinematics can have residual systematic effects on the TRF prediction. Therefore, no uncertainty is assigned to the inclusive $t\bar{t}$ production cross section in the search, while variations of the relative fractions of $t\bar{t}$ events with additional jets originating from $b$- and $c$-quarks, as well as comparisons of $t\bar{t}$ + jets kinematics with alternative predictions, are considered as systematic uncertainties related to the theory modeling of the $t\bar{t}$ + jets process, as described below. A categorization of $t\bar{t}$ + jets events is performed for the purpose of assigning systematic uncertainties associated with the modeling of heavy-flavor production in different topologies [96]. Events are categorized depending on the flavor content of additional particle jets and labeled either $t\bar{t} + b$-jets or $t\bar{t} + c$-jets, while the remaining events are labeled as $t\bar{t}$ + light-jets events, including those with no additional jets.

Detailed comparisons of $t\bar{t} + b$ production between the nominal NLO POWHEG-BOX v2 + PYTHIA 6.428 $t\bar{t}$ inclusive MC sample and an NLO prediction based on SHERPA + OPENLOOPS [56,75] (referred to as SHERPAOL) have shown that the cross sections agree within 50% [97]. Therefore, a normalization uncertainty of 50% is applied to the $t\bar{t} + b$ component of the $t\bar{t}$ + jets background obtained from the POWHEG-BOX v2 + PYTHIA 6.428 MC simulation. In the absence of an NLO prediction for the $t\bar{t} + c$ background, a 50% systematic uncertainty is also applied to the $t\bar{t} + c$ component, and the uncertainties in the $t\bar{t} + b$ and $t\bar{t} + c$ background normalizations are taken as uncorrelated. The overall normalization of all systematic uncertainties in the $t\bar{t}$ + jets prediction, except these explicit uncertainties in the $t\bar{t} + c$ and $t\bar{t} + b$ normalizations, is fixed to the nominal one and only migrations across categories and distortions to the shape of the kinematic distributions are considered.

To provide a comparison with a different parton-shower model, an alternative $t\bar{t}$ sample was generated using the same POWHEG model setup as for the nominal sample described in Sec. IV, except the PS, hadronization, underlying-event (UE) and MPI are simulated using HERWIG++ (version 2.7.1) [98] with the UEEE5 tune [99] and the corresponding CTEQ6L1 PDF set. To assess the systematic uncertainties related to the use of different models for the hard-scattering generation, while maintaining the same PS model, a sample using MadGRAPH5_ aMC@NLO [17] interfaced to HERWIG++ 2.7.1 was
generated. The effects of initial- and final-state radiation (ISR/FSR) are explored using two alternative Powheg-Box v2 + Pythia 6.428 samples, one with $h_{\text{damp}}$ set to $2 \times m_{\text{top}}$, the renormalization and factorization scales set to half the nominal value and using the PERUGIA2012 high-variation UE tune, giving more radiation, and one with the PERUGIA2012 low-variation UE tune, $h_{\text{damp}} = m_{\text{top}}$ and the renormalization and factorization scales set to twice the nominal value, giving less radiation [100]. The $\mu_R$ and $\mu_F$ scale variations and the $h_{\text{damp}}$ variations are kept correlated, since the two proposed variations cover the full set of uncertainties obtained by changing the scales and the resummation damping parameter independently.

Previous studies have seen that NNLO calculations provide better agreement with data than NLO calculation, particularly for the top-quark $p_T$ distribution [101]. Hence, an uncertainty in the modeling of the top-quark $p_T$ distribution is evaluated by taking the full difference between applying and not applying the reweighting to match the predictions at NNLO accuracy in QCD [102,103] of the top-quark $p_T$ distribution. This uncertainty only affects the $t\bar{t}$ + light-jets and $t\bar{t} + 1\ell$ events, for which NNLO predictions have been derived in literature.

In the case of $t\bar{t} + \geq 1b$ events, an uncertainty is assigned by comparing the NLO prediction in the four-flavor scheme of $t\bar{t} + \geq 1b$ including parton shower [97] based on SherpaOL with the nominal NLO Powheg-Box v2 + Pythia 6.428 inclusive $t\bar{t}$ MC sample with a five-flavor scheme, by means of a generator-level reweighting, as detailed in Ref. [96]. This reweighting is performed separately for each of the $t\bar{t} + \geq 1b$ subcategories in such a way that their inter-normalization and the shape of the relevant kinematic distributions are at NLO accuracy, while preserving the nominal $t\bar{t} + \geq 1b$ cross section in Powheg-Box v2 + Pythia 6.428. Additional uncertainties are assessed for those contributions of $t\bar{t} + \geq 1b$ background which are not part of the NLO prediction, namely from MPI or FSR from top-quark decay products. They are assessed via the alternative radiation samples described above.

Uncertainties affecting the modeling of the $W/Z + \text{jets}$ background include 5% scale uncertainty from their respective normalizations to the theoretical NNLO cross sections [104]. An additional 24% normalization uncertainty is added in quadrature for each additional inclusive jet-multiplicity bin, based on a comparison among different algorithms for merging LO matrix elements and parton showers [105]. Therefore, normalization uncertainties of 54% and 59% are assigned for events with exactly five jets and at least six jets, respectively. These normalization uncertainties are taken as correlated (uncorrelated) across jet multiplicities within signal regions (efficiency extraction regions). Uncertainties affecting the modeling of the single-top-quark background include an uncertainty of $+5\%$ and $-4\%$ in the total cross section estimated as a weighted average of the theoretical uncertainties in $t$, $Wt$- and $s$-channel production [84–86].

Uncertainties in the diboson background normalization include 5% from the NLO cross sections [106], as well as an additional 24% normalization uncertainty added in quadrature for each additional inclusive jet-multiplicity bin: this assumes that two of the jets originate from the $W/Z$ decays, as in $WW/WW \rightarrow \ell\ell\ell\ell$. Recent comparisons between data and Sherpa 2.1.1 for $WZ(\rightarrow \ell\ell\ell\ell)+\geq 4$ jets show agreement within the experimental uncertainty of approximately 40% [107], which further justifies the above uncertainties. Uncertainties in the $t\bar{t} + V$ and $t\bar{t} + H$ normalizations are $\pm 15\%$ and $^{+10\%}_{-13\%}$, respectively, from the uncertainties in their respective NLO cross sections [88,89,108,109].

For the determination of the SM $t\bar{t}t\bar{t}$ production signal strength, no uncertainty is assigned to the theoretical cross section for this process. In the extraction of the exclusion limits on $t\bar{t}t\bar{t}$ production via BSM models, the SM $t\bar{t}t\bar{t}$ process is considered as background and a conservative 50% normalization uncertainty is assigned to the total $t\bar{t}t\bar{t}$ cross section, taking into account the uncertainties in both its production cross section and possible acceptance and shape variations [16,17].

Uncertainties in the data-driven fake or nonprompt lepton background estimate include contributions from the limited sample size in data, particularly at high jet and $b$-tag multiplicities, from the uncertainty in the real and fake efficiencies extracted from data in dedicated control regions (e.g., selected with a requirement on either the maximum $E_T^{\text{miss}}$ or $m_W$), as well as from the extrapolation from these control regions to the analysis regions, as detailed in Ref. [92]. Based on comparisons between data and the total prediction in these control regions, the normalization uncertainties assumed for this background are 50% (100%) for events with a central (forward) electron, and 50% for muons, taken to be uncorrelated across regions with different mass-tagged RCLR jet multiplicities and between electron and muon channels. No explicit shape uncertainty is assigned due to the associated large statistical uncertainties. These uncertainties are uncorrelated between bins in the final discriminant distribution and effectively cover possible shape uncertainties.

**VIII. RESULTS**

Following the statistical method presented below, four-top-quark production signals are searched for by performing a binned profile likelihood fit to the $H_T^{\text{had}}$ distribution simultaneously in the 12 signal regions in the single-lepton channel and 8 signal regions in the dilepton channel, using a total of 20 final-state topologies. The single-lepton and dilepton channels are combined in order to gain sensitivity to different four-top-quark production signals.
A. Statistical interpretation

For each search, the $H_{\text{had}}$ distributions across all regions considered are jointly analyzed to test for the presence of a signal predicted by the benchmark scenarios. The statistical interpretation uses a binned likelihood function $\mathcal{L}(\mu, \theta)$ constructed as a product of Poisson probability terms over all bins considered in each search (namely, all $H_{\text{had}}$ bins in the 20 signal regions defined in Fig. 1). The likelihood function depends on the signal-strength parameter $\mu$, a multiplicative factor that scales the number of expected signal events, and $\theta$, a set of nuisance parameters (NPs) that encode the effect of systematic uncertainties on the signal.
and background expectations, which are implemented in the likelihood function as Gaussian, log-normal or Poisson constraints. Individual sources of systematic uncertainty are considered to be uncorrelated. Correlations of a given systematic uncertainty are maintained across processes and channels. The statistical uncertainty of the prediction, which incorporates the statistical uncertainty of the MC and background expectations, is included in the likelihood in the form of events and of the data-driven fake and nonprompt lepton estimate, is included in the likelihood in the form of additional nuisance parameters, one for each of the included bins.

The test statistic $q_{\mu}$ is defined as the profile likelihood ratio: $q_{\mu} = -2 \ln(L(\mu, \hat{\mu}, \hat{\theta})/L(\hat{\mu}, \hat{\theta}))$, where $\hat{\mu}$ and $\hat{\theta}$ are the values of the parameters that maximize the likelihood function (with the constraint $0 \leq \hat{\mu} \leq \mu$), and $\hat{\theta}$ are the values of the NPs that maximize the likelihood function for a given value of $\mu$. The test statistic $q_{\mu}$ is implemented in the RooFit package [110,111]. In the absence of any significant excess above the background expectation, upper limits on the signal production cross section for each of the signal scenarios considered in Sec. IVA are derived by using $q_{\mu}$ and the CL$_{s}$ method [112,113]. For a given signal scenario, values of the production cross section (parametrized by $\mu$) yielding $\text{CL}_{s} < 0.05$, where $\text{CL}_{s}$ is computed using the asymptotic approximation [114], are excluded at $> 95\%$ C.L.

**B. Comparison between data and prediction in signal regions after the fit to data**

A binned likelihood fit to the data is performed in the 12 signal regions in the single-lepton channel and 8 signal regions in the dilepton channel, leading to good agreement between data and postfit estimates. Comparisons of the $H_{\infty}^{\text{had}}$ distributions between data and the total SM...
Table II shows the postfit impact of the largest sources of systematic uncertainty on the signal strength $\mu$ after the simultaneous fit to data in the single-lepton and dilepton channels. The leading sources of systematic uncertainty vary depending on the analysis region considered. The largest contributions are due to the uncertainty associated with the choice of $t\bar{t} +$ jets parton shower and hadronization model and that of the $t\bar{t} +$ jets NLO generator, as well as large statistical uncertainties associated with the background prediction.

### C. Limits on four-top-quark production in the single-lepton and dilepton channel

No significant excess of events above the SM background prediction, excluding the SM $t\bar{t}t\bar{t}$ production, is found. In the case of $t\bar{t}t\bar{t}$ production with SM kinematics, an observed (expected) 95% C.L. upper limit on the production cross section of 47 fb (33 fb) is obtained, corresponding to an upper limit on $\sigma(t\bar{t}t\bar{t})$ relative to the SM prediction of 5.1 (3.6). The SM fitted signal strength $\mu$, after combination of the single-lepton and dilepton channels, is measured to be $1.7^{+1.9}_{-1.7}$.

The search is used to set limits on BSM four-top-quark production via an EFT model (see Sec. IV). For setting limits on this BSM model, the SM $t\bar{t}t\bar{t}$ process is considered as a background. In the case of $t\bar{t}t\bar{t}$ production via an EFT model with a four-top-quark contact interaction, an observed (expected) 95% C.L. upper limit on the production cross section of 21 fb (22 fb) is obtained. The cross-section limit for the contact interaction case is lower than in the SM because the contact interaction tends to result in final-state objects with slightly larger momenta (see e.g., Fig. 3). The upper limit on the

![Figure 9](https://example.com/figure9.png)

**FIG. 9.** (a) Summary of the 95% C.L. upper limits on $\sigma(t\bar{t}t\bar{t})$ relative to the SM prediction in the individual channels and for the combination. The observed limits (solid black lines) are shown together with the expected limits in the background-only hypothesis (dashed black lines) and in the SM signal-plus-background hypothesis case (dashed red lines). One- and two-standard-deviation uncertainty bands around the expected limits in the background-only hypothesis are also shown. (b) Summary of the signal-strength measurements in the individual channels and for the combination. The statistical uncertainties are evaluated from a fit to the data performed with all the nuisance parameters associated with systematic uncertainties fixed to their postfit values from the nominal fit.
production cross section can be translated into an observed (expected) limit on the free parameter of the model $|C_{4f}|/\Lambda^2 < 1.9 \text{ TeV}^{-2}$ (1.9 TeV$^{-2}$).

**D. Combination with the same-sign dilepton and multilepton final-state search**

The ATLAS Collaboration has carried out a search for new physics using 36.1 fb$^{-1}$ of $pp$ collisions at $\sqrt{s} = 13 \text{ TeV}$ in the same-sign dilepton and multilepton final states (referred to as “SS dilepton/trilepton” channel) [20]. In order to improve the sensitivity to final states containing four top quarks, the results of the search in single-lepton events or dilepton events with two opposite-sign charged leptons reported in Sec. VIII C (referred to as “single lepton/OS dilepton” channel) are combined with the results from the complementary SS dilepton/trilepton channel.

In the combination, all the experimental systematic uncertainties (described in Sec. VII A) are treated as fully correlated between the two channels, while all the background modeling systematic uncertainties described in Sec. VII B are kept uncorrelated with those in the SS dilepton/trilepton channel. This choice is motivated by the different nature of most of the background contributions in the two channels, the different importance of the common background processes and the different techniques used for the data-driven estimates.

The expected sensitivity to the SM $t\bar{t}t\bar{t}$ production from the combination of the two searches, expressed in terms of signal significance relative to the background-only prediction, is 1.0 standard deviation, while the observed value is 2.8 standard deviations. The excess is driven by the SS dilepton/trilepton channel, where the observed (expected) SM $t\bar{t}t\bar{t}$ signal significance amounts to 3.0 (0.8) standard deviations, to be compared with the 1.0 (0.6) standard deviation found in the single lepton/OS dilepton search. The kinematic properties of the SS dilepton/trilepton events were compared with the expectations from the BSM $t\bar{t}t\bar{t}$ production benchmark models studied therein, and found to agree poorly with all of them, in particular for the $b$-tagged jet multiplicity.

Assuming no signal, an observed (expected) 95% C.L. upper limit on the SM four-top-quark production cross section of 49 fb (15 fb) is obtained. It corresponds to an upper limit on $\sigma(t\bar{t}t\bar{t})$ relative to the SM prediction of 5.3 (2.1). In the signal-plus-background hypothesis, the best-fit value of the SM cross section is found to be $\sigma_{SM} = 28.5^{+12}_{-11} \text{ fb}$, to be compared to the theoretical prediction of $9.2^{+2.9}_{-2.3}$ (scale) ±0.5 (PDF) fb [17]. Figure 9(a) shows the expected and observed upper limits on $\sigma_{SM}^{t\bar{t}t\bar{t}}$ for the two searches separately and for the combined search, while Fig. 9(b) shows a summary of the signal-strength measurements for each of the two searches and their combination. In the SS dilepton/trilepton channel the uncertainty in $\mu$ is mainly statistical, while the systematic uncertainties dominate the sensitivity of the search in the single lepton/OS dilepton channel. The probability that the results of the two searches are compatible is assessed by comparing the maximum-likelihood values for a fit performed after decorrelating the signal-strength parameters in the two channels and for the nominal combined fit with a common signal-strength parameter. The probability of obtaining a discrepancy between the two signal-strength parameters equal to or larger than the one obtained is found to be 31%.

Limits are also set for BSM $t\bar{t}t\bar{t}$ production via an EFT model with a four-top-quark contact interaction. In this benchmark scenario, the SM $t\bar{t}t\bar{t}$ sample is included as a background process. A combined observed (expected) 95% C.L. upper limit on the production cross section of 47 fb (33 fb) is obtained, which translates into an observed (expected) limit on the free parameter of the model of $|C_{4f}|/\Lambda^2 < 1.9 \text{ TeV}^{-2}$ (1.6 TeV$^{-2}$).

**IX. SUMMARY**

A search for four-top-quark production in the single-lepton and opposite-sign dilepton channels is presented. The analyzed data sample consists of 36.1 fb$^{-1}$ of proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ collected with the ATLAS detector at the Large Hadron Collider during 2015 and 2016. In order to improve the sensitivity of the search, events are categorized according to their jet, $b$-tagged jet and mass-tagged reclustered large-$R$ jet multiplicities. No significant excess of events above the SM background expectation is found. For the four-top-quark Standard Model production, an observed (expected) 95% C.L. upper limit on the production cross section of 47 fb (33 fb), corresponding to 5.1 (3.6) times the SM prediction, is obtained. The result is combined with the same-sign dilepton and multilepton final-states search carried out by ATLAS [20] and an observed (expected) upper limit of 49 fb (19 fb), corresponding to 5.3 (2.1) times the SM prediction is obtained at 95% C.L. Additionally, in the case of four-top-quark production via an EFT model with a four-top-quark contact interaction, a combined observed (expected) 95% C.L. upper limit on the production cross section of 21 fb (15 fb) is obtained.

**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; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China;
APPENDIX: COMPARISON BETWEEN DATA AND PREDICTION OF THE $H_T^{\text{had}}$ DISTRIBUTIONS IN VALIDATION REGIONS AFTER THE FIT TO DATA

Figures 10 and 11 show comparisons of the $H_T^{\text{had}}$ distributions in the validation regions between the data and the postfit prediction in the dilepton and in the single-lepton channels, respectively. The postfit prediction is obtained from the fit to data presented in Sec. VIII in the 20 signal regions and propagated to the validation regions, which are not included in the fit nor used to extract information from the data. The good level of agreement found between data and prediction in these regions is therefore an indication of the validity of the extrapolation of the results of the fit between different regions.

FIG. 10. Comparison between data and prediction of the $H_T^{\text{had}}$ distributions in the dilepton validation regions after the combined fit to data in the single-lepton and dilepton channels. The $t\bar{t}$ + jets background is estimated with the data-driven method. The $t\bar{t} + V$ and $t\bar{t} + H$ processes are denoted $t\bar{t} + V$ and $t\bar{t} + H$. Contributions from $W/Z + \text{jets}$, single-top, diboson and multijet backgrounds are combined into a single background source referred to as “Non-$t\bar{t}$”. The hashed area represents the combined statistical and systematic uncertainties of the prediction. The last bin in all figures contains the overflow. The lower panel shows the ratio between the data and the total prediction, including the SM $t\bar{t}t\bar{t}$ signal scaled by the best-fit signal strength. An arrow indicates that the point is off scale.
FIG. 11. Comparison between data and prediction of the $H_T^{\text{had}}$ distributions in the single-lepton validation regions after the combined fit to data in the single-lepton and dilepton channels. The $t\bar{t}$ + jets background is estimated with the data-driven method. The $t\bar{t} + V$ and $t\bar{t} + H$ processes are denoted $t\bar{t} + H/V$. Contributions from $W/Z +$ jets, single-top, diboson and multijet backgrounds are combined into a single background source referred to as “Non-$t\bar{t}$”. The hashed area represents the combined statistical and systematic uncertainties of the prediction. The last bin in all figures contains the overflow. The lower panel shows the ratio between the data and the total prediction, including the SM $t\bar{t}t\bar{t}$ signal scaled by the best-fit signal strength. An arrow indicates that the point is off scale.


[109] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trocsanyi, \(t\bar{t}W^\pm\) and \(t\bar{t}Z\) hadroproduction at NLO accuracy in QCD with parton shower and Hadronization effects, J. High Energy Phys. 11 (2012) 056.


SEARCH FOR FOUR-TOP-QUARK PRODUCTION IN THE...

PHYS. REV. D 99, 052009 (2019)


(ATLAS Collaboration)

1Department of Physics, University of Adelaide, Adelaide, Australia
2Physics Department, SUNY Albany, Albany, New York, USA
3Department of Physics, University of Alberta, Edmonton, Alberta, Canada
4Department of Physics, Ankara University, Ankara, Turkey
4aIstanbul Aydin University, Istanbul, Turkey
4cDivision of Physics, TOBB University of Economics and Technology, Ankara, Turkey
5LAPP, Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS/IN2P3, Annecy, France
6High Energy Physics Division, Argonne National Laboratory, Argonne, Illinois, USA
7Department of Physics, University of Arizona, Tucson, Arizona, USA
8Department of Physics, University of Texas at Arlington, Arlington, Texas, USA
9Physics Department, National and Kapodistrian University of Athens, Athens, Greece
10Department of Physics, National Technical University of Athens, Zografou, Greece
11Department of Physics, University of Texas at Austin, Austin, Texas, USA
12Bahcesehir University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey
12bIstanbul Bilgi University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey
12cDepartment of Physics, Bogazici University, Istanbul, Turkey
12dDepartment of Physics Engineering, Gaziantep University, Gaziantep, Turkey
13Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan
14Institut de Física d’Altes Energies (IFAE), Barcelona Institute of Science and Technology, Barcelona, Spain
15aInstitute of High Energy Physics, Chinese Academy of Sciences, Beijing, China
15bPhysics Department, Tsinghua University, Beijing, China
15cDepartment of Physics, Nanjing University, Nanjing, China
15dUniversity of Chinese Academy of Science (UCAS), Beijing, China
15eInstitute of Physics, University of Belgrade, Belgrade, Serbia
16Department for Physics and Technology, University of Bergen, Bergen, Norway
17Physics Division, Lawrence Berkeley National Laboratory and University of California, Berkeley, California, USA
19Institut für Physik, Humboldt Universität zu Berlin, Berlin, Germany
20Albert Einstein Center for Fundamental Physics and Laboratory for High Energy Physics, University of Bern, Bern, Switzerland
21School of Physics and Astronomy, University of Birmingham, Birmingham, United Kingdom
22Centro de Investigaciones, Universidad Antonio Nariño, Bogota, Colombia
23Department of Physics, Nanjing University, Nanjing, China
24INFN Sezione di Bologna, Bologna, Italy
25Institute of Physics, University of Belgrade, Belgrade, Serbia
26Department of Physics, University of California, Berkeley, California, USA
27aTransilvania University of Brasov, Brasov, Romania
27bHoria Hulubei National Institute of Physics and Nuclear Engineering, Bucharest, Romania
27cDepartment of Physics, Alexandru Ioan Cuza University of Iasi, Iasi, Romania
27dNational Institute for Research and Development of Isotopic and Molecular Technologies, Bucharest, Romania
27eUniversity Politehnica Bucharest, Bucharest, Romania
27fWest University in Timisoara, Timisoara, Romania
28Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovak Republic
29Department of Subnuclear Physics, Institute of Experimental Physics of the Slovak Academy of Sciences, Kosice, Slovak Republic
29aPhysics Department, Brookhaven National Laboratory, Upton, New York, USA
30Department de Física, Universidad de Buenos Aires, Buenos Aires, Argentina
31Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
31aDepartment of Physics, University of Cape Town, Cape Town, South Africa