Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

DOI
10.1007/JHEP11(2021)118

Publication date
2021

Document Version
Final published version

Published in
Journal of High Energy Physics

License
CC BY

Citation for published version (APA):
Measurement of the $t\bar{t}t\bar{t}$ production cross section in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS collaboration

E-mail: atlas.publications@cern.ch

ABSTRACT: A measurement of four-top-quark production using proton-proton collision data at a centre-of-mass energy of 13 TeV collected by the ATLAS detector at the Large Hadron Collider corresponding to an integrated luminosity of 139 fb$^{-1}$ is presented. Events are selected if they contain a single lepton (electron or muon) or an opposite-sign lepton pair, in association with multiple jets. The events are categorised according to the number of jets and how likely these are to contain $b$-hadrons. A multivariate technique is then used to discriminate between signal and background events. The measured four-top-quark production cross section is found to be $26^{+17}_{-15}$ fb, with a corresponding observed (expected) significance of 1.9 (1.0) standard deviations over the background-only hypothesis. The result is combined with the previous measurement performed by the ATLAS Collaboration in the multilepton final state. The combined four-top-quark production cross section is measured to be $24^{+7}_{-6}$ fb, with a corresponding observed (expected) signal significance of 4.7 (2.6) standard deviations over the background-only predictions. It is consistent within 2.0 standard deviations with the Standard Model expectation of $12.0 \pm 2.4$ fb.

KEYWORDS: Hadron-Hadron scattering (experiments), Top physics

ArXiv ePrint: 2106.11683

https://doi.org/10.1007/JHEP11(2021)118
1 Introduction

Being the heaviest known elementary particle of the Standard Model (SM), the top quark has a large coupling to the SM Higgs boson and is predicted to have large couplings to hypothetical new particles in many models of physics beyond the Standard Model. For this reason, it is particularly important to study rare processes involving the top quark. The production of four top quarks, $tttt$, is one of these processes which has not yet been
Examples of tree-level Feynman diagrams for SM $t\bar{t}t\bar{t}$ signal (left and middle) and one of the main backgrounds, $t\bar{t}$ production in association with $b$-jets (right).

observed. The $t\bar{t}t\bar{t}$ cross section could be enhanced, for instance, by gluino pair production as in supersymmetric theories [1, 2], by pair production of scalar gluons [3, 4], or by the production of a heavy scalar or pseudoscalar boson in association with a top-quark pair ($t\bar{t}$) in type-II two-Higgs-doublet models [5–7]. The $t\bar{t}t\bar{t}$ cross section is also sensitive to both the magnitude and the charge conjugation and parity properties of the Yukawa coupling of the top quark to the Higgs boson [8, 9], as well as to various four-fermion couplings in the context of the effective field theory framework [10, 11]. Within the SM, the $t\bar{t}t\bar{t}$ cross section in proton-proton ($pp$) collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV is predicted to be $\sigma_{t\bar{t}t\bar{t}} = 12.0$ fb at next-to-leading order (NLO) in QCD including NLO electroweak corrections [12], with a relative combined uncertainty of 20% dominated by the missing higher order QCD correction evaluated by varying the renormalisation and factorisation scale choices. The uncertainty from the PDF and $\alpha_s$ choice in the $t\bar{t}t\bar{t}$ cross section was estimated to be about 6.3% using the PDF4LHC prescription. An example of a Feynman diagram for SM $t\bar{t}t\bar{t}$ QCD production is shown in figure 1 (left). The electroweak $t\bar{t}t\bar{t}$ contribution is illustrated in figure 1 (middle) with an example of a Feynman diagram where a Higgs boson acts as an off-shell mediator.

The $t\bar{t}t\bar{t}$ events can give rise to several different final states depending on the hadronic or semileptonic decay mode of each of the top quarks. The final states can be grouped according to the number of electrons or muons from the semileptonic top-quark decays, including those from subsequent leptonic $\tau$ decays. The final state with two leptons\(^1\) of the same electric charge or with more than two leptons is referred to as the 2LSS/3L final state. This final state contains 13% of all produced $t\bar{t}t\bar{t}$ events and features a low level of background contamination. The final state with one lepton or two oppositely charged leptons (referred to as the 1L/2LOS final state) accounts for a larger fraction, capturing 57% of all produced $t\bar{t}t\bar{t}$ events. However, this final state suffers from a large irreducible background that is mostly composed of $t\bar{t}$ production in association with additional jets ($t\bar{t}+\text{jets}$). It is complementary to the 2LSS/3L final state and poses different challenges. The main challenge lies in the proper evaluation of the dominant background from $t\bar{t}b\bar{b}$ events with additional jets, which has significant theoretical uncertainty. An example Feynman diagram for this background is shown in figure 1 (right).

\(^1\)In the rest of this article, ‘lepton’ refers exclusively to an electron or muon.
ATLAS and CMS have already searched for $t\bar{t}t\bar{t}$ production in 13 TeV $pp$ collisions at the Large Hadron Collider (LHC). The most recent ATLAS result focused on the 2LSS/3L final state using 139 fb$^{-1}$ of data and led to the first evidence for this process with an observed (expected) significance of 4.3 (2.4) standard deviations and a measured cross section of $\sigma_{t\bar{t}t\bar{t}} = 24^{+7}_{-5}$ fb [13]. The previous ATLAS search in the 1L/2LOS final state using 36.1 fb$^{-1}$ set an observed (expected) 95% confidence level (CL) upper limit on $\sigma_{t\bar{t}t\bar{t}}$ of 47 fb (33 fb) [14]. CMS also set a 95% CL upper limit on the $t\bar{t}t\bar{t}$ production cross section in this final state of 48 fb using a 35.8 fb$^{-1}$ data set [15]. The latest CMS search with 137 fb$^{-1}$ of data in the 2LSS/3L final state lead to an observed (expected) significance for a $t\bar{t}t\bar{t}$ signal of 2.6 (2.7) standard deviations and a measured cross section of $\sigma_{t\bar{t}t\bar{t}} = 12.6^{+5.8}_{-5.2}$ fb [16].

This article presents a search for $t\bar{t}t\bar{t}$ production in the 1L/2LOS final state using the full data set of $pp$ collision data at $\sqrt{s} = 13$ TeV corresponding to 139 fb$^{-1}$. The result is then combined with that obtained in the 2LSS/3L final state using the same data set [13].

2 The ATLAS detector

The ATLAS experiment [17] at the LHC is a multipurpose particle detector with a forward-backward symmetric cylindrical geometry and a near $4\pi$ coverage in solid angle.\footnote{ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates $(r,\phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$.} It consists of an inner tracking detector (ID) surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic (EM) and hadron calorimeters, and a muon spectrometer (MS). The inner tracking detector covers the pseudorapidity range $|\eta| < 2.5$. It consists of silicon pixel, silicon microstrip, and transition radiation tracking detectors. Lead/liquid-argon (LAr) sampling calorimeters provide EM energy measurements with high granularity. A steel/scintillator-tile hadron calorimeter covers the central pseudorapidity range $(|\eta| < 1.7)$. The endcap and forward regions are instrumented with LAr calorimeters for EM and hadronic energy measurements up to $|\eta| = 4.9$. The MS surrounds the calorimeters and is based on three large air-core toroidal superconducting magnets with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. The MS includes a system of precision tracking chambers and fast detectors for triggering. A two-level trigger system is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to accept events at a maximum rate of nearly 100 kHz. This is followed by a software-based trigger that reduces the accepted event rate to 1 kHz, on average, depending on the data-taking conditions [18]. An extensive software suite [19] is used for real and simulated data reconstruction and analysis, for operation and in the trigger and data acquisition systems of the experiment.
3 Object and event selection

Events are selected from $\sqrt{s} = 13$ TeV $pp$ collision data collected by the ATLAS detector in the period between 2015 and 2018. Only events for which all detector subsystems were operational are considered. The data set corresponds to an integrated luminosity of 139 fb$^{-1}$ [20, 21]. Events were required to fire single-electron or single-muon triggers, with minimum $p_T$ thresholds varying from 20 to 26 GeV depending on the lepton flavour and the data-taking period. Triggers with minimum $p_T$ thresholds include isolation requirements [22, 23].

Events are required to have at least one vertex reconstructed from at least two ID tracks with transverse momenta $p_T > 0.4$ GeV. The primary vertex for each event is defined as the vertex with the highest sum of $p_T^2$ over all associated ID tracks [24].

Object reconstruction closely follows that of ref. [13] and is briefly summarised in the following. Electron candidates are reconstructed from energy deposits in the EM calorimeter associated with ID tracks [25]. The pseudorapidity of the calorimeter energy cluster, $\eta_{\text{cluster}}$, must satisfy $|\eta_{\text{cluster}}| < 2.47$, excluding the transition region between the barrel and the endcap calorimeters ($|\eta_{\text{cluster}}| \notin [1.37, 1.52]$). Muon candidates are reconstructed by combining tracks reconstructed in both the ID and the MS [26] and are required to have $|\eta| < 2.5$. Both the electron and muon candidates are required to have $p_T > 10$ GeV. The transverse impact parameter divided by its estimated uncertainty, $|d_0|/\sigma(d_0)$, is required to be lower than five (three) for electron (muon) candidates. The longitudinal impact parameter $z_0$ must satisfy $|z_0 \sin(\theta)| < 0.5$ mm for both lepton flavours. Electrons are required to satisfy the ‘Tight’ likelihood-based identification criterion and to be isolated according to the ‘FixedCutTight’ selection criterion [25]. Muons must satisfy the ‘Medium’ cut-based identification criterion and be isolated according to the ‘FixedCutTightTrackOnly’ selection criterion [27].

Jets are reconstructed from topological clusters [28] of energy deposits in the calorimeters using the anti-$k_t$ algorithm [29, 30] with a radius parameter of $R = 0.4$ and calibrated as described in ref. [31]. They are referred to as ‘small-$R$ jets’. These jets are required to have $p_T > 25$ GeV and $|\eta| < 2.5$. To suppress the effect of additional $pp$ collisions in the same or a nearby bunch crossing, collectively referred to as pile-up, the jets with $p_T < 120$ GeV and $|\eta| < 2.4$ must satisfy a pile-up suppression requirement based on the output of a multivariate classifier called the jet-vertex-tagger (JVT) [32]. Events are required to pass a set of quality criteria to suppress those containing any jets arising from non-collision sources or detector noise [33]. The MV2c10 multivariate algorithm [34] is used to identify jets containing $b$-hadrons. Each jet is given a score representing the likelihood of the jet to contain a $b$-hadron. A jet is $b$-tagged if the score passes a certain threshold, referred to as an operating point (OP). Four OPs are defined with average expected efficiencies for $b$-jets of 60%, 70%, 77% and 85%. A pseudo-continuous score is assigned to each jet passing these OPs, with an integer value ranging from five for jets that pass the 60% OP to two for jets passing only the 85% OP. A score of one is assigned if the jet does not pass any of the OPs.

The selected and calibrated small-$R$ jets are used as inputs for jet reclustering [35] using the anti-$k_t$ algorithm with a radius parameter of $R = 1.0$. These reclustered jets are referred to as ‘large-$R$ jets’. The calibration corrections and uncertainties for the reclustered large-$R$
jets are inherited from the small-$R$ jets [35]. A trimming procedure is applied to the reclustered jets to remove all the associated small-$R$ jets that have $p_T$ below 5% of the $p_T$ of the reclustered jet [36]. The reclustered jets are required to have $p_T > 200$ GeV and $|\eta| < 2.0$.

A sequential overlap removal procedure defined in ref. [13] is applied to ensure that the same calorimeter energy deposit or the same track is not associated with two or more different reconstructed objects.

The missing transverse momentum in the event, whose magnitude is denoted by $E_T^{\text{miss}}$, is defined as the negative vector sum of the $p_T$ of reconstructed and calibrated objects in the event [37]. This sum includes the momenta of the ID tracks matched to the primary vertex but not associated with any other objects.

The events are required to have either exactly one lepton satisfying $p_T > 28$ GeV and at least seven jets (1L channel) or exactly two oppositely charged leptons with $p_T > 28$ GeV for the leading lepton and $p_T > 10$ GeV for the subleading lepton and at least five jets (2LOS channel). In the 2LOS channel, the events with two same-flavour leptons must have a dilepton invariant mass above 15 GeV and outside the $Z$-boson mass window of 83–99 GeV. Each event is required to have at least one reconstructed lepton that matches the lepton that fired the trigger and must contain at least two $b$-tagged jets passing the 70% OP.

4 Monte Carlo samples

Monte Carlo (MC) samples of simulated events were produced to model the SM $t\bar{t}t\bar{t}$ signal and background processes. The nominal samples are identical to those used in ref. [13] and are normalised using the best theory predictions available. The NNPDF3.0NNLO [38] PDF set was used in all matrix element (ME) calculations if not stated otherwise. The top-quark mass $m_{\text{top}}$ was set to 172.5 GeV in all relevant samples.

The nominal $t\bar{t}t\bar{t}$ signal sample was generated using the MadGraph5_aMC@NLO v2.6.2 [39] generator at NLO in the strong coupling constant $\alpha_s$ with the NNPDF3.1NNLO [38] PDF set. The functional form of the renormalisation and factorisation scales were set to $\mu_r = \mu_f = m_T/4$, where $m_T$ is defined as the scalar sum of the transverse masses $\sqrt{m^2 + p_T^2}$ of the particles generated from the ME calculation, following ref. [12]. An additional sample with settings similar to the nominal was generated using a leading-order (LO) ME, which makes more efficient use of simulation resources as unlike the nominal it does not contain negative weight events. The additional sample is only used for the training of the multivariate discriminant to separate signal from background.

The nominal MC sample for $t\bar{t}$ background modelling was produced with the HVQ program [40, 41] in the POWHEGBOX v2 [40, 42–44] generator at NLO in QCD in the five-flavour scheme (5FS). The $h_{\text{damp}}$ parameter, which controls the $p_T$ of the first additional emission beyond the Born configuration, was set to $1.5m_{\text{top}}$ [45, 46]. A dedicated $t\bar{t}b\bar{b}$ sample with the highest available precision, and with the additional $b$-quarks coming from the ME, was produced at NLO QCD accuracy in the four-flavour scheme (4FS) with the POWHEGBOXRES [47] generator and OpenLoops [48, 49], with the NNPDF3.0nlonf4 [38] PDF set. For this sample, a pre-release of the implementation of this
process in PowhegBoxRes was provided by the authors [50]. The factorisation scale was set to \( \frac{1}{2} \sum_{t, \bar{t}, b, \bar{b}, j} m_{T,i} \) (where \( j \) stands for extra partons), the renormalisation scale was set to \( \sqrt{\prod_{t, \bar{t}, b, \bar{b}} m_{T,i}^2} \), and the \( h_{\text{damp}} \) parameter was set to \( \frac{1}{2} \sum_{t, \bar{t}, b, \bar{b}} m_{T,i} \) where \( m_{T,i} \) is the transverse mass of a given parton.

Single-top-quark production processes, i.e. \( tW \) associated production, \( t \)-channel and \( s \)-channel production, were modelled using the PowhegBox v2 [42–44, 51–53] generator at NLO in QCD. The \( t \)-channel process was generated in the 4FS with the NNPDF3.0NLOnf4 [38] PDF set, while for \( tW \) and \( s \)-channel processes the 5FS was used. The diagram removal scheme [54] was employed to handle the interference between \( tW \) and \( t\bar{t} \) production [46].

The \( t\bar{t}W \) events were generated using the SHERPA 2.2.1 [55] generator. The ME was calculated for up to one additional parton at NLO in QCD and up to two partons at LO QCD precision using the Comix [55] and OpenLoops [48, 49] libraries and merged with the SHERPA parton shower [56] using the MEPS@NLO prescription [57–60] and a merging scale of 30 GeV. The production of \( t\bar{t}H \) events was modelled using the PowhegBox generator at NLO in QCD. The \( t\bar{t}Z \) and \( tWZ \) events were generated with MadGraph5_aMC@NLO v2.3.3 at NLO in QCD. The other rare top-quark processes, namely \( tZ \), \( t\bar{t}WW \) and \( t\bar{t}t \) production, were all modelled using the MadGraph5_aMC@NLO generator at LO in QCD.

The production of \( V+\text{jets} \) events (where \( V = W \) or \( Z \)) was simulated with the SHERPA 2.2.1 generator using NLO-accurate MEs for up to two partons and LO-accurate MEs for up to four partons. Samples of diboson final states (\( VV \)) were also simulated with the SHERPA 2.2.1 generator.

All events generated using PowhegBox or MadGraph5_aMC@NLO were interfaced with PYTHIA 8.230 [61] for the parton shower and hadronisation, using the A14 set of tuned parameters [45] and the NNPDF2.3L0 [62] set of PDFs. Additional \( tt\bar{t} \), \( t\bar{t} \) and single-top-quark samples were produced by replacing PYTHIA 8 with HERWIG 7.04 [63, 64] for parton showering and hadronisation, using the H7UE set of tuned parameters [64] and the MMHT2014LO [65] PDF set. These samples are used to evaluate uncertainties due to the choice of parton shower and hadronisation model.

To assess the uncertainty due to the choice of generator, the \( t\bar{t} \), single-top-quark, \( t\bar{t}H \) and \( t\bar{t}W \) samples produced with the nominal generator set-ups are compared with alternative samples generated with MadGraph5_aMC@NLO for the calculation of the hard-scattering, interfaced with PYTHIA 8. An alternative sample of \( t\bar{t}Z \) events was generated with NLO MEs using SHERPA 2.2.1. All alternative samples were generated using the same PDF in the ME as in the nominal sample.

The effects of pile-up were modelled by overlaying minimum-bias events, simulated using the soft QCD processes of PYTHIA 8.186 with the A3 tune [66], on events from hard processes. For all samples of simulated events, except those generated using SHERPA, the EVTGEN 1.2.0 program [67] was used to describe the decays of bottom and charm hadrons.

The non-\( t\bar{t} \) samples were processed through the simulation [68] of the ATLAS detector geometry and response using GEANT4 [69]. A fast simulation (AtlFast 2) of the ATLAS
detector relying on parameterised showers in the calorimeter [70] was used for the $t\bar{t}t\bar{t}$ and $t\bar{t}$ samples. All samples were reconstructed using the same software used for collider data. Corrections were applied to the simulated events so that the physics objects’ selection efficiencies, energy scales and energy resolutions match those determined from data control samples.

5 Analysis strategy

In the 1L/2LOS final state, the parton-level $t\bar{t}t\bar{t}$ topology is characterised by four $b$-quarks resulting from the decay of the four top quarks and by either six or four other quarks coming from the hadronic $W$-boson decays. Consequently, the signature of the events from $t\bar{t}t\bar{t}$ production in the ATLAS detector features a high number of jets ($N_{\text{jets}}$), with 10 (8) jets in the 1L (2LOS) channel, among which four contain a $b$-hadron. This phase-space region is contaminated by a large number of background events coming almost exclusively from $t\bar{t}+\text{jets}$ production. After the preselection described in section 3, the contribution from other backgrounds is below 8% and arises mainly from $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$, single top-quark or $W/Z$ boson production in association with jets. Diboson production and rare processes including $tZq$, $tWZ$, $ttWW$ and $t\bar{t}t\bar{t}$ contribute less than 1% of all events. All small non-$t\bar{t}$ backgrounds are estimated from simulations. The contribution from the background arising from misreconstructed or non-prompt leptons is negligible in the 1L channel. This has been verified by the agreement between data and MC simulation using different isolation requirements on the leptons. In the 2LOS channel, this background is expected to be dominated by $t\bar{t}$ events with one prompt lepton. The contribution is found to be below 2% using $t\bar{t}$ MC simulations and is included in the $t\bar{t}$ background.

The nominal $t\bar{t}$ sample was generated at NLO as described in section 4, with up to one additional parton from the ME calculation. All additional jets, including the heavy-flavour ones, are produced in the parton shower. The background predictions in high $N_{\text{jets}}$ regions were found to be unreliable. Reference [71] demonstrates clear mismodelling of the number of jets not from the $t\bar{t}$ decay and the scalar sum of the transverse momenta of the reconstructed hadronically and leptonically decaying top quarks, $H_{\text{T}}^t$. In addition, the rate of $t\bar{t}$ production in association with $b$-jets was observed to be underestimated by the current MC simulations [72, 73]. Therefore, the following strategies were developed in this analysis to obtain a reliable estimate of the $t\bar{t}+\text{jets}$ background by reweighting the $t\bar{t}$ MC samples using data.

Firstly, the $t\bar{t}+\text{jets}$ MC events are classified according to the flavour of the particle jets that are not from the decay of the $t\bar{t}$ system, using the same procedure as described in ref. [74]. The particle jets are reconstructed from the simulated stable particles using the anti-$k_t$ algorithm with a radius parameter $R = 0.4$, and are required to have $p_T > 15$ GeV and $|\eta| < 2.5$. Events are labelled as $t\bar{t}+\geq 1b$ if at least one particle jet is matched within $\Delta R < 0.4$ to a $b$-hadron with $p_T > 5$ GeV that did not originate from the decay of a top quark. Similarly, events not categorised as $t\bar{t}+\geq 1b$, and where at least one particle jet is matched within $\Delta R < 0.4$ to a $c$-hadron with $p_T > 5$ GeV that did not originate from the decay of a top quark, are labelled as $t\bar{t}+\geq 1c$. The remaining events are labelled as $t\bar{t}+$light.
The selected events are categorised into different regions according to the lepton and jet multiplicities and different $b$-tagging requirements. The details of the event categorisation can be found in section 6. These regions vary in their signal-to-background ratio and the flavour composition of $t\bar{t}+\text{jets}$. Corrections to the normalisation of $t\bar{t}+\text{light}$, $t\bar{t}+\geq 1c$ and $t\bar{t}+\geq 1b$ and their kinematic distributions are derived from regions with low signal contamination using data, as described in section 7. These corrections are used to mitigate the mismodelling of $t\bar{t}+\text{jets}$ MC events in regions with a high signal-to-background ratio, where a binned profile likelihood fit is performed to extract the $t\bar{t}t\bar{t}$ signal strength. In the regions most sensitive to $t\bar{t}t\bar{t}$ production, a multivariate discriminant is used to improve the sensitivity by distinguishing between signal and background events, as discussed in section 8. Large-$R$ jets are used as proxies for hadronically decaying top quarks with high $p_T$, improving the discrimination between $t\bar{t}t\bar{t}$ signal and $t\bar{t}+\text{jets}$ background.

The different $t\bar{t}+\text{jets}$ components after the corrections are further adjusted and constrained in the profile likelihood fit. The $t\bar{t}+\geq 1b$ events are divided into further categories to define their systematic uncertainties. Events with one particle jet matched to a single $b$-hadron are labelled as $t\bar{t}+b$, those with one particle jet matched to at least two $b$-hadrons are labelled as $t\bar{t}+B$, and those with two particle jets, each matched to one $b$-hadron are labelled as $t\bar{t}+b\bar{b}$. The rest of the $t\bar{t}+\geq 1b$ events are labelled as $t\bar{t}+\geq 3b$. Section 9 documents all systematic uncertainties considered in this analysis. Details of the scheme used for the $t\bar{t}+\text{jets}$ modelling systematic uncertainties are presented in section 9.3.

6 Event categorisation

Selected events in both channels are categorised into several regions as illustrated in figure 2. The resulting regions are used for different purposes. Signal-depleted and signal-enriched regions are defined as control regions and signal regions, respectively. They are used in the profile likelihood fit to extract the signal cross section and to constrain the overall background model using the assigned systematic uncertainties. The $t\bar{t}+\text{jets}$ kinematic reweighting regions are used to extract correction factors for the $t\bar{t}+\text{jets}$ MC prediction. The validation regions are defined in order to check that the background prediction is able to describe data in regions with background composition similar to that in the signal regions.

To take advantage of the high jet multiplicity in the $t\bar{t}t\bar{t}$ signal, events in each lepton channel are first categorised according to their number of jets, from 7 (5) jets to at least 10 (8) jets in the 1L (2LOS) channel. These events are then further categorised according to the $b$-tagging requirements summarised in table 1. The requirements were chosen to provide good separation between the different flavour components of the associated jets in the $t\bar{t}+\text{jets}$ background. The 2b, 4b and $\geq 5b$ regions are defined by respectively requiring the presence of 2, 4 and at least 5 jets $b$-tagged at the 70% OP. The $\geq 5b$ regions in the 1L channel help constrain the modelling of $t\bar{t}+\geq 3b$. In the 2LOS channel, the 4b and $\geq 5b$ regions are merged into $\geq 4b$ regions because fewer events are expected. The events with 3 jets $b$-tagged at the 70% OP are further assigned to the 3bL, 3bH and 3bV regions using requirements on the number of jets $b$-tagged at the 60% and 85% OPs. The 3bL (3bH) regions are defined to have relatively lower (higher) purity of MC ‘truth’ $b$-jets amongst
Table 1. Summary of the $b$-tagging requirements for the event categorisation. Events in each category must satisfy all requirements listed in columns. $N_b^{60\%}$, $N_b^{70\%}$ and $N_b^{85\%}$ are defined as the numbers of $b$-tagged jets obtained using $b$-tagging operating points with average expected efficiencies of $60\%$, $70\%$ and $85\%$, respectively. The $3bL$ ($3bH$) requirement selects events with lower (higher) purity of MC ‘truth’ $b$-jets amongst the three jets tagged at the $70\%$ OP. The $3bV$ requirement is used to define the validation regions. The symbol ‘-' indicates that no requirement is applied.

<table>
<thead>
<tr>
<th>Name</th>
<th>$N_b^{60%}$</th>
<th>$N_b^{70%}$</th>
<th>$N_b^{85%}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2b</td>
<td>-</td>
<td>= 2</td>
<td>-</td>
</tr>
<tr>
<td>3bL</td>
<td>$\leq 2$</td>
<td>= 3</td>
<td>-</td>
</tr>
<tr>
<td>3bH</td>
<td>= 3</td>
<td>= 3</td>
<td>= 3</td>
</tr>
<tr>
<td>3bV</td>
<td>= 3</td>
<td>= 3</td>
<td>$\geq 4$</td>
</tr>
<tr>
<td>$\geq 4b$ (2LOS)</td>
<td>-</td>
<td>$\geq 4$</td>
<td>-</td>
</tr>
<tr>
<td>4b (1L)</td>
<td>-</td>
<td>= 4</td>
<td>-</td>
</tr>
<tr>
<td>$\geq 5b$ (1L)</td>
<td>-</td>
<td>$\geq 5$</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 2. Schematic view of the event categories used to select analysis regions (signal, control, validation and $tt+\text{jets}$ reweighting regions) in the 1L channel (left) and 2LOS channel (right). The axes represent the jet multiplicity and $b$-tagging requirements defined in table 1. The $3bL$ ($3bH$) $b$-tagging requirement selects events with lower (higher) purity of MC ‘truth’ $b$-jets amongst the three jets tagged at the $70\%$ OP. The $3bV$ $b$-tagging requirement is used to define the validation regions. The regions in grey are not used in the analysis.

As a result, the $3bH$ regions contain a larger fraction of $tt+b$ and $tt+B$ events, whereas the $3bL$ regions are more populated by $tt+\geq 1c$ and $tt+$light, where the third jet $b$-tagged at the $70\%$ OP is a mis-tagged $c$-jet or light-flavour jet. The $3bV$ regions are defined to be orthogonal to $3bL$ and $3bH$ regions and are used to validate the background modelling in events enriched in $tt+\geq 1b$.

$^3$MC ‘truth’ $b$-jets are defined as jets that are matched to $b$-hadrons.
Figure 3. Relative contribution from the signal and backgrounds in all signal, control and validation regions in the 1L channel (left) and 2LOS channel (right). The 3bL (3bH) b-tagging requirement selects events with lower (higher) purity of MC ‘truth’ b-jets amongst the three jets tagged at the 70% OP. The 3bV b-tagging requirement is used to define the validation regions. For the tt+jets background, the fraction is shown for each component with the finer classification. The ttll signal is normalised to the SM cross-section prediction.

A total of 21 regions are used in the profile likelihood fit, with 12 regions in the 1L channel and 9 regions in the 2LOS channel. They are defined by considering the regions with at least 8 (6) jets in the 1L (2LOS) channel and satisfying the 3bL, 3bH or ≥4b requirements. Among these regions, the ones that have at least 10 (8) jets or have 9 (7) jets and satisfy the ≥4b requirement in the 1L (2LOS) channel are defined as the signal regions. The rest of the fitted regions are defined as the control regions. A total of 6 validation regions are also defined by considering the regions with at least 8 (6) jets in the 1L (2LOS) channel and satisfying the 3bV requirement. The validation regions are not used in the fit and hence do not contribute to the signal extraction. The largest signal contamination in the validation regions is expected to be 4.4% in the (≥8j, 3bV) region in the 2LOS channel. A test found that including these regions in the fit would increase the sensitivity to the signal by 5%. However, this small gain in sensitivity is relinquished to ensure reliable background modelling. The tt+jets kinematic reweighting regions are defined by considering all the regions that satisfy the 2b requirement.

Figure 3 shows the background composition in each signal, control and validation region, as expected from the nominal tt+jets MC simulation after applying the corrections described in section 7. The contribution from the signal is included, assuming the cross section predicted by the SM. The control regions have a signal contamination of no more than 1%. The largest signal-to-background ratio evaluated from the inclusive yields in the signal regions is 6.1%.

7 Modelling of ttantijets background

This section describes the corrections applied to both the nominal tt+jets prediction and to the associated systematic uncertainties. The corrections applied to the tt+jets MC sample
include a rescaling of the $t\bar{t}$+jets flavour components and a sequential kinematic reweighting. Any possible residual mismodelling is accounted for by the systematic uncertainties in the profile likelihood fit used for the extraction of the signal strength.

7.1 $t\bar{t}$+jets flavour rescaling

The $t\bar{t}$+jets flavour rescaling adjusts the overall yields of the $t\bar{t}$+light, $t\bar{t}$+$\geq 1c$ and $t\bar{t}$+$\geq 1b$ categories. The rescaling factors are derived from a dedicated profile likelihood fit to data using the event yields in the regions defined by various $b$-tagging requirements. Events with $\geq 8j$ in the 1L channel and $\geq 6j$ in the 2LOS channel are assigned to 2b, 3bL, 3bH and $\geq 4b$ regions, using the same criteria as defined in table 1. The flavour rescaling fit exploits the different $t\bar{t}$+jets flavour fractions in the eight fitted regions. These regions have the same $N_{\text{jets}}$ requirements as the regions used in the profile likelihood fit to extract the signal. The largest signal-to-background ratio in these regions is 2.5%, estimated from MC simulation prior to the fit. Systematic uncertainties due to the tagging efficiency of $b$-jets and the mis-tag rate of $c$-jets and light-flavour jets are considered as nuisance parameters. The measured rescaling factors for $t\bar{t}$+light, $t\bar{t}$+$\geq 1c$ and $t\bar{t}$+$\geq 1b$ are $0.99 \pm 0.05$, $1.58 \pm 0.18$ and $1.33 \pm 0.06$, respectively, where the quoted uncertainties are from the statistical uncertainty of the data and from uncertainties in the $b$-tagging calibration.

The measured rescaling factors are applied to the $t\bar{t}$+jets MC sample entering the profile likelihood fit to extract the signal; however, their measured uncertainties are not used in the fit. Instead, large global normalisation uncertainties are assigned to $t\bar{t}$+HF events, as discussed in section 9.3. However, $t\bar{t}$+jets modelling uncertainties due to the choice of generator and scale also affect the flavour composition of the radiated jets. To reduce the correlation among these modelling uncertainties and the dedicated $t\bar{t}$+HF normalisation uncertainties, the $t\bar{t}$+jets flavour fraction in all modelling systematic variations is rescaled to match the nominal prediction after the flavour rescaling within the acceptance of the analysis, i.e. events in the 1L (2LOS) channel with $N_{\text{jets}} \geq 8$ (6) and $N_{b70\%} \geq 2$. The $t\bar{t}$+$\geq 1b$ events in the $t\bar{t}b\bar{b}$ 4FS sample are rescaled to have the same yield as the nominal $t\bar{t}$+$\geq 1b$ events within the acceptance. Such treatment ensures that the overall production rate of $t\bar{t}$+HF events is controlled by the global normalisation uncertainties, whilst the modelling systematic uncertainties affect only the relative changes in the yield across different analysis regions and the kinematic distributions within each region.

7.2 Sequential kinematic reweighting

Following the flavour rescaling, a sequential reweighting is used to mitigate the kinematic mismodelling observed in $t\bar{t}$+jets MC simulation. The reweighting corrects the distributions of $N_{\text{jets}}$, the number of large-$R$ jets ($N_{\text{R-jets}}$), the scalar sum of all jet and lepton $p_T$ in the event ($H_T^{\text{all}}$), and the average $\Delta R$ between any two jets ($\Delta R_{\text{avg}}^{\text{all}}$). These variables are related to the overall jet activity in the events and are observed to be mismodelled, especially the $N_{\text{jets}}$ and $H_T^{\text{all}}$ spectra. These variables capture the most representative global kinematics of the events, as well as kinematic properties of the individual jets such as their $p_T$ and angular distributions.
The $t\bar{t}+\text{jets}$ events in $\geq 3b$ regions are reweighted so the overall MC prediction matches the data in the $2b$ regions. In this procedure it is assumed that the radiation modelling deficiency in the parton shower is independent of the flavour of the radiated jets. Systematic variations of the $t\bar{t}+\text{jets}$ modelling cover possible deviations from this assumption.

The reweighting is performed in three steps, separately for the 1L and 2LOS channels. In the first step, the two-dimensional distribution of the numbers of jets and large-$R$ jets, $(N_{\text{jets}}, N_{\text{LR-jets}})$, is reweighted. The $N_{\text{jets}}$ spectrum is corrected up to $\geq 13$ ($\geq 11$) jets, while the $N_{\text{LR-jets}}$ spectrum is corrected up to $\geq 2$ ($\geq 1$) large-$R$ jets in the 1L (2LOS) channel. The reweighting factors are derived for each $(N_{\text{jets}}, N_{\text{LR-jets}})$ bin. The second step aims to correct for the mismodelling in the $H_T^{\text{all}}$ spectrum. The reweighting is performed on the reduced $H_T^{\text{all,red}}$ variable, defined as $H_T^{\text{all,red}} = H_T^{\text{all}} - (N_{\text{jets}} - N_{\text{min}}) \times 90 \text{ GeV}$, where $N_{\text{min}} = 7$ ($5$) in the 1L (2LOS) channel. The value of 90 GeV corresponds to the average $p_T$ of each additional jet, estimated as the difference between the most probable values of $H_T^{\text{all}}$ in consecutive $N_{\text{jets}}$ bins. The term subtracted from $H_T^{\text{all}}$ therefore removes the dependence of this variable on $N_{\text{jets}}$ and minimises the correlation between the first and the second reweighting steps. The reweighting factors are therefore derived in an inclusive region with $N_{\text{jets}} \geq 7$ ($5$) in the 1L (2LOS) channel, but separately for each $N_{\text{LR-jets}}$ multiplicity bin, from 0 to $\geq 2$ ($\geq 1$) large-$R$ jets. The events with 7 ($5$) jets in the 1L (2LOS) channel are included in this step to increase the statistical precision of the reweighting factors in the highest $N_{\text{LR-jets}}$ multiplicity bin. These events are not used elsewhere in the analysis. To further reduce statistical fluctuations in the binned reweighting factors of $H_T^{\text{all,red}}$, a three-parameter exponential function, $f(H_T^{\text{all,red}}) = a + b \cdot \exp(c \cdot H_T^{\text{all,red}})$, is used to fit the factors in each $N_{\text{LR-jets}}$ multiplicity bin. The final step corrects for the residual mismodelling in the angular distributions by reweighting the binned distribution of $\Delta R_{\text{avg}}$, the average $\Delta R$ between any two jets. This is performed in each $(N_{\text{jets}}, N_{\text{LR-jets}})$ bin, up to $\geq 10$ ($\geq 8$) jets and $\geq 2$ ($\geq 1$) large-$R$ jets in the 1L (2LOS) channel.

Uncertainties in the derived reweighting factors are propagated as systematic uncertainties in the $t\bar{t}+\text{jets}$ background. The uncertainty sources include the limited numbers of data and MC events and the cross-section uncertainties of the non-$t\bar{t}$ processes. An additional uncertainty is considered for the $H_T^{\text{all,red}}$ reweighting step by replacing the exponential function with a reciprocal function. A more detailed discussion of these uncertainties is presented in section 9.3.

The reweighting factors are also derived for each systematic uncertainty affecting the $t\bar{t}+\text{jets}$ prediction, such that in the $2b$ regions all systematic variations match the nominal prediction and thus the data. After applying the reweighting, the overall systematic uncertainty in the $\geq 3b$ regions is reduced.

### 7.3 Effects of the corrections to $t\bar{t}+\text{jets}$

Figure 4 illustrates the combined effect of the flavour rescaling and the sequential reweighting. The $N_{\text{jets}}$ and $H_T^{\text{all}}$ distributions for events with $\geq 8$ jets and $\geq 3$ $b$-jets in the 1L channel are shown as examples. Significant mismodelling is present in both distributions before the corrections. A large improvement in the level of agreement between data and MC simulation is achieved for both the overall normalisation and the kinematic shape. The
Figure 4. The $N_{\text{jets}}$ and $H_{\text{T}}$ distributions in the region with $\geq 8$ jets and $\geq 3$ $b$-jets in the 1L channel before (left) and after (right) the flavour rescaling and the sequential kinematic reweighting. The band includes the total pre-fit uncertainty of the MC prediction. The ratio of the data to the total MC expectation is shown in the lower panel. The last bin in all distributions includes the overflow.

flavour rescaling factors increase the overall normalisation of $t\bar{t}+\geq 1c$ and $t\bar{t}+\geq 1b$, which dominate the deficiency of MC events in the $\geq$3$b$ regions. The sequential reweighting increases the jet multiplicity while decreasing the total energy of the event and the average distance between the jets predicted by the $t\bar{t}$+jets simulation. The total uncertainty in the MC prediction is also reduced after the sequential reweighting is applied coherently to the $t\bar{t}$+jets systematic variations. The level of agreement after the reweighting of $\Delta R_{\text{avg}}$ is equally good, though the effect of the correction is smaller.
8 Signal discrimination

A multivariate analysis is performed in the signal regions to discriminate \( t\bar{t}t\bar{t} \) signal from the large background. It uses boosted decision trees (BDTs) \[75\] combining several input observables to build an output score maximising the separation between signal and the total background. The LO \( t\bar{t}t\bar{t} \) simulated signal sample is used in the training, while the BDT performance is evaluated on the NLO \( t\bar{t}t\bar{t} \) sample. For the distributions of kinematic variables used by the multivariate discriminant, good agreement was observed when comparing them in LO and NLO simulation. The background sample is composed of the simulated \( t\bar{t}+jets \) events after the data-driven corrections are applied and the non-\( t\bar{t} \) backgrounds as predicted by the simulation.

The BDTs are trained in six regions using fourteen variables. The six regions are defined by events with at least three \( b \)-tagged jets and 8 (6), 9 (7) and at least 10 (8) jets in the 1L (2LOS) channel. The BDTs use global event variables, the kinematic properties of reconstructed objects and pairs of objects, jet \( b \)-tagging information, the multiplicity and substructure of large-\( R \) jets, and the \( E_T^{\text{miss}} \). The most powerful discriminating variable in all regions is the sum of the pseudo-continuous \( b \)-tagging score over the six jets with the highest score in the event. The definition of the pseudo-continuous \( b \)-tagging score can be found in section 3. In the highest \( N_{\text{jets}} \) regions in the two channels, where an inclusive requirement on the number of jets is applied, the jet count is an important discriminating variable. The minimum distance \( \Delta R \) among all pairs of \( b \)-tagged jets also provides good discrimination in the majority of regions since the spatial separation of \( b \)-jets in the \( t\bar{t}t\bar{t} \) signal is larger on average than in the main \( t\bar{t}+\geq 1b \) background. Other variables include \( H_T^{\text{all}} \) calculated using all reconstructed jets and leptons in the event, centrality \( \sum_i p_T^i / \sum_i E_t^i \), where the sum runs over all reconstructed jets and leptons in the event, leading jet \( p_T \), minimum \( \Delta R \) among all pairs of \( b \)-tagged jets and leptons, average \( \Delta R \) between all pairs of jets, the invariant mass of the triplet of jets that has the smallest \( \Delta R \), the \( E_T^{\text{miss}} \), and the transverse mass of the \( W \) boson, \( m_T(\ell, E_T^{\text{miss}}) \), in the single-lepton channel.\(^5\) Additional variables are related to large-\( R \) jets: the number of large-\( R \) jets with a mass above 100 GeV, the sum of the first \( k_t \) splitting scale \( d_{12} \) of all large-\( R \) jets,\(^6\) and the sum of the second \( k_t \) splitting scale \( d_{23} \) of all large-\( R \) jets.

The modelling of the input variables was checked in the control and validation regions before and after the fit by propagating the fitted parameters obtained from the fit in the control and signal regions to the validation regions. Figure 5 shows the modelling of the sum of the pseudo-continuous \( b \)-tagging score in each lepton channel before performing the fit. The distribution in each lepton channel is shown for an inclusive selection of at least three \( b \)-tagged jets and at least 9 (7) jets in the 1L (2LOS) channel. These regions include

\(^4\)The \( \Delta R \) of a triplet of jets is defined as \( \Delta R_{ijk} = \sqrt{\Delta R_{ij}^2 + \Delta R_{ik}^2 + \Delta R_{jk}^2} \), where \( i, j, k \) are the indices of the three jets.

\(^5\)The transverse mass is defined as \( \sqrt{2p_T E_T^{\text{miss}} (1 - \cos \Delta \phi)} \), where \( \Delta \phi \) is the azimuthal angle between the lepton and \( E_T^{\text{miss}} \).

\(^6\)The \( k_t \) splitting scale \( d_{ij} \) is defined as the recombination distance between the jet constituents from a \( k_t \) algorithm with radius parameter \( R \): \( d_{ij} = \min(p_{T i}^2, p_{T j}^2) \times \Delta R_{ij}^2 / R^2 \).
Figure 5. Pre-fit comparison between data and prediction for the distributions of the sum of the pseudo-continuous $b$-tagging score over the six jets with the highest score in the event for the 1L channel (left) and the 2LOS channel (right) in the regions with $\geq 3$ $b$-jets and $\geq 9$ (7) jets in the 1L (2LOS) channel. The $t\bar{t}$+jets background is corrected using data. The band includes the total uncertainty of the pre-fit computation. The dashed red line shows the signal distribution normalised to the background yield. The ratio of the data to the total pre-fit expectation is shown in the lower panel. The last bin contains overflow events.

events from the validation, control and signal regions where the BDT score distribution is used. Taking into account all uncertainties, no significant discrepancy between data and the predicted background was found.

9 Systematic uncertainties

Various sources of systematic uncertainty affect the estimated signal and background rates, including those related to the luminosity, the identification and reconstruction of the physics objects, referred to as experimental uncertainties, and the modelling of the signal and background processes. In the following, a brief description of the sources of systematic uncertainty is provided. Particular emphasis is placed on the uncertainties related to the $t\bar{t}$ background prediction, which has the largest impact on the sensitivity of the measurement.

9.1 Experimental uncertainties

The experimental uncertainties are the same as in ref. [13] and are briefly summarised here. The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [20], obtained using the LUCID-2 detector [21] for the primary luminosity measurements. An uncertainty in the pile-up simulation is derived from the uncertainty in the scale factors used to adjust the MC pile-up to the data pile-up profile. Uncertainties in the modelling of electrons and muons in MC simulation arise from their momentum scale calibration and resolution, as
well as trigger, reconstruction, identification, and isolation efficiencies. Uncertainties in the modelling of jets are primarily related to their energy scale (JES) and resolution (JER). The JES uncertainty is decomposed into a set of 30 uncorrelated components referred to as eigenvectors (EV), with contributions from pile-up, jet flavour composition and single-particle response [76]. The JER uncertainty is represented by nine components [77]. An uncertainty in the efficiency to pass the JVT requirement for suppressing pile-up jets is also considered [32]. The $b$-tagging efficiencies and mis-tagging rates are measured in data using the same methods as those described in refs. [34, 78, 79]. The uncertainties in the $b$-tagging calibration are determined separately for $b$-jets, $c$-jets and light-flavour jets. They account for differences between data and simulation, depending on $p_T$ for $b$- and $c$-jets, and on $p_T$ and $\eta$ for light-flavour jets, and are divided into 85 components (45 for $b$-jets, 20 for $c$-jets and 20 for light-flavour jets). The $E_{\text{miss}}$ uncertainty due to a possible miscalibration of its soft-track component is derived from data-MC comparisons of the $p_T$ balance between the hard and soft $E_{\text{miss}}$ components [37].

9.2 Signal modelling uncertainties

Various sources of modelling uncertainty are considered for the $t\bar{t}t\bar{t}$ signal. An uncertainty related to the missing higher-order QCD corrections is estimated by varying the renormalisation ($\mu_r$) and the factorisation ($\mu_f$) scales simultaneously by factors of 2.0 and 0.5 relative to the nominal value. The effect of the PDF variations on the signal MC prediction was evaluated following the PDF4LHC prescription [80] and found to be negligible. The uncertainty due to the choice of parton shower and hadronisation model is estimated by comparing the nominal $t\bar{t}t\bar{t}$ sample with the alternative that uses HERWIG 7 to simulate the PS, as described in section 4.

9.3 Uncertainties in the $t\bar{t}+\text{jets}$ background

The systematic uncertainties affecting the $t\bar{t}+\text{jets}$ background modelling are summarised in table 2. They are applied to each flavour component of the $t\bar{t}$ background separately, i.e. treated as uncorrelated, to account for the variation in flavour composition of the regions included in the fit and for possible differences in the modelling of the $t\bar{t}+\geq 1b$, $t\bar{t}+\geq 1c$, and $t\bar{t}+\text{light}$ processes.

An uncertainty of 50% in the normalisation of $t\bar{t}+\geq 1c$ events as well as of the different subcategories of the $t\bar{t}+\geq 1b$ events ($t\bar{t}+b$, $t\bar{t}+b\bar{b}$, $t\bar{t}+B$, and $t\bar{t}+\geq 3b$) is applied [72]. The uncertainties due to the choice of generator and PS model used to simulate the inclusive $t\bar{t}$ sample are evaluated by comparing the nominal $t\bar{t}$ sample with the alternative $t\bar{t}$ samples, detailed in section 4. They are split into shape and migration components, with the former affecting only the shape of the distributions in each fit region and the latter changing the yields in different regions. These uncertainties are applied to each of the $t\bar{t}+\geq 1b$ subcategories as detailed in table 2.

Uncertainties due to missing higher-order QCD corrections are estimated by separately varying the renormalisation and the factorisation scales by factors of 2.0 and 0.5 in the nominal $t\bar{t}$ sample. Additionally, uncertainties in the amounts of initial- and final-state radiation (ISR and FSR) from the PS are assessed by respectively varying the corresponding
2b regions which is subtracted before deriving the reweighting. The cross sections of these

The difference compared to the exponential function is smaller than 20% in any fitted

exponential function with a three-parameter reciprocal function,

the two channels. Five additional systematic uncertainties are obtained by replacing the ex-

contour of the nominal fit to the exponential function, resulting in five systematic uncer-

numbers of data and MC events in each of the reweighted bins of the (\(N_{\text{jets}}, N_{\text{LR-jets}}\) and \(\Delta R_{\text{avg}}\), variables result in 120 systematic uncertainties. Their effect on \(H_T^{\text{all}}\) and the BDT
discriminants is less than 1% in any bin of the corresponding distribution.

The uncertainty in the \(H_T^{\text{all,red}}\) reweighting is evaluated using a 68% confidence interval
contour of the nominal fit to the exponential function, resulting in five systematic uncer-
tainties corresponding to the five fitted functions in each of the \(N_{\text{LR-jets}}\) multiplicity bins in the
two channels. Five additional systematic uncertainties are obtained by replacing the ex-
ponential function with a three-parameter reciprocal function, \(f(H_T^{\text{all,red}}) = a + b H_T^{\text{all,red}}\). The

difference compared to the exponential function is smaller than 20% in any fitted
regions, with a larger deviation at high or low \(H_T^{\text{all,red}}\). The net effect of the systematic
uncertainties in the \(H_T^{\text{all,red}}\) reweighting is below 5% across the \(H_T^{\text{all}}\) and BDT distributions.

The reweighting factors depend on the contribution from the non-\(t\bar{t}\) processes in the
2b regions which is subtracted before deriving the reweighting. The cross sections of these
processes are varied by ±50% simultaneously, and the reweighting factors are rederived.
The resulting difference is taken as an additional systematic uncertainty, ranging from 3% to 6%, in the reweighting factors.

\footnotetext[7]{The A14 tune’s variations of \(a^s_{\text{ISR}}(m_Z)\) are usually referred to as Var3c, and correspond to \(a^s_{\text{ISR}}(m_Z) = 0.127 - 0.012\) [45].}

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>Description</th>
<th>Components (number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t\bar{t}+\geq 1b) normalisation</td>
<td>±50%</td>
<td>(t\bar{t}+b, t\bar{t}+bb, t\bar{t}+B, t\bar{t}+\geq 3b) (4)</td>
</tr>
<tr>
<td>(t\bar{t}+\geq 1c) normalisation</td>
<td>±50%</td>
<td>(t\bar{t}+\geq 1c) (1)</td>
</tr>
<tr>
<td>Generator choice</td>
<td>Powheg vs MadGraph5_aMC@NLO</td>
<td>((t\bar{t}+\text{light}, t\bar{t}+\geq 1c, t\bar{t}+b, t\bar{t}+bb, t\bar{t}+B, t\bar{t}+\geq 3b) (\otimes ) (shape, migration) (12))</td>
</tr>
<tr>
<td>PS choice</td>
<td>Pythia 8 vs Herwig 7</td>
<td>((t\bar{t}+\text{light}, t\bar{t}+\geq 1c, t\bar{t}+b, t\bar{t}+bb, t\bar{t}+B, t\bar{t}+\geq 3b) (\otimes ) (shape, migration) (12))</td>
</tr>
<tr>
<td>Renormalisation scale</td>
<td>Varying (\mu_r) in Powheg</td>
<td>(t\bar{t}+\text{light, } t\bar{t}+\geq 1c, t\bar{t}+\geq 1b) (3)</td>
</tr>
<tr>
<td>Factorisation scale</td>
<td>Varying (\mu_f) in Powheg</td>
<td>(t\bar{t}+\text{light, } t\bar{t}+\geq 1c, t\bar{t}+\geq 1b) (3)</td>
</tr>
<tr>
<td>ISR</td>
<td>Varying (a_{\text{ISR}}^s) (PS) in Pythia 8</td>
<td>(t\bar{t}+\text{light, } t\bar{t}+\geq 1c, t\bar{t}+\geq 1b) (3)</td>
</tr>
<tr>
<td>FSR</td>
<td>Varying (\mu_t) (PS) in Pythia 8</td>
<td>(t\bar{t}+\text{light, } t\bar{t}+\geq 1c, t\bar{t}+\geq 1b) (3)</td>
</tr>
<tr>
<td>5FS vs 4FS</td>
<td>PowhegBoxRes (4FS) vs PowhegBox (5FS)</td>
<td>(t\bar{t}+b, t\bar{t}+bb, t\bar{t}+B, t\bar{t}+\geq 3b) (4)</td>
</tr>
</tbody>
</table>

Table 2. Summary of the sources of systematic uncertainty for the \(t\bar{t}+\text{jets}\) modelling. The last
column of the table lists the uncorrelated components of each systematic uncertainty source. All
systematic uncertainty sources are treated as uncorrelated across the \(t\bar{t}+\text{jets}\) components. For
generator and PS choices, each \(t\bar{t}+\text{jets}\) component is further decomposed into a shape component and a migration component. The number of uncorrelated components for each physics source is shown in parentheses.
9.4 Modelling uncertainties in non-$t\bar{t}+\text{jets}$ backgrounds

An uncertainty of 5.5\% in the total cross section of the three single-top-quark production modes is included [81–83]. Similar to the $t\bar{t}$ background, uncertainties associated with the PS and hadronisation model and the generator choice are evaluated by comparing the nominal POWHEG + PYTHIA 8 sample for each process with alternative samples produced with POWHEG + HERWIG 7 and MadGraph5_aMC@NLO + PYTHIA 8. The uncertainty due to the treatment of the interference between $tW$ and $t\bar{t}$ production at NLO is evaluated by comparing the nominal $tW$ sample produced using the diagram removal scheme with an alternative sample produced with the same generator but using the diagram subtraction scheme [54].

Modelling uncertainties in the $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$ processes are evaluated in a similar way. Uncertainties of 60\%, 15\%, and 20\% are applied to the $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$ cross sections, respectively [13, 84, 85]. The PDF uncertainty in the $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$ processes is negligible compared to the uncertainty applied to their cross section. The uncertainty due to the choice of generator and PS is assessed by comparing the nominal simulation of each of these processes with the alternative one described in section 4. Additional uncertainties of 10\%, 20\% and 30\% for events with respectively 9 (7), 10 (8) and at least 11 (9) jets are applied to account for the production of additional jets in these processes in the 1L (2LOS) channel. These uncertainty values are based on the level of mismodelling of the $t\bar{t}+\text{jets}$ $N_{\text{jets}}$ distribution observed in the 2b region. They are also consistent with the modelling uncertainties evaluated in the 2LSS/3L final state for these backgrounds using a full set of MC modelling uncertainties [13].

An uncertainty of 60\% is assumed for the $V+\text{jets}$ production cross section. It is estimated by adding a 24\% uncertainty in quadrature for each additional jet based on a comparison among different algorithms for merging LO matrix elements and parton showers [86].

For each of the other small background processes a conservative cross-section uncertainty of 50\% is applied.

10 Results

The analysis uses a binned profile likelihood fit to extract the $t\bar{t}t\bar{t}$ signal strength with systematic uncertainties parameterised as constrained nuisance parameters. The regions used in the fit are described in section 6. In all control regions, the binned $H_{\text{T}}^{\text{all}}$ distribution is used as the input to the profile likelihood fit, except for the (8j, $\geq 5b$) region in the 1L channel where the event count is used due to the small number of events and the low signal-to-background ratio in this region. To improve the sensitivity to the $t\bar{t}t\bar{t}$ signal, the distribution of the BDT score described in section 8 is used in the signal regions. Each bin in these regions is represented by a Poisson probability term for the observed data, with the expectation value provided by the MC prediction with the corrections to $t\bar{t}+\text{jets}$. The expected yield depends on the signal strength, $\mu$, defined as the ratio of the measured
The production cross section to that predicted by the SM, $\sigma_{t\bar{t}t\bar{t}}/\sigma_{t\bar{t}}^{\text{SM}}$, and the systematic uncertainties, $\theta$. A binned likelihood function, $L(\mu, \theta)$, is constructed as the product of the Poisson probability terms of all fitted bins. Systematic uncertainties are taken into account in the likelihood function, each as a nuisance parameter constrained by a Gaussian prior. The uncertainties due to the limited number of generated MC events, also referred to as the MC statistical uncertainties, are implemented as additional nuisance parameters with Poisson priors, one for each fitted bin. The MC prediction is fitted to the data by adjusting $\mu$ and $\theta$ simultaneously to maximise the likelihood function. The statistical inferences of the fit results are derived using the likelihood ratio, $\lambda(\mu) = L(\mu, \hat{\theta}) / L(\tilde{\mu}, \tilde{\theta})$, where $\hat{\mu}$ and $\tilde{\theta}$ are the values of $\mu$ and $\theta$ that maximise the likelihood function and $\tilde{\theta}$ represents the values that maximise the likelihood function given a specific $\mu$ [87]. The uncertainty in the best-fit $\mu$ value is determined by finding the $\mu$ values that correspond to $-2 \ln \lambda(\mu) = 1$. The test statistic $q_0 = -2 \ln \lambda(0)$ is used to check the compatibility of the background-only hypothesis ($\mu = 0$) with the observed data, and consequently the significance of the fitted signal. The statistical analysis is performed using the RooStat framework [88, 89].

Figure 6 shows the observed event yield compared with the prediction in each control and signal region in each lepton channel before the fit (‘pre-fit’) and after the fit (‘post-fit’). The level of agreement between data and the prediction is improved post-fit due to the adjustment of the nuisance parameters. The overall uncertainty in the prediction is also significantly reduced as a result of the constraints from the data and the correlations among the nuisance parameters. Figures 7 and 8 show the distribution of the BDT score in the signal regions in both lepton channels. Good agreement is observed between data and the post-fit prediction in all fitted regions. The goodness of fit was evaluated using a likelihood-ratio test, comparing the likelihood value from the nominal fit with the one obtained from a saturated model, built with one free-floating normalisation factor for each fitted bin [90]. The probability that the fit model is compatible with the observed data is 59%. The post-fit background modelling is further checked in validation regions that are not used in the fit. The definition of these regions is described in section 6. Figure 9 shows the distribution of the BDT score in four of those regions.

The best-fit value for the signal strength is

$$\mu = 2.2 \pm 0.7 \, \text{(stat.)} \pm^{+1.5}_{-1.0} \, \text{(syst.)} = 2.2 \pm^{+1.6}_{-1.2},$$

which takes into account the additional 20% uncertainty in $\sigma_{t\bar{t}t\bar{t}}^{\text{SM}}$ in the denominator of the $\mu$ parameter. In order to check the compatibility of the 1L and 2LOS channels, an alternative fit is performed with two independent signal-strength parameters, one for each of the two channels. The best-fit values of the two signal strengths are $2.9 \pm^{+1.6}_{-1.2}$ for the 1L channel and $1.3 \pm^{+1.7}_{-1.5}$ for the 2LOS channel. The probability of obtaining a discrepancy between these two signal-strength parameters equal to or larger than the one observed is 39%. The compatibility between the signal regions was also checked by performing a fit using a different signal-strength parameter in each region. The fitted values are found to be consistent within uncertainties.
Figure 6. Comparison of predicted and observed event yields in each control and signal region in the 1L channel (top) and in the 2LOS channel (bottom) before the fit (left) and after the fit (right). The $tt+\text{jets}$ background is corrected at the pre-fit level using data. The band includes the total pre- or post-fit uncertainties. The dashed red line shows the signal distribution normalised to the total background yield. The ratio of the data to the total pre- or post-fit prediction is shown in the lower panel.

After excluding the uncertainty in the SM $tt\bar{t}t$ cross section, the $tt\bar{t}t$ production cross section is measured to be

$$\sigma_{tt\bar{t}t} = 26 \pm 8 \, \text{(stat.)} \pm 15 \, \text{(syst.)} \, \text{fb} = 26^{+17}_{-13} \, \text{fb}.$$ 

The statistical uncertainty is obtained from a fit where all nuisance parameters are fixed to their post-fit values. The systematic uncertainty is determined by subtracting in quadrature the statistical uncertainty from the total uncertainty. The measured cross section is consistent with the SM predicted cross section of $\sigma_{tt\bar{t}t}^{\text{SM}} = 12.0 \pm 2.4 \, \text{fb}$ computed at NLO in QCD including NLO electroweak corrections [12]. The observed significance of $tt\bar{t}t$ production relative to the background-only prediction is 1.9 standard deviations, while 1.0 standard deviations is expected. The observed (expected) $p$-value is 3% (17%), corresponding to the probability of obtaining a result at least as signal-like as observed if no signal is present.

Figure 10 shows the event yield in data compared with the post-fit signal ($S$) and total background ($B$) prediction. It is ordered by the signal-to-background ratio of the discriminant bins from all fitted regions. The predictions are shown for the best-fit signal
Figure 7. Comparison between data and post-fit prediction for the distributions of the BDT score in each signal region in the 1L channel. The band includes the total uncertainty of the post-fit computation. The dashed red line shows the signal distribution normalised to the background yield. The ratio of the data to the total post-fit computation is shown in the lower panel.
strength and for the SM prediction. The observed data shows a signal-like excess over the background, with a higher probability of compatibility with the best-fit signal strength of \( \mu = 2.2 \) than with the SM prediction of \( \mu = 1.0 \) in the high \( \log_{10}(S/B) \) region.

Table 3 lists the contributions from the various groups of systematic uncertainties to the fitted uncertainty in the measured \( tt\bar{t}t \) production cross section, together with the total systematic uncertainty, the statistical uncertainty and the total uncertainty. The impacts of the individual uncertainties on the best-fit value of \( \mu \) are presented in figure 11, ranked according to their post-fit impacts. The 20 systematic uncertainties with the largest
Figure 9. Comparison between data and post-fit prediction for the distributions of the BDT score in each validation region. The band includes the total uncertainty of the post-fit computation. The dashed red line shows the signal distribution normalised to the background yield. The ratio of the data to the total post-fit computation is shown in the lower panel.

Impact are shown. The impacts of the individual and grouped systematic uncertainties are evaluated using different methods, as described in the corresponding captions.

Apart from the uncertainties in the $t\bar{t}t\bar{t}$ signal, the largest contribution is from $t\bar{t}+\geq 1b$ modelling uncertainties. The major uncertainties in this category are due to the choice of flavour scheme and MC generator used to model the $t\bar{t}+b\bar{b}$ background, which is dominant in all signal regions. This reflects the large difference between the different predictions, which is expected given our limited knowledge of the phase space being probed. The contribution from $t\bar{t}+\geq 1c$ modelling uncertainties is dominated by the $t\bar{t}+\geq 1c$ cross-section.
uncertainty. Although the fraction of $t\bar{t}+\geq 1c$ events is less than 10% in the signal regions, this uncertainty acquires a large anti-correlation with the $t\bar{t}+b\bar{b}$ cross-section uncertainty because of the interplay between $t\bar{t}+\geq 1c$ and $t\bar{t}+b\bar{b}$ in the 3b regions. The $t\bar{t}+\geq 1c$ cross-section uncertainty is constrained by 50% due to the substantial impact of this uncertainty on the regions with loose $b$-tagging requirements, where abundant data is available to provide the constraint. The $t\bar{t}+b\bar{b}$ cross-section uncertainty does not have a large post-fit impact on $\mu$ by itself. This is because it is constrained by 70% relative to its pre-fit uncertainty due to the high purity of $t\bar{t}+b\bar{b}$ and the sufficient amount of data in the $\geq 4b$ regions. The uncertainties in the jet energy scale and resolution also contribute substantially to the overall uncertainty. This is due to the high $N_{\text{jets}}$ multiplicity feature of the $tt\bar{t}\bar{t}$ signal. The uncertainties in the $b$-tagging efficiencies are also an important source. The most important contribution is from ‘light jets mis-tag rates EV0’, which is the leading component in the 20 uncertainties related to the mis-tag rates of light-flavour jets. This component mainly affects the migration of the different $t\bar{t}+\text{jets}$ components across the regions defined using different $b$-tagging requirements, subsequently affecting the $t\bar{t}+b\bar{b}$ yields in the signal regions, which are highly correlated with the fitted signal strength.

**Figure 10.** Observed and expected event yields as a function of $\log_{10}(S/B)$, where $S$ and $B$ are the post-fit signal and total background yields, respectively. The bins in all fitted regions are ordered and grouped in bins of $\log_{10}(S/B)$. The signal is shown for both the best-fit signal strength, $\mu = 2.2$, and the SM prediction, $\mu = 1.0$. The lower panel shows the ratio of the data to the post-fit background prediction, compared with the signal-plus-background prediction with the best-fit signal strength and the SM prediction. The shaded band represents the total uncertainty in the background prediction.
Figure 11 also illustrates the best-fit values of the nuisance parameters corresponding to the 20 most important systematic uncertainties. A number of nuisance parameters are pulled from their nominal value in the likelihood fit. However, all observed pulls are smaller than 50\% of the corresponding prior uncertainty. The pull on the ‘$t\bar{t}t\bar{t}$ PS choice’ is driven by the discrepancy between data and MC simulation in the 2LOS channel, ($\geq 8j$, $\geq 4b$) region. The slight pull on this nuisance parameter reduces the $t\bar{t}t\bar{t}$ yield in this region by 3\%. The effects in other fitted regions are either comparable or smaller. The ‘light jets mis-tag rates EV0’ nuisance parameter is pulled to mitigate the discrepancy between data and prediction by adjusting the relative normalisation of the different $t\bar{t}$+jets components across the fitted regions. All other pulls, including those not shown in figure 11, are mainly on $t\bar{t}$+jets modelling uncertainties. The combination of these pulls is found to correct for the mismodelling in the shape and normalisation of the $t\bar{t}$+jets background prediction. Similar pulls are observed in a background-only fit when removing the bins with a signal-to-background ratio larger than 5\%. Additional tests were performed with pseudo-data sets constructed using various alternative $t\bar{t}$+jets predictions and injected signal. No bias in the extracted $\mu$ was seen in the fits to these pseudo-data sets.

11 Combination with the same-sign dilepton and multilepton final state

The measurement of $t\bar{t}t\bar{t}$ production in the 1L/2LOS final state presented in this article is combined with the ATLAS measurement in the 2LSS/3L final state [13] using the same data set. The combination is performed via a simultaneous profile likelihood fit in all regions of both analyses, with all systematic uncertainties included, and the combined $t\bar{t}t\bar{t}$ production cross section is extracted. The events in the two analyses are statistically independent by construction due to the different lepton selection criteria.

The systematic uncertainties have a significant impact on the sensitivity of both the 1L/2LOS and 2LSS/3L analyses. However, because the most relevant systematic uncertainties in these final states are different, the assumptions about the correlations of the uncertainties have a negligible effect on the combined result.

The uncertainties related to the modelling of the $t\bar{t}$+jets background are treated as uncorrelated between the 1L/2LOS and 2LSS/3L final states. The contribution of $t\bar{t}$+jets events in the 2LSS/3L final state is much smaller than that in the 1L/2LOS final state, and arises from misreconstructed or non-prompt leptons, while it is the dominant irreducible background in the 1L/2LOS final state. Its estimation and the treatment of relevant uncertainties is therefore different in the two final states. The experimental uncertainties are treated as fully correlated between the two final states since both analyses use the same reconstructed objects and data set. Theoretical modelling uncertainties in the non-$t\bar{t}$ backgrounds and the $t\bar{t}t\bar{t}$ signal are also fully correlated except for the normalisation of the $t\bar{t}W$ background, which is treated differently in the 1L/2LOS and 2LSS/3L final states. In the latter, $t\bar{t}W$ is the most important background and its normalisation is a free parameter of the fit, while in the former it is a small background and its normalisation is taken from MC simulation with a large uncertainty motivated by the normalisation factor obtained in the 2LSS/3L analysis. For this reason, the $t\bar{t}W$ normalisation is uncorrelated between the
<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>$\Delta \sigma_{\bar{t}t\bar{t}}$ [fb]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Signal Modelling</strong></td>
<td></td>
</tr>
<tr>
<td>$\bar{t}t \bar{t}t$ modelling</td>
<td>+8</td>
</tr>
<tr>
<td><strong>Background Modelling</strong></td>
<td></td>
</tr>
<tr>
<td>$\bar{t}t + \geq 1b$ modelling</td>
<td>+8</td>
</tr>
<tr>
<td>$\bar{t}t + \geq 1c$ modelling</td>
<td>+5</td>
</tr>
<tr>
<td>$\bar{t}t + \text{jets reweighting}$</td>
<td>+4</td>
</tr>
<tr>
<td>Other background modelling</td>
<td>+4</td>
</tr>
<tr>
<td>$\bar{t}t + \text{light modelling}$</td>
<td>+2</td>
</tr>
<tr>
<td><strong>Experimental</strong></td>
<td></td>
</tr>
<tr>
<td>Jet energy scale and resolution</td>
<td>+6</td>
</tr>
<tr>
<td>$b$-tagging efficiency and mis-tag rates</td>
<td>+4</td>
</tr>
<tr>
<td>MC statistical uncertainties</td>
<td>+2</td>
</tr>
<tr>
<td>Luminosity</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Other uncertainties</td>
<td>&lt; 1</td>
</tr>
<tr>
<td><strong>Total systematic uncertainty</strong></td>
<td>+15</td>
</tr>
<tr>
<td><strong>Statistical uncertainty</strong></td>
<td>+8</td>
</tr>
<tr>
<td><strong>Total uncertainty</strong></td>
<td>+17</td>
</tr>
</tbody>
</table>

**Table 3.** The contribution from different systematic uncertainties to the measured $\bar{t}t\bar{t}t$ production cross section, $\sigma_{\bar{t}t\bar{t}t}$, grouped into categories. For each uncertainty source, the fit is repeated with the corresponding group of nuisance parameters fixed to their best-fit values, $\hat{\theta}$. The contribution from each source, $\Delta \sigma_{\bar{t}t\bar{t}t}$, is then evaluated by subtracting in quadrature the uncertainty in $\sigma_{\bar{t}t\bar{t}t}$ obtained in this fit from that of the full fit. The contributions from individual groups are compared with the total systematic uncertainty and the statistical uncertainty. The total systematic uncertainty is different from the sum in quadrature of the different groups due to correlations among nuisance parameters in the fit.
Figure 11. The nuisance parameters ranked according to their post-fit impacts on the best-fit value of $\mu$. Only the 20 nuisance parameters with the largest impacts are shown. The empty (solid) blue rectangles illustrate the pre-fit (post-fit) impacts on the parameter of interest $\mu$, corresponding to the top axis. The pre-fit (post-fit) impact of each nuisance parameter, $\Delta \mu$, is calculated as the difference in the fitted value of $\mu$ between the nominal fit and the fit when fixing the corresponding nuisance parameter to $\hat{\theta} \pm \Delta \hat{\theta}$ ($\hat{\theta} \pm \Delta \hat{\theta}$), where $\hat{\theta}$ is the best-fit value of the nuisance parameter and $\Delta \theta$ ($\Delta \hat{\theta}$) is its pre-fit (post-fit) uncertainty. The black points show the best-fit values of the nuisance parameters, with the error bars representing the post-fit uncertainties. Each nuisance parameter is shown relative to its nominal value, $\theta_0$, and in units of its pre-fit uncertainty.
two final states. The systematic uncertainties associated with the reducible backgrounds in the 2LSS/3L final state and uncertainties related to the data-driven corrections to the $t\bar{t}+\text{jets}$ background in the 1L/2LOS final state are uncorrelated.

The $t\bar{t}t\bar{t}$ signal strength from the combination of the two final states is measured to be

$$\mu = 2.0 \pm 0.4 \text{ (stat.)} \pm 0.7 \text{ (syst.)} = 2.0 \pm 0.6.$$  

The best-fit $t\bar{t}t\bar{t}$ signal strengths for the 1L/2LOS and 2LSS/3L final states fitted separately, and for the combination of the two final states, are summarised in figure 12. The fitted $\mu$ values are consistent between the two final states, which are dominated by different backgrounds and different uncertainties. The combined SM $t\bar{t}t\bar{t}$ production cross section is measured to be

$$\sigma_{t\bar{t}t\bar{t}} = 24 \pm 4 \text{ (stat.)} \pm 5 \text{ (syst.)} \text{ fb} = 24 \pm 7 \text{ fb.}$$

The measured cross section is consistent within 2.0 standard deviations with the SM prediction of $\sigma_{t\bar{t}t\bar{t}}^{\text{SM}} = 12.0 \pm 2.4 \text{ fb}$ computed at NLO in QCD including NLO electroweak corrections [12]. The observed (expected) significance of the combined measured $t\bar{t}t\bar{t}$ signal is $4.7\sigma$ ($2.6\sigma$) above the background-only hypothesis. The significance is improved compared to the observed (expected) significance of $4.3\sigma$ ($2.4\sigma$) obtained in the 2LSS/3L analysis.

12 Conclusion

A measurement of four-top-quark production using 139 fb$^{-1}$ of $\sqrt{s} = 13$ TeV proton-proton collision data collected by the ATLAS detector at the Large Hadron Collider is presented.

The selection in the 1L/2LOS final state targets events with a single lepton or an opposite-sign lepton pair accompanied by multiple jets. The measured four-top-quark production cross section is found to be $26 \pm 8 \text{ (stat.)} \pm 15 \text{ (syst.)} \text{ fb}$, corresponding to an observed (expected) signal significance of $1.9 \ (1.0)$ standard deviations over the background-only
hypothesis. The measurement uncertainty is dominated by the systematic uncertainties, particularly the modelling uncertainty of the $t\bar{t}t\bar{t}$ signal and the $t\bar{t}+\text{HF}$ background.

The result in the 1L/2LOS final state is combined with the previous measurement performed by the ATLAS Collaboration in the multilepton final state with the same data set. The combined four-top-quark production cross section is measured to be $\sigma_{t\bar{t}t\bar{t}} = 24 \pm 4 \text{ (stat.)} \pm 5 \text{ (syst.) fb}$. It is consistent within 2.0 standard deviations with the Standard Model expectation of $\sigma_{t\bar{t}t\bar{t}}^{\text{SM}} = 12.0 \pm 2.4 \text{ fb}$. The observed (expected) significance of the combined $t\bar{t}t\bar{t}$ cross section is 4.7 (2.6) standard deviations above the background-only hypothesis. It improves upon the observed (expected) significance of 4.3 (2.4) standard deviations from the analysis in the multilepton final state.

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 and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRI, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; JINR; MES of Russia and NRC KI, Russian Federation; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DSI/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; COST, ERC, ERDF, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; Norwegian Financial Mechanism 2014-2021, Norway; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, 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 (U.K.) and BNL (U.S.A.), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. [91].
Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References


[72] ATLAS collaboration, Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton-proton collisions at $\sqrt{s} = 13$TeV with the ATLAS detector, JHEP 04 (2019) 046 [arXiv:1811.12113] [INSPIRE].

[73] CMS collaboration, Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in $pp$ collisions at $\sqrt{s} = 13$TeV, Phys. Lett. B 803 (2020) 135285 [arXiv:1909.05306] [INSPIRE].


[75] A. Hocker et al., TMVA — toolkit for multivariate data analysis, physics/0703039 [INSPIRE].


Department of Physics, University of Adelaide, Adelaide, Australia

Department of Physics, University of Alberta, Edmonton AB, Canada

Department of Physics, Ankara University, Ankara, Turkey

Department of Physics, Istanbul Aydin University, Application and Research Center for Advanced Studies, Istanbul, Turkey

Department of Physics, TOBB University of Economics and Technology, Ankara, Turkey

LAPP, Univ. Savoie Mont Blanc, CNRS/IN2P3, Annecy, France

High Energy Physics Division, Argonne National Laboratory, Argonne IL, United States of America

Department of Physics, University of Arizona, Tucson AZ, United States of America

Department of Physics, University of Texas at Arlington, Arlington TX, United States of America

Physics Department, National and Kapodistrian University of Athens, Athens, Greece

Physics Department, National Technical University of Athens, Zografou, Greece

Department of Physics, University of Texas at Austin, Austin TX, United States of America

Bahcesehir University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey

Istanbul Bilgi University, Faculty of Engineering and Natural Sciences, Istanbul, Turkey

Department of Physics, Bogazici University, Istanbul, Turkey

Department of Physics Engineering, Gaziantep University, Gaziantep, Turkey

Institut de Física d’Altes Energies (IFAE), Barcelona Institute of Science and Technology, Barcelona, Spain

Institute of High Energy Physics, Chinese Academy of Sciences, Beijing, China

Physics Department, Tsinghua University, Beijing, China

Department of Physics, Nanjing University, Nanjing, China

University of Chinese Academy of Sciences (UCAS), Beijing, China

Institute of Physics, University of Belgrade, Belgrade, Serbia

Department for Physics and Technology, University of Bergen, Bergen, Norway
Kyoto University of Education, Kyoto; Japan
Research Center for Advanced Particle Physics and Department of Physics, Kyushu University, Fukuoka; Japan
Instituto de Física La Plata, Universidad Nacional de La Plata and CONICET, La Plata; Argentina
Physics Department, Lancaster University, Lancaster; United Kingdom
Oliver Lodge Laboratory, University of Liverpool, Liverpool; United Kingdom
Department of Experimental Particle Physics, Jožef Stefan Institute and Department of Physics, University of Ljubljana, Ljubljana; Slovenia
School of Physics and Astronomy, Queen Mary University of London, London; United Kingdom
Department of Physics, Royal Holloway University of London, Egham; United Kingdom
Department of Physics and Astronomy, University College London, London; United Kingdom
Louisiana Tech University, Ruston LA; United States of America
Fysikum, Lunds universitet, Lund; Sweden
Centre de Calcul de l’Institut National de Physique Nucléaire et de Physique des Particules (IN2P3), Villeurbanne; France
Departamento de Física Teorica C-15 and CIAFF, Universidad Autónoma de Madrid, Madrid; Spain
Institut für Physik, Universität Mainz, Mainz; Germany
School of Physics and Astronomy, University of Manchester, Manchester; United Kingdom
CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille; France
Department of Physics, University of Massachusetts, Amherst MA; United States of America
Department of Physics, McGill University, Montreal QC; Canada
School of Physics, University of Melbourne, Victoria; Australia
Department of Physics, University of Michigan, Ann Arbor MI; United States of America
Department of Physics and Astronomy, Michigan State University, East Lansing MI; United States of America
B.I. Stepanov Institute of Physics, National Academy of Sciences of Belarus, Minsk; Belarus
Research Institute for Nuclear Problems of Byelorussian State University, Minsk; Belarus
Group of Particle Physics, University of Montreal, Montreal QC; Canada
P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow; Russia
National Research Nuclear University MEPhI, Moscow; Russia
D.V. Skobeltsyn Institute of Nuclear Physics, M.V. Lomonosov Moscow State University, Moscow; Russia
Fakultät für Physik, Ludwig-Maximilians-Universität München, München; Germany
Max-Planck-Institut für Physik (Werner-Heisenberg-Institut), München; Germany
Graduate School of Science and Kobayashi-Maskawa Institute, Nagoya University, Nagoya; Japan
Department of Physics and Astronomy, University of New Mexico, Albuquerque NM; United States of America
Institute for Mathematics, Astrophysics and Particle Physics, Radboud University/Nikhef, Nijmegen; Netherlands
Nikhef National Institute for Subatomic Physics and University of Amsterdam, Amsterdam; Netherlands
Department of Physics, Northern Illinois University, DeKalb IL; United States of America
(a) Budker Institute of Nuclear Physics and NSU, SB RAS, Novosibirsk; (b) Novosibirsk State University Novosibirsk; Russia
Institute for High Energy Physics of the National Research Centre Kurchatov Institute, Protvino; Russia
Institute for Theoretical and Experimental Physics named by A.I. Alikhanov of National Research Centre “Kurchatov Institute”, Moscow; Russia
Department of Physics, New York University, New York NY; United States of America
Ochanomizu University, Otsuka, Bunkyo-ku, Tokyo; Japan
Ohio State University, Columbus OH; United States of America
Department of Physics, Tokyo Institute of Technology, Tokyo; Japan
Tomsk State University, Tomsk; Russia
Department of Physics, University of Toronto, Toronto ON; Canada
Department of Physics, University of Illinois, Urbana IL; United States of America
Department of Physics, University of British Columbia, Vancouver BC; Canada
Department of Physics and Astronomy, York University, Toronto ON; Canada
Division of Physics and Tomonaga Center for the History of the Universe, Faculty of Pure and Applied Sciences, University of Tsukuba, Tsukuba; Japan
Department of Physics and Astronomy, Tufts University, Medford MA; United States of America
Department of Physics and Astronomy, University of California Irvine, Irvine CA; United States of America
Department of Physics and Astronomy, University of Uppsala, Uppsala; Sweden
Department of Physics, University of Illinois, Urbana IL; United States of America
Instituto de Física Corpuscular (IFIC), Centro Mixto Universidad de Valencia — CSIC, Valencia; Spain
Department of Physics, University of British Columbia, Vancouver BC; Canada
Department of Physics and Astronomy, University of Victoria, Victoria BC; Canada
Fakultät für Physik und Astronomie, Julius-Maximilians-Universität Würzburg, Würzburg; Germany
Department of Physics, University of Warwick, Coventry; United Kingdom
Waseda University, Tokyo; Japan
Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot; Israel
Department of Physics, University of Wisconsin, Madison WI; United States of America
Fakultät für Mathematik und Naturwissenschaften, Fachgruppe Physik, Bergische Universität Wuppertal, Wuppertal; Germany
Department of Physics, Yale University, New Haven CT; United States of America

 Also at Borough of Manhattan Community College, City University of New York, New York NY; United States of America
 Also at Bruno Kessler Foundation, Trento; Italy
 Also at Center for High Energy Physics, Peking University; China
 Also at Centro Studi e Ricerche Enrico Fermi; Italy
 Also at CERN, Geneva; Switzerland
 Also at CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille; France
 Also at Département de Physique Nucléaire et Corpusculaire, Université de Genève, Genève; Switzerland
 Also at Departament de Física de la Universitat Autònoma de Barcelona, Barcelona; Spain
 Also at Department of Financial and Management Engineering, University of the Aegean, Chios; Greece
 Also at Department of Physics and Astronomy, Michigan State University, East Lansing MI; United States of America
 Also at Department of Physics and Astronomy, University of Louisville, Louisville, KY; United States of America
 Also at Department of Physics, Ben Gurion University of the Negev, Beer Sheva; Israel
 Also at Department of Physics, California State University, East Bay; United States of America
 Also at Department of Physics, California State University, Sacramento; United States of America
 Also at Department of Physics, King’s College London, London; United Kingdom
 Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg; Russia
 Also at Department of Physics, University of Fribourg, Fribourg; Switzerland
 Also at Faculty of Physics, M.V. Lomonosov Moscow State University, Moscow; Russia
 Also at Faculty of Physics, Sofia University, ‘St. Kliment Ohridski’, Sofia; Bulgaria
 Also at Giresun University, Faculty of Engineering, Giresun; Turkey
Also at Hellenic Open University, Patras; Greece

Also at Instituto Catalana de Recerca i Estudis Avancats, ICREA, Barcelona; Spain

Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest; Hungary

Also at Institute of Particle Physics (IPP); Canada

Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku; Azerbaijan

Also at Instituto de Fisica Teorica, IFT-UAM/CSIC, Madrid; Spain

Also at Istanbul University, Dept. of Physics, Istanbul; Turkey

Also at Joint Institute for Nuclear Research, Dubna; Russia

Also at Moscow Institute of Physics and Technology State University, Dolgoprudny; Russia

Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg; Germany

Also at The City College of New York, New York NY; United States of America

Also at TRIUMF, Vancouver BC; Canada

Also at Universita di Napoli Parthenope, Napoli; Italy

Also at University of Chinese Academy of Sciences (UCAS), Beijing; China

Also at Yeditepe University, Physics Department, Istanbul; Turkey

* Deceased