EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: JHEP

CERN-EP-2022-170
2nd November 2022

Search for $t\bar{t}H/A \to t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

A search for a new heavy scalar or pseudo-scalar Higgs boson (H/A) produced in association with a pair of top quarks, with the Higgs boson decaying into a pair of top quarks ($H/A \to t\bar{t}$) is reported. The search targets a final state with exactly two leptons with same-sign electric charges or at least three leptons. The analysed dataset corresponds to an integrated luminosity of 139 fb^{-1} of proton–proton collisions collected at a centre-of-mass energy of 13 TeV with the ATLAS detector at the LHC. Two multivariate classifiers are used to separate the signal from the background. No significant excess of events over the Standard Model expectation is observed. The results are interpreted in the context of a type-II two-Higgs-doublet model. The observed (expected) upper limits at 95% confidence level on the $t\bar{t}H/A$ production cross-section times the branching ratio of $H/A \to t\bar{t}$ range between 14 (10) fb and 6 (5) fb for a heavy Higgs boson with mass between 400 GeV and 1000 GeV, respectively. Assuming that only one particle, either the scalar H or the pseudo-scalar A, contributes to the $t\bar{t}t\bar{t}$ final state, values of $\tan\beta$ below 1.2 or 0.5 are excluded for a mass of 400 GeV or 1000 GeV, respectively. These exclusion ranges increase to $\tan\beta$ below 1.6 or 0.6 when both particles are considered.

Contents

1 Introduction

The discovery of the Higgs boson by the ATLAS and CMS Collaborations in 2012 [1, 2] and later precision measurements [3, 4] of Higgs boson production and decay properties established the Standard Model (SM) of particle physics as an effective description of nature up to the TeV energy scale.

Within the SM, the Brout–Englert–Higgs mechanism [5–10] is responsible for generating the mass of the gauge bosons via electroweak symmetry breaking (EWSB). The Higgs boson emerges from the EWSB as the only physical spin-0 CP-even particle of the SM, while the remaining components of the Higgs field are absorbed into the longitudinal components of the gauge bosons.

In many 'beyond the Standard Model' (BSM) scenarios, the Higgs sector is extended to incorporate new degrees of freedom. A well-motivated and minimal extension of the SM paradigm is provided by two-Higgs-doublet models (2HDM) [11] where the Higgs sector consists of two complex doublets: a mixture of the two doublets plays the same role as the SM Higgs field and generates a Higgs boson (h); the other mixture produces three Higgs bosons – one neutral CP-even (H), one neutral CP-odd (A), and one charged ($H^{\pm}$). A generic CP-conserving 2HDM with natural flavour conservation, and a lighter CP-even Higgs boson playing the role of the SM Higgs boson, has five free parameters: three Higgs boson masses (m_H, m_A, $m_{H^{\pm}}$), the mixing angle of the two CP-even Higgs fields (α), and the ratio of the vacuum expectation values of the two Higgs doublets ($\tan\beta$). A common realisation of the 2HDM is the Minimal Supersymmetric Standard Model (MSSM) [12, 13]. Similar Higgs sectors arise in axion models [14].

Existing constraints from direct searches for heavy neutral bosons by the ATLAS and CMS Collaborations [15–25], as well as precision measurements of the production cross-sections and decay rate of the SM Higgs boson, restrict the available parameter-space to the so-called 'alignment limit', $\sin(\beta-\alpha) \to 1$. In this limit the h couplings are the same as for the SM Higgs boson.

For heavy neutral Higgs bosons with masses more than twice the top-quark mass, the dominant decay mode is $H/A \to t\bar{t}$. Inclusive searches for $H/A \to t\bar{t}$ are challenging because of destructive interference with the SM background, $gg \to t\bar{t}$, that largely dilutes a resonant peak in the $t\bar{t}$ invariant mass spectrum [26, 27]. An alternative approach is to search for H/A production in association with third-generation quarks [28]; thanks to its sizeable cross-section and striking signature, the $t\bar{t}H/A$ production mode provides a promising channel, characterised by large experimental acceptance and low SM background rate.

This paper presents a search for a new heavy scalar or pseudo-scalar Higgs boson, H/A, produced in association with a pair of top quarks, with the Higgs boson decaying into a pair of top quarks, $t\bar{t}H/A(\to t\bar{t})$. The Feynman diagram of this process is shown in Figure 1. The mass of the heavy Higgs boson is assumed to be between 400 GeV and 1000 GeV, where a large $H/A \to t\bar{t}$ branching ratio and small H/A total widths are expected. The search targets a final state with exactly two leptons[1] with same-sign electric charges or at least three leptons (SSML). This particular signature is experimentally favoured due to the low level of background contamination, with the main contribution originating from the SM production of four top quarks and of $t\bar{t}$ in association with a W boson ($t\bar{t}W$), Z boson ($t\bar{t}Z$), or Higgs boson ($t\bar{t}H$). Other significant sources of background are events where one of the leptons has a mis-assigned charge and events which contain a fake/non-prompt lepton. Backgrounds from multiboson, $t\bar{t}WW$, single-top-quark or other rare top-quark processes are expected to be minor. The analysed dataset corresponds to an integrated luminosity of 139 fb^{-1} of proton–proton (pp) collisions collected at a centre-of-mass energy of $\sqrt{s}$ =13 TeV with the ATLAS detector at the LHC.

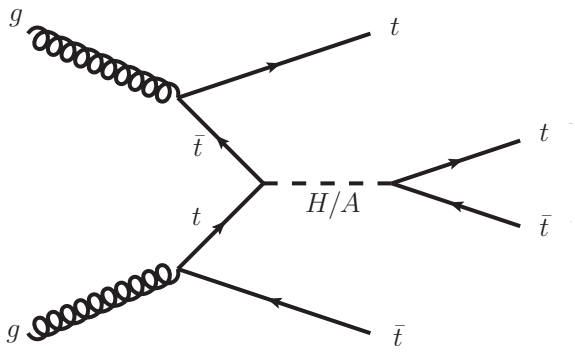


Figure 1: Feynman diagram showing the production of a heavy scalar or pseudo-scalar Higgs boson, H/A, produced in association with a pair of top quarks, with the Higgs boson decaying into a pair of top quarks.

Previous searches for $t\bar{t}H/A(\to t\bar{t})$ in the SSML channel were performed by the ATLAS and CMS Collaborations [29, 30]. A similar search for BSM $t\bar{t}t\bar{t}$ events using an alternative experimental signature featuring exactly one charged lepton or two opposite-sign leptons was performed by the CMS Collaboration [31]. Other related searches include those looking for SM production of four top quarks [30, 32, 33]. The ATLAS and CMS measurements of SM four-top-quark production found the cross-section to be 24^{+7}_{-6} fb and 13^{+11}_{-9} fb, respectively, compared to a SM expectation of $\sigma_{t\bar{t}t\bar{t}} = 12.0 \pm 2.4$ fb [34].

[1] In this paper, leptons refer to either electrons or muons, which can include those that come from τ-lepton decays.

The results obtained in this search are interpreted in the context of a type-II 2HDM; upper limits are placed on the $t\bar{t}H/A$ production cross-section times the branching ratio of $H/A \to t\bar{t}$ as well as on $\tan\beta$ as a function of the heavy Higgs boson's mass. Due to a significantly larger dataset and improved analysis techniques, including the use of multivariate classifiers based on boosted decision trees (BDTs) to distinguish the signal from the SM background, the expected sensitivity of the present $t\bar{t}H/A(\to t\bar{t})$ search exceeds that of the previous ATLAS $t\bar{t}H/A(\to t\bar{t})$ search with 36 fb^{-1} [29] by about a factor of four.

This paper is structured as follows. The ATLAS detector is presented in Section 2. The data and samples of simulated events are described in Section 3. The object and event selections are detailed in Section 4. The analysis strategy is presented in Section 5. This is followed by a description of the systematic uncertainties in Section 6. The statistical analysis and the results are presented in Sections 7 and 8, respectively. Finally, the conclusion is presented in Section 9.

2 ATLAS detector

The ATLAS detector [35] at the LHC covers nearly the entire solid angle around the collision point.[2] It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadron calorimeters, and a muon spectrometer incorporating three large superconducting air-core toroidal magnets with eight coils each.

The inner-detector system (ID) is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four measurements per track, the first hit normally being in the insertable B-layer (IBL) installed before Run 2 [36, 37]. It is surrounded by the silicon microstrip tracker (SCT), which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$. The TRT also provides electron identification information based on the fraction of hits (typically 30 in total) above a higher energy-deposit threshold corresponding to transition radiation.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) calorimeters, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters. Hadron calorimetry is provided by the steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadron endcap calorimeters. The solid angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic energy measurements respectively.

The muon spectrometer (MS) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by the superconducting air-core toroidal magnets. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. A set of precision chambers covers the region $|\eta| < 2.7$ with three layers of monitored drift tubes, complemented by cathode-strip chambers in the forward region, where the background is highest. The muon trigger system

[2] 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, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

covers the range $|\eta| < 2.4$ with resistive-plate chambers in the barrel, and thin-gap chambers in the endcap regions.

Interesting events are selected by the first-level trigger system implemented in custom hardware, followed by selections made by algorithms implemented in software in the high-level trigger [38]. The first-level trigger accepts events from the 40 MHz bunch crossings at a rate below 100 kHz, which the high-level trigger further reduces in order to record events to disk at about 1 kHz.

An extensive software suite [39] is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

3 Data and samples of simulated events

This analysis uses pp collision data collected between 2015 and 2018 by the ATLAS detector at $\sqrt{s} = 13$ TeV. The full dataset corresponds to an integrated luminosity of 139 fb^{-1}. Only data recorded during stable beam conditions in which all detector subsystems were operational [40] are included. Events were collected using single-lepton or dilepton triggers. The trigger p_{T} threshold depended on the lepton flavour and the data-taking period [41, 42]. The lowest p_{T} threshold for the single-lepton trigger varied from 20 to 26 GeV, while for the dilepton trigger it varied from 7 to 24 GeV.

Monte Carlo (MC) simulation samples are used for the different signal and background processes. The generated events were processed through the simulation [43] of the ATLAS detector geometry and response based on Geant4 [44], and through the same reconstruction software as data. Corrections are applied to the simulated events so that particle selection efficiencies, energy scales and energy resolutions match those determined from data. For the parton shower and hadronisation, unless otherwise stated, samples generated with MadGraph5_aMC@NLO [45] and Powheg Box v2 [46–52] were interfaced to Pythia 8.230, 8.210 or 8.212 [53], using the A14 set of tuned parameters [54] and the NNPDF2.3lo [55] parton distribution function (PDF) set. The decays of bottom and charm hadrons were simulated using the EvtGen 1.6.0 or EvtGen 1.2.0 program [56].

Signal MC samples were generated using the MadGraph5_aMC@NLO 2.3.3 generator at leading order (LO) with the NNPDF3.1lo [55] PDF set. In order to account for the spin correlation of the particles in the final state, the full matrix-element for the $2 \to 12$ parton-scattering process was generated. The signal MC events were generated assuming the production of $t\bar{t}H(\to t\bar{t})$ in a type-II 2HDM. A total of seven signal MC samples were generated with m_H ranging from 400 GeV to 1000 GeV with regular 100 GeV spacing. The width of the heavy Higgs boson is assumed to be small throughout the mass range, varying from 5 GeV to 30 GeV, consistent with the expected width of a heavy Higgs boson in the type-II 2HDM for $\tan\beta \sim 1$. Interference between the signal and SM four-top-quark production was not simulated. In the mass range of interest for this search, the interference leads to a change of up to 20% in the cross-section for a signal with $m_H = 400$ GeV and a width of 50 GeV, and the kinematic properties of the signal are found to be consistent between A and H. For $\tan\beta \sim 1$, the $t\bar{t}H(\to t\bar{t})$ and $t\bar{t}A(\to t\bar{t})$ production cross-sections are found to differ by 1% at most.

The SM production of $t\bar{t}t\bar{t}$ events was modelled in the following way. The nominal sample was produced with the MadGraph5_aMC@NLO 2.3.3 generator, which provides matrix elements at next-to-leading order (NLO) in the strong coupling constant α_{s} with the NNPDF3.1nlo PDF set. The functional form of the renormalisation and factorisation scales was set to $\mu_{\mathrm{r}} = \mu_{\mathrm{f}} = m_{\mathrm{T}}/4$ where m_{T} is defined as the scalar

sum of the transverse masses $\sqrt{m^2 + p_{\mathrm{T}}^2}$ of the particles generated from the matrix element calculation, following Ref. [34]. The decay of the top quark was simulated at LO using MadSpin [57, 58] to preserve all spin correlations. To avoid the use of negative weights present in the nominal NLO sample in the training of the multivariate discriminant used to separate SM $t\bar{t}t\bar{t}$ events from background (see Section 5.2), a sample was produced with similar generator settings, but at LO. An additional $t\bar{t}t\bar{t}$ sample was produced with the Pythia 8.230 parton shower of the nominal sample replaced by Herwig 7.04 [59, 60] to evaluate the impact of using a different parton shower and hadronisation model. The H7UE set of tuned parameters [60] and the MMHT2014lo PDF set [61] were used. A sample of $t\bar{t}t\bar{t}$ events modelled using the Sherpa 2.2.10 [62] generator with the NNPDF3.0nnlo [55] PDF set is used to assess the uncertainty associated with the choice of event generator. A sample including electroweak (EW) corrections [63] was generated with the same settings as in Sherpa 2.2.10, but using Sherpa 2.2.11, in order to evaluate the systematic uncertainty associated with the missing EW corrections in the nominal sample. The production of $t\bar{t}t\bar{t}$ events is normalised to a cross-section of 12 fb computed at NLO in QCD including EW corrections [34].

The production of $t\bar{t}W$ events was modelled at NLO in QCD using the Sherpa 2.2.10 generator with the NNPDF3.0nnlo PDF set, including up to one extra parton at NLO and up to two extra partons at LO. The additional partons were matched and merged with the Sherpa parton shower based on Catani–Seymour dipole factorisation [64] using the MEPS@NLO prescription [65–68] with a merging scale of 30 GeV. The virtual QCD corrections for matrix elements at NLO accuracy were provided by the OpenLoops 2 [69–71] library. The renormalisation and factorisation scales were set to $\mu_{\mathrm{r}} = \mu_{\mathrm{f}} = m_{\mathrm{T}}/2$. The production of $t\bar{t}W$ events with only EW corrections was modelled at LO in QCD using the same generator and PDF set as in their QCD-only production. The $t\bar{t}W$ cross-section predicted by these MC generator settings, including EW corrections, is 639 fb. The impact of the systematic uncertainty associated with the choice of generator is evaluated with an alternative $t\bar{t}W$ sample generated at NLO in QCD with no additional partons using the MadGraph5_aMC@NLO 2.3.3 generator with the NNPDF3.0nlo PDF set. The production of $t\bar{t}W$ events in MadGraph5_aMC@NLO with only EW corrections was modelled at LO in QCD with the NNPDF2.3lo PDF set.

The production of $t\bar{t}(Z/\gamma^*)$ events was modelled using the Sherpa 2.2.1 generator at NLO with the NNPDF3.0nnlo PDF set. The invariant mass of the lepton pair was required to satisfy $m_{\ell^+\ell^-} > 5$ GeV. For the low mass range, i.e. $m_{\ell^+\ell^-} \in (1, 5)$ GeV, events were modelled using the MadGraph5_aMC@NLO 2.3.3 generator at NLO with the NNPDF3.0nlo PDF set. These two samples were combined and together they form the '$t\bar{t}(Z/\gamma^*)$ (high mass)' sample. In order to assess the uncertainty associated with the choice of generator, a MadGraph5_aMC@NLO sample generated with exactly the same settings as the nominal low-mass-range sample but with $m_{\ell^+\ell^-} > 5$ GeV was used.

The production of $t\bar{t}H$ events was modelled using the Powheg Box v2 generator, which provided matrix elements at NLO in the strong coupling constant α_{s} in the five-flavour scheme with the NNPDF3.0nlo PDF set. The cross-section was calculated at NLO QCD and NLO EW accuracy using MadGraph5_aMC@NLO as reported in Ref. [72]. An alternative sample generated with MadGraph5_aMC@NLO 2.3.3 is used to evaluate the impact of the uncertainty associated with the generator choice.

The production of $t\bar{t}$ and single-top-quark events was modelled using the Powheg Box v2 generator at NLO in QCD with the NNPDF3.0nlo PDF set. For the $t\bar{t}$ sample, the h_{damp} parameter[3] was set to $1.5\, m_{\mathrm{top}}$ [73]. The diagram removal scheme [74] was used to remove interference and overlap between $t\bar{t}$ and tW production. For these two processes, the contribution from internal photon conversions ($\gamma^* \to \ell^+\ell^-$)

[3] The h_{damp} parameter is a resummation damping factor and one of the parameters that controls the matching of Powheg matrix elements to the parton shower and thus effectively regulates the high-p_{T} radiation against which the $t\bar{t}$ system recoils.

with $m_{\ell^+\ell^-} < 1$ GeV was modelled by QED multiphoton radiation via the parton shower. This contribution is referred to in the following as the 'Low m_{γ^*}' sample.

The production of tWZ, tZ, $t\bar{t}t$, $t\bar{t}WW$ and other rare top-quark processes, namely the $t\bar{t}ZZ$, $t\bar{t}WZ$, $t\bar{t}HH$, and $t\bar{t}WH$ processes, was modelled using the MadGraph5_aMC@NLO generator. The tWZ sample was modelled at NLO, while the remaining processes were modelled at LO in QCD. The contribution of $t\bar{t}WW$ is normalised to the NLO QCD theoretical cross-section [72].

The WH and ZH processes were modelled using the Pythia 8.230 generator with the A14 tune and NNPDF2.3lo PDF set and normalised to their theoretical cross-sections calculated at NNLO QCD and NLO EW accuracies [75–81].

The production of Z+jets, W+jets, diboson (VV) and triboson (VVV) events was modelled with the Sherpa 2.2.1 or 2.2.2 generator depending on the process. The NNPDF3.0nnlo set of PDFs was used, along with the dedicated set of tuned parton-shower parameters developed by the Sherpa authors.

The effect of multiple interactions in the same and neighbouring bunch crossings (pile-up) was modelled by overlaying the simulated hard-scattering event with inelastic pp events generated with Pythia 8.186 [82] using the NNPDF2.3lo set of PDFs and the A3 set of tuned parameters [83].

4 Object and event selections

Events are required to contain at least one primary vertex (PV) reconstructed from at least two ID tracks, each with $p_{\mathrm{T}} > 0.5$ GeV. If more than one PV candidate satisfies these criteria, then the PV with the largest sum of p_{T}^2 over all associated ID tracks is selected [84].

Electron candidates are reconstructed from energy deposits in the electromagnetic calorimeter matched to a track in the ID [85]. The identification working point for the nominal selection in this analysis is 'TightLH' [85]. Only electron candidates with $p_{\mathrm{T}} > 28$ GeV and within $|\eta| < 2.47$, excluding the calorimeter transition region $1.37 < |\eta| < 1.52$, are selected. Electrons are required to be well isolated using criteria based on the properties of the topological clusters in the calorimeter and of ID tracks around the reconstructed electron. Standard requirements on the longitudinal (z_0) and transverse (d_0) impact parameters are used to select electrons originating from the primary vertex. The requirements are $|z_0 \sin\theta| < 0.5$ mm and $|d_0|/\sigma(d_0) < 5$.

Muon candidates are reconstructed by combining tracks in the ID with tracks in the MS [86]. The identification working point used in this analysis is 'Medium' [86]. Only muon candidates with $p_{\mathrm{T}} > 28$ GeV and within $|\eta| < 2.5$ are selected. Muons are required to satisfy isolation requirements based on the properties of ID tracks around the reconstructed muon. Similarly to electrons, requirements on the longitudinal and transverse impact parameters, $|z_0 \sin\theta| < 0.5$ mm and $|d_0|/\sigma(d_0) < 3$, are also applied.

An additional requirement is imposed on electrons in the $e^\pm e^\pm$ and $e^\pm \mu^\pm$ channels to reduce the background coming from electron charge misidentification. This requirement is based on a BDT discriminant which uses the calorimeter and tracking variables to remove approximately 90% of electrons with the wrong charge assignment while selecting 98% of electrons with the correctly measured charge [85].

The constituents for jet reconstruction are identified by combining measurements from both the ID and the calorimeter using a particle-flow algorithm [87]. Jet candidates are reconstructed from these particle-flow objects using the anti-k_t algorithm [88, 89] with a radius parameter of $R = 0.4$. They are calibrated using

simulation with corrections obtained from data using in situ techniques [90]. Only jet candidates with $p_{\mathrm{T}} > 25$ GeV and within $|\eta| < 2.5$ are selected. To reduce the effect of pile-up, each jet with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ is required to satisfy the 'Tight' working point of the jet-vertex tagger (JVT) [91] criteria used to identify the jets as originating from the selected primary vertex. A set of quality criteria are also applied to reject events containing at least one jet arising from non-collision sources or detector noise [92].

Jets containing b-hadrons are identified (b-tagged) using the DL1r algorithm [93]. It consists of a deep-learning neural network based on the distinctive features of the b-hadrons in terms of the impact parameters of tracks and the displaced decay vertices reconstructed in the ID. The input to the DL1r network also includes discriminating variables constructed by a recurrent neural network [94], which exploits the spatial and kinematic correlations between tracks originating from the same b-hadron. A jet is b-tagged if the DL1r score is above 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%, as determined in simulated $t\bar{t}$ events. A jet is considered b-tagged if it passes the OP corresponding to 77% average efficiency for jets containing b-hadrons, with a misidentification rate of 1/130 (1/4.9) for light-flavour (charm) jets. Correction factors are applied to the simulated event samples to compensate for differences between data and simulation in the b-tagging efficiency for b-jets, c-jets and light-flavour jets. The correction for b-jets is derived from $t\bar{t}$ events with final states containing two leptons, and the corrections are consistent with unity with uncertainties at the level of a few percent over most of the jet p_{T} range [95].

An overlap removal procedure is applied to ensure that the same calorimeter energy deposit or the same track is not used in two different objects. First, any electron found to share a track with another electron with higher p_{T} is removed. Next, electrons sharing their track with a muon candidate are removed. Then, jets are removed if they are within $\Delta R = 0.2$ of an electron. Subsequently, in order to remove electrons arising from b- or c-decays, electrons are removed if they are within a $\Delta R = 0.4$ of a jet. Next, jets within $\Delta R = 0.2$ of a muon are removed if they have less than three tracks. Finally, in order to remove muons arising from b- or c-decays, muons are removed if their tracks are within $\Delta R = 0.4 + 10\ \mathrm{GeV}/p_{\mathrm{T}}^{\mu}$ of any remaining jets.

The missing transverse momentum of the event, $\vec{p}_{\mathrm{T}}{}^{\mathrm{miss}}$, is defined as the negative vector sum of the p_{T} of all selected and calibrated objects in the event. This sum includes a term to account for the transverse momenta of ID tracks matched to the selected PV but which are not associated with any of the selected objects in the event [96]. The magnitude of $\vec{p}_{\mathrm{T}}{}^{\mathrm{miss}}$ is denoted by $E_{\mathrm{T}}^{\mathrm{miss}}$.

Events are required to have exactly two leptons with the same electric charge or at least three leptons without any charge requirement. Each event must have at least one reconstructed lepton that matches a lepton that satisfied the trigger requirements. In order to reduce the background coming from electron charge misidentification in the same-sign dilepton channel, and the contamination from the Z boson decay in the trilepton channel, a Z-veto rejecting $m_{\ell\ell} \in [81, 101]$ GeV is applied, where $m_{\ell\ell}$ is checked for the two electrons defining the $e^{\pm}e^{\pm}$ channel, and for all opposite-sign same-flavour lepton pairs in the trilepton channel. In addition, events with an $e^{\pm}e^{\pm}$ final state are also required to satisfy $m_{ee} > 15$ GeV to reduce the background coming from low-mass resonances with electron charge misidentification.

5 Analysis strategy

The analysis strategy follows the one used in the previous ATLAS $t\bar{t}t\bar{t}$ measurement [33], with dedicated control regions to constrain the dominant background processes, and multivariate techniques to separate

the signal from background. The background estimation is detailed in Section 5.1.

A signal-enriched region is defined by exploiting the high multiplicity of light-flavour and b-tagged jets as well as the high overall momentum of final-state jets and leptons in $t\bar{t}t\bar{t}$ events. Thus, events are required to have at least six jets, among which at least two are b-tagged, and $H_{\mathrm{T}} > 500$ GeV, where H_{T} is defined as the scalar sum of the p_{T} of all leptons and jets in the event. This signal region (SR) is referred to as the *baseline SR* in the following.

The signal is separated from the SM background by using two sequential BDT classifiers (see Section 5.2). These BDTs are trained in the *baseline SR*, inclusively in both lepton flavour and multiplicity. The first one, the 'SM BDT', is used to separate SM $t\bar{t}t\bar{t}$ events from the remaining backgrounds. Due to the similar kinematics of the signal and SM $t\bar{t}t\bar{t}$ events, signal events are expected to be located in the same region of the BDT score distribution as the SM $t\bar{t}t\bar{t}$ events. Therefore, the final signal region is a part of the *baseline SR*, and it is defined by the requirement SM BDT > 0.55. This definition is referred to as the *BSM SR* in the following. The full selection for the *BSM SR* can be found in Table 1. The second classifier, the 'BSM pBDT', is a mass-parameterised BDT that discriminates signal events from all backgrounds. To test for the presence of a signal, the distribution of the BSM pBDT score in the *BSM SR* is fitted jointly with the control regions defined in Section 5.1. For the $t\bar{t}H(\rightarrow t\bar{t})$ signal described in Section 3, the acceptance times the efficiency to pass the *BSM SR* requirements ranges from 2% to 3% for the different signal hypotheses.

The fit was initially validated through extensive studies using fits to real data where bins of the SM BDT score with a signal contamination above 5% were excluded (referred to as blinding requirements). The robustness of the model for systematic uncertainties is established by verifying the stability of the fitted background when varying assumptions about some of the leading sources of uncertainty. After this, the blinding requirements on the SM BDT are removed in the data and a fit under the signal-plus-background hypothesis is performed to all analysis regions. Further checks involve the comparison of the fitted nuisance parameters before and after removal of the blinding requirements, and their values are found to be consistent confirming that the signal-rich analysis bins do not contribute to the background estimation. The modelling of a number of kinematic variables in all considered analysis regions is also studied before and after the fit, confirming that the fit consistently improves the modelling of all of them.

5.1 Background estimation

Several SM processes can mimic a final state of two leptons with the same electric charge or three leptons. They can be classified into two different categories depending on the origin of the lepton:

Irreducible: All selected leptons are prompt. They originate mainly from W or Z boson decays. The main contribution in the signal region is given by $t\bar{t}t\bar{t}$, $t\bar{t}W$, $t\bar{t}(Z/\gamma^*)$ (high mass) and $t\bar{t}H$ production. Smaller contributions come from VV, VVV and VH production, and from rare processes ($t\bar{t}VV$, tWZ, tZq and $t\bar{t}t$). This background is estimated using the samples of simulated events described in Section 3 and normalised to the theory cross-sections. The $t\bar{t}W$ contribution is split between its QCD and EW components.[4] Because the normalisation of $t\bar{t}W$ has been found to be underestimated in the simulation in recent measurements [97], the normalisation of the QCD component is corrected using data in a dedicated control region as explained in Section 5.1.1. The Low m_{γ^*} background is also included in this category. Since the normalisation of this background might not be correct

[4] The $t\bar{t}W$ QCD and EW samples are calculated at different orders: QCD is NLO in QCD, and EW is LO in QCD. Therefore, the QCD and EW components are decorrelated and treated as two independent samples with their own systematic uncertainties.

in simulation, its normalisation is also estimated in a dedicated control region as explained in Section 5.1.1. This search targets BSM production of four top quarks, so the background originating from SM four-top-quark production is normalised to its SM prediction with a Gaussian constraint given by the theoretical uncertainty.

Reducible: At least one of the leptons is fake/non-prompt or it is a prompt lepton with its charge misidentified. It originates mainly from $t\bar{t}$+jets, V+jets and tW+jets production. The fake/non-prompt lepton background is estimated together with the $t\bar{t}W$ QCD and Low m_{γ^*} backgrounds by using the template fit method described in Section 5.1.1. The background coming from electron charge misidentification (QmisID) is estimated using data-driven techniques as described in Section 5.1.2.

5.1.1 Fake/non-prompt lepton background, Low m_{γ^*} and $t\bar{t}W$ QCD normalisations

Fake/non-prompt lepton background arises from different sources:

- Events with one lepton from a detector-material (or beam pipe) photon conversion (Mat. Conv.).
- Events with one electron (muon) from a heavy-flavour meson decay (HF e/μ).
- Events with one lepton from a light-meson decay or a jet misidentified as a lepton (LF).
- Events with one fake/non-prompt lepton not arising from any of the above (other fake).

The true generator-level information from the $t\bar{t}$+jets, V+jets and single-top samples described in Section 3 is used to separate the different sources of fake/non-prompt leptons. The last two components are very small and are fully estimated with samples of simulated events. The contributions from detector-material photon conversions and HF e/μ are estimated with the template fit method. This method relies on the simulation to model the kinematic distributions of fake/non-prompt leptons from different sources and uses dedicated control regions to determine their normalisations.

Several control regions were defined, with each region designed to maximise the background component which is mainly fitted in that region. The variable to be fitted in each region was chosen so as to provide the most discrimination for the targeted background. The normalisation of the Low m_{γ^*} irreducible background is mainly estimated in the control region enriched in background from detector-material photon conversions. A dedicated control region was defined for $t\bar{t}W$ QCD production. An extra control region was defined so as to be as close as possible to the *BSM SR*. In total, five control regions are used in the analysis, closely following the definitions in Ref. [33]. They are described in Table 1 and summarised below.

- 'CR Conv': It is enriched in detector-material photon conversions and Low m_{γ^*} background. Events are required to have an $e^{\pm}e^{\pm}$ or $e^{\pm}\mu^{\pm}$ pair. For each electron in the event, the invariant mass of the system formed by the track associated with the electron and the closest track at the conversion (primary) vertex m_{ee}^{CV} (m_{ee}^{PV}) is computed. The conversion vertex is defined at the point where the track from the electron and its closest track in ΔR have the same ϕ. Virtual photons lead to a lepton pair originating from the primary vertex, having a low $m_{ee}^{\mathrm{PV}} \sim m_{\gamma^*}$. Detector-material photon conversions have a large conversion radius, and the track extrapolation induces a large apparent m_{ee}^{PV}. The control region is then obtained by selecting events with low m_{ee}^{CV} and fitting the m_{ee}^{PV} distribution to separate the detector-material photon conversions and Low m_{γ^*} background from each other. Events are also required to have four or five jets, at least one b-tagged jet, and low H_{T}.

- 'CR HF e' ('CR HF μ'): It is enriched in events with one electron (muon) coming from heavy-flavour decay. The selection targets $t\bar{t}$ dilepton decays with an extra non-prompt lepton. This region is then defined by selecting events with three leptons, namely eee and $ee\mu$ ($e\mu\mu$ and $\mu\mu\mu$) for CR HF e (CR HF μ), low H_{T} and exactly one b-jet. The number of events in this region is used as the fitted variable.
- 'CR $t\bar{t}W$': It is enriched in $t\bar{t}W$ events. Events are required to have an $e^{\pm}\mu^{\pm}$ or $\mu^{\pm}\mu^{\pm}$ pair, at least four jets and at least two b-jets. In order to reduce the background coming from electron charge misidentification, events containing electrons with $|\eta| > 1.5$ are removed. This region is also required to be orthogonal to the 'CR Conv' region and to the *baseline SR*. The fitted variable is the scalar sum of the lepton p_{T}.
- 'CR lowBDT': It is not enriched in any particular background, but rather used as a control region which is very close to the *BSM SR*. Events are required to be in the *baseline SR*, but with SM BDT < 0.55. The fitted variable is the SM BDT output score.

These control regions are fitted simultaneously with the *BSM SR* to determine both the strength of any BSM $t\bar{t}t\bar{t}$ signal and the five normalisation factors: $\lambda_{t\bar{t}W\,\mathrm{QCD}}$ for $t\bar{t}W$ QCD production, $\lambda_{\mathrm{Mat.\ Conv.}}$ for the background from detector-material photon conversions, $\lambda_{\mathrm{Low}\ m_{\gamma^*}}$ for the contribution from virtual photons leading to e^+e^- pairs, and $\lambda_{\mathrm{HF}\,e}$ ($\lambda_{\mathrm{HF}\,\mu}$) for the non-prompt electron (muon) background from heavy-flavour decays.

Table 1: Definition of the signal region and control regions used in the analysis. The first column shows the region name as used in the text. The event selection requirements are defined in the middle columns. The last column shows the fitted variable in each region. The variable m_{ee}^{CV} (m_{ee}^{PV}) is defined as the invariant mass of the system formed by the track associated with the electron and the closest track at the conversion (primary) vertex. N_{j} (N_{b}) indicates the jet (b-tagged jet) multiplicity. H_{T} is defined as the scalar sum of the transverse momenta of the isolated leptons and jets. The *baseline SR* is equal to the *BSM SR* + CR lowBDT.

Region	Channel	N_{j}	N_{b}	Other selection requirements	Fitted variable
CR Conv	$e^{\pm}e^{\pm}$ || $e^{\pm}\mu^{\pm}$	$4 \leq N_{\mathrm{j}} < 6$	≥ 1	$m_{ee}^{\mathrm{CV}} \in [0, 0.1]$ GeV $200 < H_{\mathrm{T}} < 500$ GeV	m_{ee}^{PV}
CR HF e	eee || $ee\mu$		$= 1$	$100 < H_{\mathrm{T}} < 250$ GeV	Yield
CR HF μ	$e\mu\mu$ || $\mu\mu\mu$		$= 1$	$100 < H_{\mathrm{T}} < 250$ GeV	Yield
CR $t\bar{t}W$	$e^{\pm}\mu^{\pm}$ || $\mu^{\pm}\mu^{\pm}$	≥ 4	≥ 2	$m_{ee}^{\mathrm{CV}} \notin [0, 0.1]$ GeV, $\lvert\eta(e)\rvert < 1.5$ for $N_{\mathrm{b}} = 2$, $H_{\mathrm{T}} < 500$ GeV or $N_{\mathrm{j}} < 6$; for $N_{\mathrm{b}} \geq 3$, $H_{\mathrm{T}} < 500$ GeV	$\sum p_{\mathrm{T}}^{\ell}$
CR lowBDT	SS+3L	≥ 6	≥ 2	$H_{\mathrm{T}} > 500$ GeV, SM BDT < 0.55	SM BDT
BSM SR	SS+3L	≥ 6	≥ 2	$H_{\mathrm{T}} > 500$ GeV, SM BDT ≥ 0.55	BSM pBDT

5.1.2 Electron charge misidentification background

Background from electron charge misidentification is relevant only in the same-sign dilepton channel. It arises when the sign of the electric charge of one of the two leptons in the selected same-sign pair is misreconstructed either because of bremsstrahlung photon emission followed by its conversion ($e^{\pm} \to e^{\pm}\gamma \to e^{\pm}e^{+}e^{-}$) or due to mismeasured track curvature. In the signal region, it mainly comes from $t\bar{t}$+jets production. Due to the low probability of bremsstrahlung for muons and because of the large lever

arm of the MS, the muon charge misidentification rate is very low. Thus, this background is only relevant for the $e^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$ channels.

The probability for an electron to have its charge incorrectly reconstructed is measured in a data sample of dielectron events with invariant mass m_{ee} within 10 GeV of the Z boson mass. The sideband method is used to subtract the background contamination. The electron charge misidentification probability is calculated in bins of electron $|\eta|$ and p_{T}. In order to apply it to the conversion region defined in Section 5.1.1, it is also parameterised in bins of m_{ee}^{PV}. A likelihood fit that adjusts these binned probabilities is used to find the best agreement with the observed numbers of same-charge and opposite-charge electron pairs. The electron charge misidentification rates vary from 0.002% for low-p_{T} electrons ($p_{\mathrm{T}} \leq 60$ GeV) at $|\eta| \leq 0.6$, to ~10% for high-p_{T} electrons ($p_{\mathrm{T}} \geq 200$ GeV) at $|\eta| \in [2.3, 2.5]$.

To estimate the event yields in the $e^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$ channels, ee and $e\mu$ events are selected using all the criteria applied in the analysis, except that the leptons are required to have opposite charges. Then, the final background yield is obtained by weighting these opposite-sign dilepton events by the probability of one electron charge being misreconstructed.

5.2 Signal discrimination

Multivariate techniques are used to separate the signal from the SM backgrounds. This is done through two sequential BDT classifiers: the first one, the background rejection BDT, namely SM BDT, separates SM $t\bar{t}t\bar{t}$ events from other SM backgrounds. Then, the second one, the BSM mass-parameterised BDT (BSM pBDT) discriminates between BSM $t\bar{t}t\bar{t}$ events and all background. The BSM pBDT is parameterised as a function of the mass of the heavy Higgs boson by introducing the mass as a labelled input in the training [98]. Both the SM BDT and BSM pBDT are trained in the *baseline SR* with the XGBoost (Extreme Gradient Boosting) algorithm [99]. In both cases, the input variables are optimised to maximise the integral under the receiver operating characteristic (ROC) curve of each BDT.

The input variables for the SM BDT include those from the SM $t\bar{t}t\bar{t}$ search [33], except that the highest-ranked variable, the sum of the pseudo-continuous b-tagging score,[5] is not over all jets but instead only over the four jets with the highest score. The jet multiplicity is also a new input variable for the discriminant, and it is introduced to further distinguish SM $t\bar{t}t\bar{t}$ events from the other SM backgrounds. The distributions of these two variables are shown in Figure 2. Other input variables for the SM BDT are the minimum distance ΔR between two leptons among all possible pairs, the leading lepton p_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$, the p_{T} of the leading and sub-leading jets, the p_{T} of the sixth jet, the p_{T} of the leading b-jet, the scalar sum of the transverse momenta over all leptons and jets except the leading p_{T} jet, the sum of distances ΔR between two leptons for all possible pairs, and the minimum distance ΔR between a jet and a b-jet among all possible pairs. The SM BDT score is used to split the *baseline SR*: the high SM BDT region defines the *BSM SR* (SM BDT ≥ 0.55), while the low part defines the CR lowBDT (SM BDT < 0.55). The SM BDT is used as an input variable to build the BSM pBDT and as the fitted variable in the CR lowBDT.

The BSM pBDT is used in the *BSM SR*, but because there are too few events it is trained in the *baseline SR*. Since the *BSM SR* is $t\bar{t}t\bar{t}$ enriched, the training is done after reweighting the SM $t\bar{t}t\bar{t}$ events to mimic their fraction of the total background in the *BSM SR*. The other backgrounds are reweighted so that they preserve the total background yield. The SM BDT score is the most discriminating input variable for the BSM

[5] The pseudo-continuous b-tagging score is an integer from 1 to 5 assigned to a jet, based on the operating point of the b-tagging algorithm it passes, with a value of 5 assigned to the jet most similar to a b-jet [95].

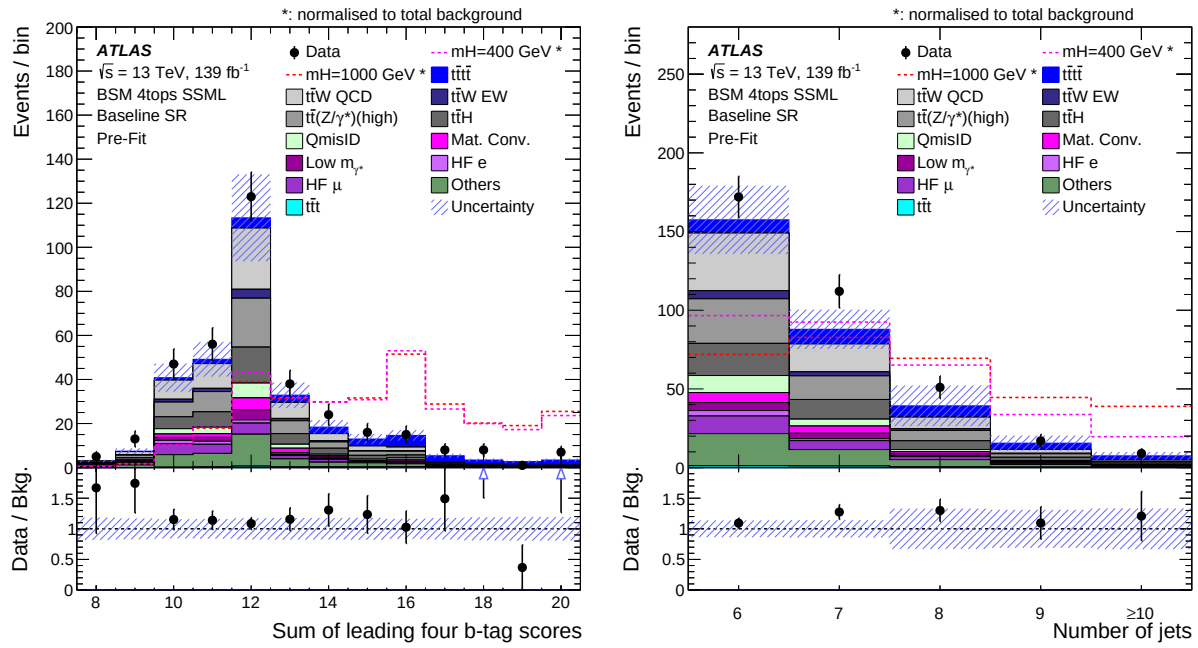
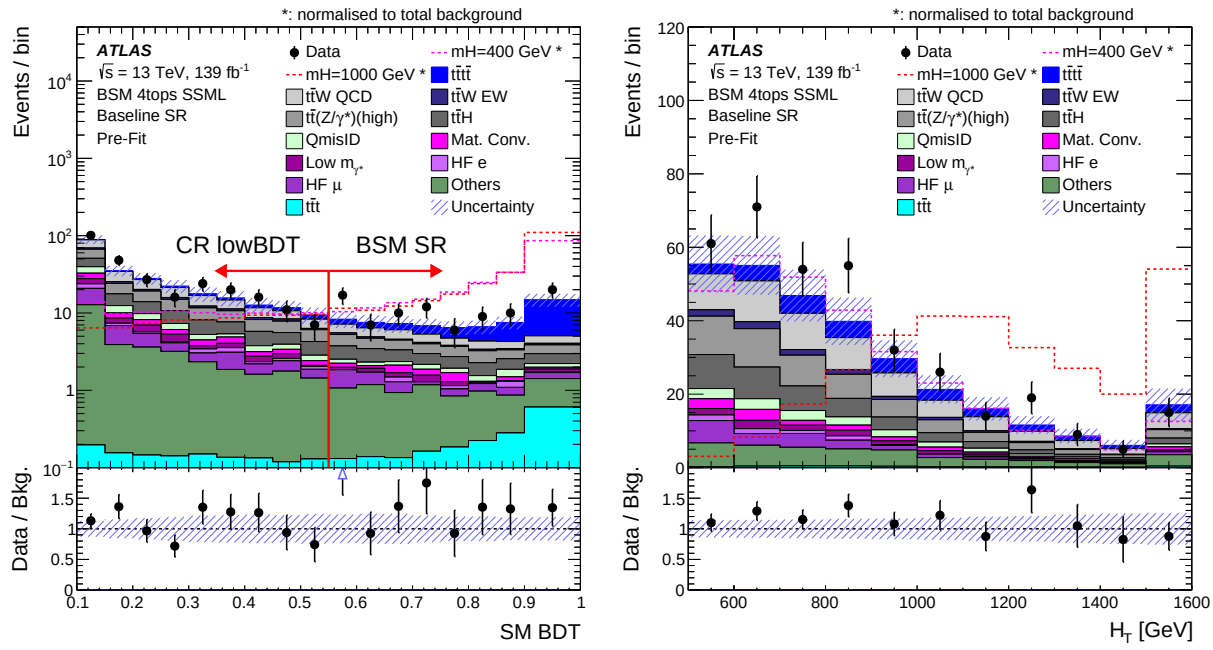


Figure 2: Pre-fit comparison between data and background in the *baseline SR* for two of the variables used as input for the SM BDT: (a) the sum of the leading four jets' b-tagging scores and (b) the number of jets. The signal samples corresponding to m_H = 400 GeV and 1000 GeV are also shown. The signal distributions are normalised to the total background to compare the shapes. The dashed band includes the total background uncertainty. The first and last bins contain underflow and overflow events, respectively. Upward arrows indicate that the value is out of the plotted range of the y-axis.

pBDT. The second most discriminating is H_{T}. The distributions of these two variables in the *baseline SR* are shown in Figure 3. Other variables used as input are the event shape variable associated with hadronic activity (sphericity) [100], the sphericity in the transverse plane, the minimum distance ΔR between two leptons among all possible pairs, the sum of distances ΔR between two leptons for all possible pairs, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ of the event. The BSM pBDT output score is used as the fitted variable in the *BSM SR*.

6 Systematic uncertainties

Systematic uncertainties arise from the reconstruction of the various physics objects and from theoretical and/or modelling uncertainties affecting the predictions for both the background and signal processes. They can be classified as either experimental uncertainties or modelling uncertainties of the signal, irreducible, and reducible backgrounds. These uncertainties manifest themselves as uncertainties in both the overall yield and shape of the final observable, and they are treated as fully correlated across all fit regions.



(a) (b)

Figure 3: Pre-fit comparison between data and background in the *baseline SR* for two of the variables used as input for the BSM pBDT: (a) the SM BDT score and (b) H_{T}. The signal samples corresponding to m_H = 400 GeV and 1000 GeV are also shown. The signal distributions are normalised to the total background to compare the shapes. The dashed band includes the total background uncertainty. The first and last bins contain underflow and overflow events, respectively. Upward arrows indicate that the value is out of the plotted range of the y-axis.

6.1 Experimental uncertainties

The uncertainty in the integrated luminosity of the combined 2015–2018 dataset is 1.7% [101], obtained using the LUCID-2 detector [102] for the primary luminosity measurements. The uncertainties related to the reweighting factors that correct the pile-up profile in simulations to match the one in data are also included.

Uncertainties in the modelling of leptons arise from their momentum resolution and scale, as well as the trigger, reconstruction, identification and isolation efficiencies used to correct for the difference between the simulation and data [41, 42, 85, 86].

Uncertainties in the modelling of jets come from their energy scale and resolution, containing the effect of jet flavour composition, single-particle response, and pile-up [90]. The uncertainty in the efficiency to pass the JVT requirement for pile-up suppression is also applied [103]. Uncertainties from the calibration of the b-tagging efficiencies, including the efficiencies of tagging b-jets as well as the rates of mis-tagging c-jets and light-flavour jets are also considered, and decomposed respectively into 45, 20 and 20 independent eigenvectors [93].

The uncertainty in $E_{\mathrm{T}}^{\mathrm{miss}}$ arising from a possible mis-calibration of its soft-track component is also included [96].

6.2 Uncertainties in modelling the signal and irreducible background

Uncertainties in modelling the signal arise from the PDF and from missing higher-order QCD corrections. The former is evaluated following the PDF4LHC recommendations [104], while the latter is estimated by varying the renormalisation and factorisation scales simultaneously by factors of 2.0 or 0.5 relative to the central value.

Several sources of uncertainty are considered for the background coming from $t\bar{t}t\bar{t}$ production. The uncertainties related to missing higher-order QCD corrections and to the PDF are evaluated in the same way as for the signal. The uncertainty associated with the choice of parton shower and hadronisation model is estimated by comparing the nominal prediction with the alternative sample generated with MadGraph5_aMC@NLO matched to Herwig instead of Pythia. The uncertainty related to the choice of generator is obtained by comparing the nominal sample with the one generated with Sherpa 2.2.10. The uncertainty associated with the lack of EW corrections in the nominal sample affects only the shape of the distributions, and it is evaluated by comparing the Sherpa 2.2.11 (QCD-only) and Sherpa 2.2.11 (QCD+EW)[6] simulated samples. A separate 20% uncertainty computed at NLO in QCD including EW corrections [34] is applied to the total cross-section.

Several sources of uncertainty affect the modelling of $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$ production. Uncertainties associated with the generator are estimated by comparing the predictions from the nominal samples with those from alternative samples generated with MadGraph5_aMC@NLO as described in Section 3. The uncertainties due to missing higher-order QCD corrections are evaluated in the same way as for the signal. An uncertainty of 12% (10%) is applied to the $t\bar{t}Z$ ($t\bar{t}H$) total cross-section [72]. A 1% uncertainty from the PDF, estimated following the PDF4LHC prescription [104], is applied to both the $t\bar{t}Z$ and $t\bar{t}H$ processes. No cross-section uncertainty is assigned to $t\bar{t}W$ QCD production because its normalisation is estimated in the fit (see Section 5.1). An uncertainty of 20% [105] is applied to the $t\bar{t}W$ EW cross-section. In order to cover the difference between data and prediction in the $t\bar{t}W$ validation region defined in Section 8, where a data excess is observed for high jet multiplicities, an additional 133% (208%) uncertainty is assigned to $t\bar{t}W$ QCD production with seven (eight or more) jets. Events arising from $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$ production can enter the signal region if they have additional heavy-flavour jets. As these processes are difficult to model in the simulation, an uncertainty of 50% is assigned to the events with an additional true b-jet and a separate 50% uncertainty is assigned to the events with two or more additional true b-jets. Those estimates are taken from Ref. [33], and treated as uncorrelated between the three different processes because of the different set-ups used to simulate the events.

The uncertainty in the $t\bar{t}t$ production cross-section is set to 100% as done in Ref. [33]. Following the same motivation given earlier about the uncertainties for processes with additional heavy-flavour jets for $t\bar{t}V$ samples, an extra uncertainty of 50% is applied to $t\bar{t}t$ events with at least one additional true b-jet.

An uncertainty of 30% is applied to the cross-sections of the tZ and tWZ single-top-quark processes [106, 107]. The uncertainty in the diboson cross-section is set to 40% following studies of WZ+b production. For the rare top-quark processes, i.e. $t\bar{t}WW$, $t\bar{t}ZZ$, $t\bar{t}WZ$, $t\bar{t}HH$, and $t\bar{t}WH$, an uncertainty of 50% is applied to the cross-section [29]. For the remaining small backgrounds, an uncertainty of 50% is assigned to the cross-section. An additional uncertainty of 50% is applied to all small backgrounds except $t\bar{t}t$. This uncertainty is applied to events with one additional true b-jet, and separately to events with at least two additional true b-jets.

[6] EW corrections are included in the Sherpa 2.2.11 sample as additional weights. Thus, when using the QCD-only version the EW weights are not used.

6.3 Uncertainties in modelling the reducible background

The uncertainties in the background due to charge misidentification arise from the uncertainties in the measurement of the electron charge misidentification rates. The following contributions are considered: the statistical uncertainty from the likelihood fit to data, the difference between the rates extracted in simulated events with the likelihood method and the true rates, and the changes in the measured rates when the size of the sideband regions used to subtract the background from the Z-peak is varied from 8 GeV to 12 GeV.

The uncertainties in the detector-material photon conversion and Low m_{γ^*} backgrounds only come from the shape of the distributions used in the template fit. No uncertainty is assigned to the normalisation of these processes since the overall normalisations of these backgrounds are estimated in the fit. The shape uncertainty is estimated in a region enriched in $Z(\rightarrow \mu\mu)+\gamma$ events by comparing data with the Powheg+Pythia 8 simulation of $Z(\rightarrow \mu\mu)+\gamma$/jets production. The uncertainty is supposed to cover the extrapolation from the 'CR Conv', i.e. events with $m_{ee}^{\text{CV}} \in [0, 0.1]$ GeV, to the regions with events with larger m_{ee}^{CV}. The uncertainty was estimated to be 50% (100%) for the detector-material photon conversion (Low m_{γ^*}) background, and it is applied to events with $m_{ee}^{\text{CV}} > 0.1$ GeV in all control and signal regions.

Similarly to the detector-material photon conversion and Low m_{γ^*} backgrounds, the only uncertainty in the heavy-flavour non-prompt lepton background comes from the shape of the distributions used in the fit. This uncertainty is estimated by a bin-by-bin comparison of the data and the post-fit background prediction in every region used in the analysis, after enhancing the contribution of non-prompt leptons. This is done by using a loose lepton selection with the isolation requirements dropped and the identification criteria relaxed. The shape uncertainty is treated as correlated among all regions. It is derived separately for electrons and muons.

The uncertainty assigned to the normalisation of the background coming from light-flavour non-prompt leptons is 100%. This value was derived from the level of agreement between data and simulation in a region enriched in light-flavour non-prompt leptons. For the remaining sources of fake/non-prompt leptons, an uncertainty of 30% is applied as done in Ref. [33].

Other uncertainties affecting the shape of the distributions of the background coming from fake/non-prompt leptons arise from the modelling of the heavy-flavour content of the main source of reducible background, $t\bar{t}$+jets production. Based on the measurement of $t\bar{t}$ production with additional heavy-flavour jets [108], an uncertainty of 30% is assigned to events with three true b-jets, and a separate 30% uncertainty is assigned to events with at least four true b-jets.

7 Statistical analysis

The presence of a signal is tested for by fitting the BSM pBDT score in the *BSM SR* jointly with the control regions as described in Section 5. The statistical analysis uses a binned likelihood function $\mathcal{L}(\mu, \boldsymbol{\theta}, \boldsymbol{\lambda})$ constructed as a product of Poisson probability terms over all bins considered in the search. This function depends on the following: the signal-strength parameter μ, defined as a factor multiplying the expected yield of $t\bar{t}H/A(\rightarrow t\bar{t})$ signal events; $\boldsymbol{\theta}$, a set of nuisance parameters that encode the effect of systematic uncertainties on the signal and background expectations; and $\boldsymbol{\lambda}$, a set of multiplicative factors that normalise the background expectations. All nuisance parameters are subject to Gaussian constraints in the likelihood. Therefore, the expected total number of events in a given bin depends on μ, $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$.

For a given value of μ, the nuisance parameters $\boldsymbol{\theta}$ and the normalisation factors $\boldsymbol{\lambda}$ allow variations of the expectations for signal and background according to the corresponding systematic uncertainties, and their fitted values result in the deviations from the nominal expectations that globally provide the best fit to the data. This procedure allows a reduction of the impact of systematic uncertainties on the search sensitivity by taking advantage of the highly populated background-dominated bins included in the likelihood fit. Simulation statistical uncertainties in each bin are taken into account by dedicated parameters in the fit, which are modelled with Poisson constraints. The best-fit branching fraction is obtained by performing a binned likelihood fit to the data under the signal-plus-background hypothesis, maximising the likelihood function $\mathcal{L}(\mu, \boldsymbol{\theta}, \boldsymbol{\lambda})$ over μ, $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$.

The test statistic q_μ is defined as the profile likelihood ratio, $q_\mu = -2\ln(\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}_\mu, \hat{\boldsymbol{\lambda}}_\mu)/\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}}_{\hat{\mu}}, \hat{\boldsymbol{\lambda}}_{\hat{\mu}}))$, where $\hat{\mu}$, $\hat{\boldsymbol{\theta}}_{\hat{\mu}}$ and $\hat{\boldsymbol{\lambda}}_{\hat{\mu}}$ are the values of the parameters that maximise the likelihood function (subject to the constraint $0 \leq \hat{\mu} \leq \mu$); $\hat{\boldsymbol{\theta}}_\mu$ and $\hat{\boldsymbol{\lambda}}_\mu$ are respectively the values of the nuisance parameters and normalisation factors that maximise the likelihood function for a given value of μ. The test statistic q_μ is evaluated with the RooFit package [109, 110]. A related test statistic is used to determine whether the observed data is compatible with the background-only hypothesis by setting $\mu = 0$ in the profile likelihood ratio and leaving $\hat{\mu}$ unconstrained: $q_0 = -2\ln(\mathcal{L}(0, \hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\lambda}}_0)/\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}))$. The p-value representing the level of agreement between the data and the background-only hypothesis is estimated by integrating the distribution of q_0, based on the asymptotic formulae in Ref. [111], above the q_0 value observed in the data. Upper limits on μ are derived by using q_μ in the $\mathrm{CL_s}$ method [112, 113]. For a given signal scenario, values of μ yielding $\mathrm{CL_s} < 0.05$, where $\mathrm{CL_s}$ is computed using the asymptotic approximation [111], are excluded at ≥95% confidence level (CL).

8 Results and interpretation

The signal strength for every signal hypothesis and the normalisation factors for the background processes described in Section 5.1.1 are determined via a binned likelihood fit performed simultaneously in all signal and control regions defined in Table 1.

For the *BSM SR*, the binning of the BSM pBDT distribution is optimised for every signal hypothesis to provide the best discrimination between the tested signal and the background, avoiding the presence of bins with no contribution from the major backgrounds. The results obtained for the background-only fit shown in this section are the ones obtained when the fitted distribution in the signal region is the BSM pBDT score assigned to the background events when the signal hypothesis is $m_H = 400$ GeV. Similar results are obtained for other signal hypotheses.

The normalisation factors from the background-only fit to data for the different background processes are shown in Table 2. The data and post-fit background comparison for the distributions of the discriminating variables fitted in the control regions is shown in Figure 4. Good agreement between data and post-fit background is observed. Pre-fit and post-fit background yields in the *BSM SR* are shown in Table 3. The signal yields corresponding to the pre-fit estimate in a 2HDM with $\tan\beta = 1$ are shown for two signal hypotheses. The total post-fit background yield is less than one standard deviation from the SM prediction. For the $t\bar{t}W$ QCD background, the large difference between the pre-fit and post-fit yields arises not only from the change in its normalisation factor $\lambda_{t\bar{t}W\,\mathrm{QCD}}$, but also from several nuisance parameters associated with this process, which are found to be pulled from their nominal values in the fit. These are the ones related to the generator, the scale variations, and the systematic uncertainty assigned to events with eight or

more jets. The measured normalisation factor $\lambda_{t\bar{t}W\,\mathrm{QCD}}$ is found to agree with that in Ref. [97]. No other nuisance parameters were found to be significantly pulled or constrained in the fit.

The modelling of the $t\bar{t}W$ QCD background was validated by using the charge asymmetry of $t\bar{t}W$ production. The $t\bar{t}W$ validation region is defined to have at least four jets, at least two of which are b-tagged, and the charges of selected leptons are examined. The difference between the number of events with a positive sum of lepton charges and the number of events with a negative sum is calculated, and the SM BDT distribution of this difference is shown in Figure 5. Good agreement between the data and post-fit background is observed within the uncertainties.

Table 2: Normalisation factors for the different background processes obtained from the background-only fit. The fit is done simultaneously in all signal and control regions. The fitted distribution in the *BSM SR* is the BSM pBDT score assigned to the background events when the signal hypothesis is $m_H = 400$ GeV. The uncertainties include both statistical and systematic uncertainties.

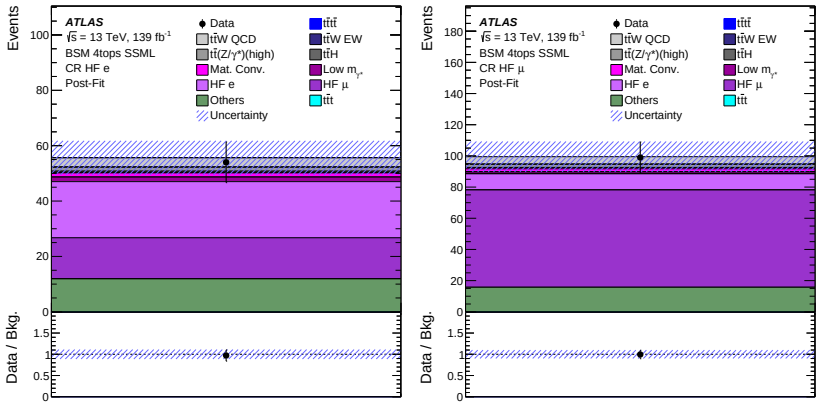
Parameter	$\lambda_{t\bar{t}W\,\mathrm{QCD}}$	$\lambda_{\mathrm{Mat.\ Conv.}}$	$\lambda_{\mathrm{Low}\ m_{\gamma^*}}$	$\lambda_{\mathrm{HF}\ e}$	$\lambda_{\mathrm{HF}\ \mu}$
Value	1.3 ± 0.3	1.5 ± 0.5	0.6 ± 0.5	0.9 ± 0.4	1.0 ± 0.2

The different post-fit BSM pBDT distributions corresponding to the signal benchmarks $m_H = 400$ and 1000 GeV in the *BSM SR* for the background-only fit are shown in Figure 6. Under the signal-plus-background hypothesis the best-fit signal cross-section ranges between 4^{+6}_{-5} fb and 2^{+2}_{-2} fb for a heavy Higgs boson with mass between 400 GeV and 1000 GeV, respectively. No significant excess of events above the SM prediction is observed. The results are then interpreted in the context of a type-II 2HDM. The observed and expected upper limits on the $t\bar{t}H/A$ cross-section times branching fraction of $H/A \to t\bar{t}$ at 95% CL as a function of $m_{H/A}$ are shown in Figure 7. The observed upper limits range between 14 fb and 6 fb for the studied mass range. These upper limits on the cross-section can be translated into limits in the $\tan\beta$ vs $m_{H/A}$ plane. Two different scenarios are considered: one where the scalar H and pseudo-scalar A bosons have equal masses and both contribute to BSM $t\bar{t}t\bar{t}$ production, and the other where only the scalar H boson contributes. The observed and expected 95% CL exclusion regions for these scenarios are shown in Figure 8. If both particles contribute to BSM $t\bar{t}t\bar{t}$ production, the excluded values of $\tan\beta$ range between 1.6 and 0.6 for $m_{H/A}$ between 400 GeV and 1000 GeV, respectively. When considering only the scalar H boson, the values of $\tan\beta$ below 1.2 and 0.5 are excluded for the same mass range. In the parameter space studied, the limits for the pseudo-scalar A boson alone are similar to those for the scalar H boson.

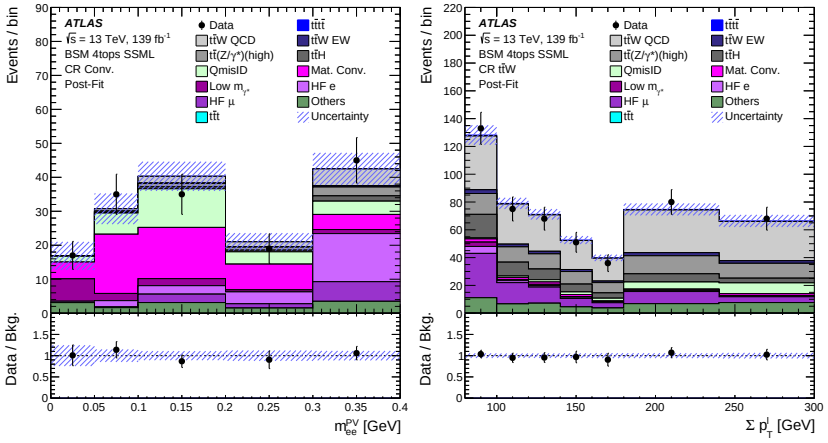
The robustness of the results was checked by reproducing the fit results using data split by data-taking period, or by splitting the *BSM SR* into events in the same-sign dilepton channel and events in the trilepton channel. An additional test was performed by using same-sign dilepton events that are only positively charged or only negatively charged. The tests show that the response of the fit is stable with regard to the fitted nuisance parameters and signal strength.

The impact of the different sources of systematic uncertainty on the signal strength is summarised in Table 4. The systematic uncertainties having the highest impact on the best-fit μ are the ones associated with the dominant backgrounds in the *BSM SR*. Thus, the highest-ranked systematic uncertainties are the ones related to the modelling of SM $t\bar{t}t\bar{t}$ and $t\bar{t}W$.

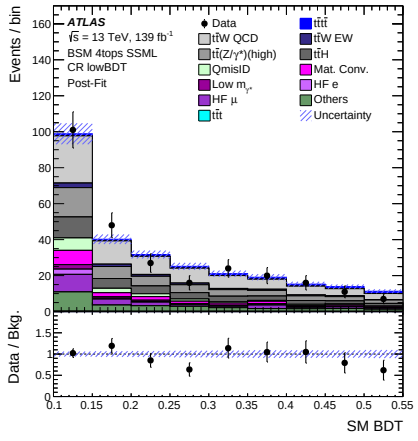
All of the results described above were obtained with the $t\bar{t}t\bar{t}$ contribution constrained to the SM prediction. The binned likelihood fit was also performed with the goal of measuring the normalisation factor for $t\bar{t}t\bar{t}$ events assuming no contribution from the BSM $t\bar{t}t\bar{t}$ signal, and by fitting the SM BDT distribution



(a) CR HF e (b) CR HF μ



(c) CR Conv (d) CR $t\bar{t}W$



(e) CR lowBDT

Figure 4: Data and post-fit background comparison for the distributions of the discriminating variables fitted in the control regions obtained with the background-only fit. The fit is done simultaneously in all signal and control regions. The fitted distribution in the *BSM SR* is the BSM pBDT score assigned to the background events when the signal hypothesis is m_H = 400 GeV. The band includes the total uncertainty of the post-fit estimate. The lower panel shows the ratio of the data to the post-fit background. The first and last bins of (c), (d) and (e) contain underflow and overflow events, respectively.

Table 3: Pre-fit and post-fit background yields in the *BSM SR* obtained with the background-only fit. The fit is done simultaneously in all signal and control regions. The fitted distribution in the *BSM SR* is the BSM pBDT score assigned to the background events when the signal hypothesis is $m_H = 400$ GeV. The signal yields correspond to the pre-fit estimate in the type-II 2HDM for $\tan\beta = 1$ and for $m_H = 400$ GeV and 1000 GeV, assuming only the production of a new scalar H boson. The number of data events is also shown. The total systematic uncertainty differs from the sum in quadrature of the different uncertainties because of correlations.

Process	Pre-fit	Post-fit
$t\bar{t}t\bar{t}$	22.3 ± 5.3	25.8 ± 5.3
$t\bar{t}W$ QCD	9.4 ± 9.6	17.2 ± 6.9
$t\bar{t}W$ EW	1.3 ± 0.5	1.4 ± 0.6
$t\bar{t}WW$	1.8 ± 1.0	1.9 ± 1.0
$t\bar{t}(Z/\gamma^*)$ (high mass)	8.5 ± 2.1	9.2 ± 2.2
$t\bar{t}H$	7.2 ± 1.7	7.8 ± 1.7
QmisID	2.1 ± 0.2	2.1 ± 0.2
Mat. Conv.	1.8 ± 0.6	3.0 ± 1.2
Low m_{γ^*}	1.2 ± 0.6	0.8 ± 0.8
HF e	0.6 ± 0.7	0.6 ± 0.7
HF μ	2.7 ± 1.1	2.9 ± 1.2
LF	1.1 ± 1.2	0.4 ± 1.0
Other fake	1.1 ± 0.7	1.3 ± 0.7
tZ, tWZ	0.9 ± 0.3	0.9 ± 0.3
VV, VH, VVV	0.3 ± 0.1	0.3 ± 0.1
$t\bar{t}t$	1.9 ± 1.9	2.3 ± 2.1
$t\bar{t}WZ, t\bar{t}ZZ, t\bar{t}WH, t\bar{t}HH$	1.3 ± 0.7	1.4 ± 0.8
Total background	65.6 ± 13.4	79.5 ± 6.7
$t\bar{t}H(\rightarrow t\bar{t})$, $m_H = 400$ GeV	38.6 ± 2.4	–
$t\bar{t}H(\rightarrow t\bar{t})$, $m_H = 1000$ GeV	4.4 ± 0.2	–
Data	91	

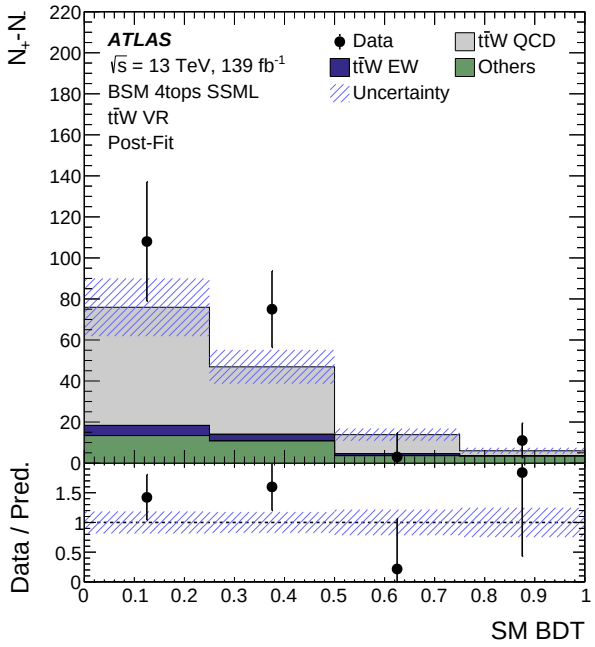
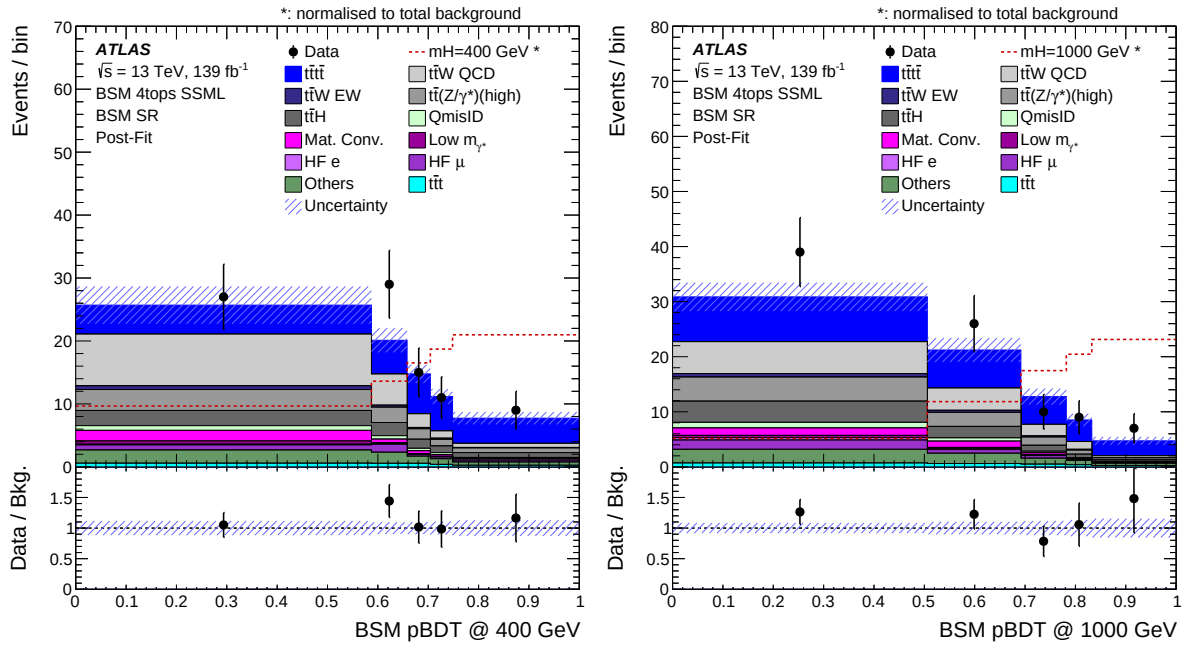


Figure 5: Data and post-fit background comparison obtained with the background-only fit in the validation region defined for $t\bar{t}W$ events. The fit is done simultaneously in all signal and control regions. The y-axis shows the difference between the number of events with a positive sum of the charges of the selected leptons and the number of events with a negative sum. The band includes the total uncertainty of the post-fit estimate. The lower panel shows the ratio of the data to the post-fit background.

in the *baseline SR*. The best-fit value is found to be compatible with the observation from the previous ATLAS $t\bar{t}t\bar{t}$ cross-section measurement [33].

Table 4: Post-fit impact of the different systematic uncertainties on the signal strength μ assuming $m_H = 400$ GeV, grouped in categories. For each uncertainty source, the fit is repeated with the corresponding group of nuisance parameters fixed to their best-fit values. The square of each group's contribution is evaluated as the difference of the squares of the full-fit uncertainty and the uncertainty obtained in the repeated fit. The contributions from individual groups are compared with the total systematic uncertainty and the statistical uncertainty. The total uncertainty is different from the sum in quadrature of the components due to correlations among nuisance parameters in the fit.

Uncertainty source	$\Delta\mu$	
Signal modelling		
$t\bar{t}H(\rightarrow t\bar{t})$	+0.01	−0.00
Background modelling		
$t\bar{t}t\bar{t}$	+0.17	−0.17
$t\bar{t}W$	+0.07	−0.07
$t\bar{t}t$	+0.06	−0.05
Non-prompt leptons	+0.05	−0.05
$t\bar{t}Z$	+0.05	−0.05
$t\bar{t}H$	+0.03	−0.03
Other background	+0.03	−0.02
Instrumental		
Jet uncertainties	+0.12	−0.09
Jet flavour tagging (b-jets)	+0.05	−0.04
Jet flavour tagging (light-flavour jets)	+0.04	−0.03
Luminosity	+0.03	−0.02
Jet flavour tagging (c-jets)	+0.02	−0.02
Other experimental uncertainties	+0.02	−0.02
MC statistical uncertainty		
Simulation sample size	+0.04	−0.04
Total systematic uncertainty	+0.31	−0.28
Statistical		
HF, Mat. Conv., and Low $m_{\gamma*}$ normalisation	+0.05	−0.04
$t\bar{t}W$ QCD normalisation	+0.05	−0.04
Total statistical uncertainty	+0.35	−0.32
Total uncertainty	+0.46	−0.41



(a) (b)

Figure 6: Data and post-fit background comparison obtained with the background-only fit to the *BSM SR* for the BSM pBDT distribution used for (a) $m_H = 400$ GeV and (b) $m_H = 1000$ GeV. The fit is done simultaneously in all signal and control regions. The band includes the total uncertainty of the post-fit estimate. The respective signal hypothesis is also shown. The signal is normalised to the total background for better visibility. The lower panel shows the ratio of the data to the post-fit background. The binning of the BSM pBDT is optimised for every signal hypothesis to provide the best discrimination between the tested signal and the background, avoiding the presence of bins with no contribution from the major backgrounds.

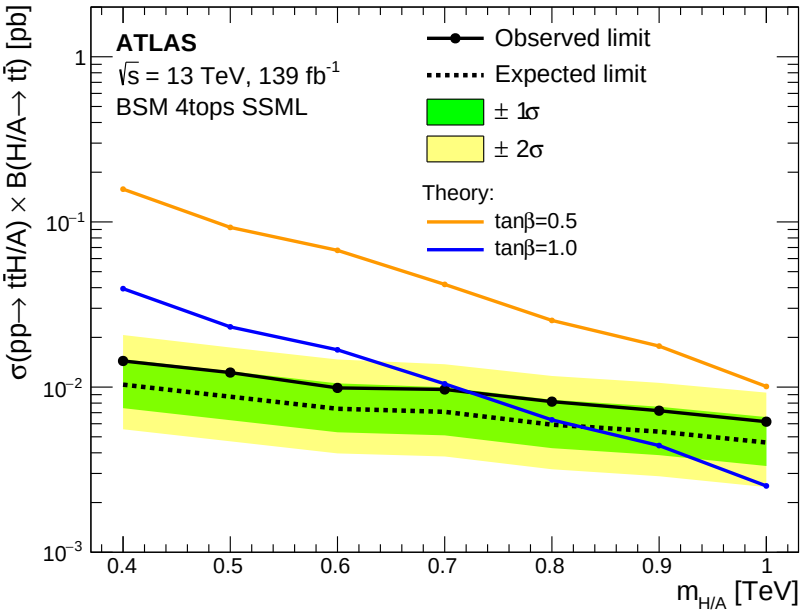
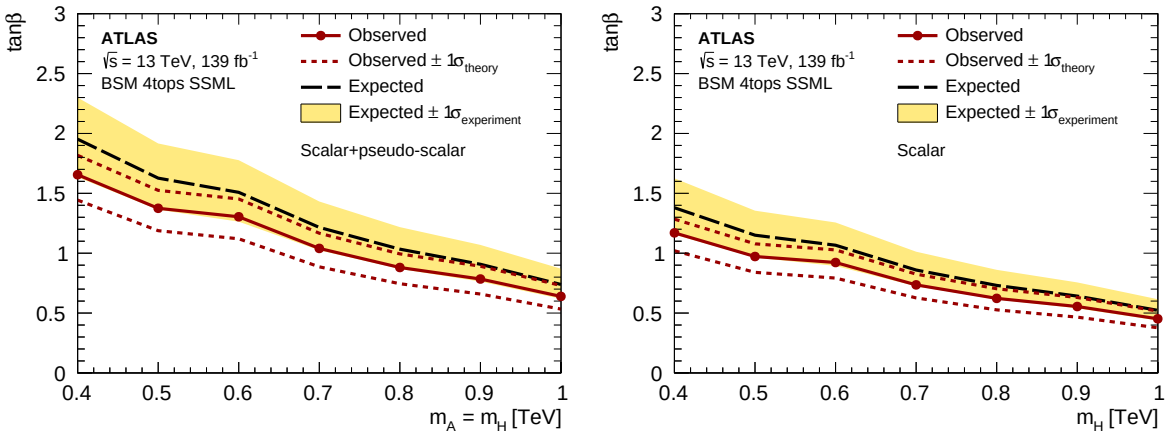


Figure 7: Observed (black solid line) and expected (black dashed line) 95% CL upper limits on the $t\bar{t}H/A$ cross-section times branching fraction of $H/A \to t\bar{t}$ as a function of $m_{H/A}$. The limits are estimated assuming that both a heavy scalar H boson and a pseudo-scalar A boson contribute to the $t\bar{t}t\bar{t}$ final state and have the same mass $m_H = m_A$. The green (yellow) band shows the $\pm 1\sigma$ ($\pm 2\sigma$) variation of the expected limits. Theoretical predictions for two values of $\tan\beta$ are shown.



(a) (b)

Figure 8: Observed (red line) and expected (black dashed line) exclusion regions at 95% CL in the $\tan\beta$ vs $m_{H/A}$ plane (a) assuming that both a heavy scalar H boson and heavy pseudo-scalar A boson contribute to the $t\bar{t}t\bar{t}$ final state and have the same mass $m_H = m_A$ or (b) assuming that only the scalar H boson contributes. The limits for the pseudo-scalar A boson alone are similar to those in (b). The yellow band shows the $\pm 1\sigma$ variation of the expected exclusion limit. The region below the red solid line is excluded at 95% CL. The exclusion regions are derived in the context of a type-II 2HDM. Variations of the theory cross-section are also shown as dashed red lines.

9 Conclusion

A search for a new heavy Higgs boson in the process $t\bar{t}H/A \rightarrow t\bar{t}t\bar{t}$ in the same-sign dilepton channel and the multilepton channel is presented. The search makes use of 139 fb^{-1} of proton–proton collision data collected by the ATLAS detector at the LHC at a centre-of-mass energy of 13 TeV.

No significant excess of events over the Standard Model expectation is observed. The results are then interpreted in the context of the type-II 2HDM, assuming no interference with the Standard Model production of four-top-quarks. The expected sensitivity of the present $t\bar{t}H/A(\rightarrow t\bar{t})$ search exceeds that of the previous ATLAS $t\bar{t}H/A(\rightarrow t\bar{t})$ search with 36 fb^{-1} [29] by about a factor of four. The observed (expected) upper limits at 95% CL on the $t\bar{t}H/A$ cross-section times branching ratio of $H/A \rightarrow t\bar{t}$ range between 14 (10) fb and 6 (5) fb for a heavy Higgs boson with mass between 400 GeV and 1000 GeV, respectively. These upper limits on the cross-section are translated into limits in the $\tan\beta$ vs $m_{H/A}$ plane. Assuming that only one particle, either the scalar H boson or the pseudo-scalar A boson, contributes to the $t\bar{t}t\bar{t}$ final state, the values of $\tan\beta$ below 1.2 or 0.5 are excluded for a mass of 400 GeV or 1000 GeV, respectively. These exclusion ranges increase to $\tan\beta$ below 1.6 or 0.6 when both particles are considered.

References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex].

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex].

[3] ATLAS Collaboration, *Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton–proton collision data at $\sqrt{s}$ = 13 TeV collected with the ATLAS experiment*, Phys. Rev. D **101** (2020) 012002, arXiv: 1909.02845 [hep-ex].

[4] CMS Collaboration, *Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 421, arXiv: 1809.10733 [hep-ex].

[5] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321, ed. by J. C. Taylor.

[6] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[7] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508, ed. by J. C. Taylor.

[8] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585, ed. by J. C. Taylor.

[9] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156.

[10] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (1967) 1554, ed. by J. C. Taylor.

[11] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) 1, arXiv: 1106.0034 [hep-ph].

[12] P. Fayet, *Supersymmetry and Weak, Electromagnetic and Strong Interactions*, Phys. Lett. B **64** (1976) 159.

[13] P. Fayet, *Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions*, Phys. Lett. B **69** (1977) 489.

[14] J. E. Kim, *Light Pseudoscalars, Particle Physics and Cosmology*, Phys. Rept. **150** (1987) 1.

[15] ATLAS Collaboration, *Search for a heavy Higgs boson decaying into a Z boson and another heavy Higgs boson in the $\ell\ell bb$ and $\ell\ell WW$ final states in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 396, arXiv: 2011.05639 [hep-ex].

[16] ATLAS Collaboration, *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, (2021), arXiv: 2112.11876 [hep-ex].

[17] ATLAS Collaboration, *Search for resonances decaying into photon pairs in* $139\,fb^{-1}$ *of* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. B **822** (2021) 136651, arXiv: 2102.13405 [hep-ex].

[18] ATLAS Collaboration, *Search for heavy resonances decaying into a pair of* Z *bosons in the* $\ell^+\ell^-\ell'^+\ell'^-$ *and* $\ell^+\ell^-\nu\bar{\nu}$ *final states using* $139\,fb^{-1}$ *of proton–proton collisions at* $\sqrt{s} = 13$, *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 332, arXiv: 2009.14791 [hep-ex].

[19] ATLAS Collaboration, *Search for Heavy Higgs Bosons Decaying into Two Tau Leptons with the ATLAS Detector Using* pp *Collisions at* $\sqrt{s} = 13$ *TeV*, Phys. Rev. Lett. **125** (2020) 051801, arXiv: 2002.12223 [hep-ex].

[20] ATLAS Collaboration, *Search for heavy neutral Higgs bosons produced in association with* b*-quarks and decaying into* b*-quarks at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. D **102** (2020) 032004, arXiv: 1907.02749 [hep-ex].

[21] CMS Collaboration, *Search for a heavy Higgs boson decaying to a pair of* W *bosons in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **03** (2020) 034, arXiv: 1912.01594 [hep-ex].

[22] CMS Collaboration, *Search for new neutral Higgs bosons through the* $H \to ZA \to \ell^+\ell^- b\bar{b}$ *process in* pp *collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **03** (2020) 055, arXiv: 1911.03781 [hep-ex].

[23] CMS Collaboration, *Search for a heavy pseudoscalar Higgs boson decaying into a* 125 *GeV Higgs boson and a* Z *boson in final states with two tau and two light leptons at* $\sqrt{s} = 13$ *TeV*, JHEP **03** (2020) 065, arXiv: 1910.11634 [hep-ex].

[24] CMS Collaboration, *Search for MSSM Higgs bosons decaying to* $\mu^+\mu^-$ *in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Phys. Lett. B **798** (2019) 134992, arXiv: 1907.03152 [hep-ex].

[25] CMS Collaboration, *Search for a heavy pseudoscalar boson decaying to a* Z *and a Higgs boson at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79** (2019) 564, arXiv: 1903.00941 [hep-ex].

[26] ATLAS Collaboration, *Search for heavy Higgs bosons* A/H *decaying to a top quark pair in* pp *collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, Phys. Rev. Lett. **119** (2017) 191803, arXiv: 1707.06025 [hep-ex].

[27] CMS Collaboration, *Search for heavy Higgs bosons decaying to a top quark pair in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **04** (2020) 171, arXiv: 1908.01115 [hep-ex].

[28] N. Craig, J. Hajer, Y.-Y. Li, T. Liu and H. Zhang, *Heavy Higgs bosons at low* $\tan\beta$*: from the LHC to 100 TeV*, JHEP **01** (2017) 018, arXiv: 1605.08744 [hep-ph].

[29] ATLAS Collaboration, *Search for new phenomena in events with same-charge leptons and* b*-jets in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **12** (2018) 039, arXiv: 1807.11883 [hep-ex].

[30] CMS Collaboration, *Search for production of four top quarks in final states with same-sign or multiple leptons in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **80** (2020) 75, arXiv: 1908.06463 [hep-ex].

[31] CMS Collaboration, *Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **11** (2019) 082, arXiv: 1906.02805 [hep-ex].

[32] ATLAS Collaboration, *Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **11** (2021) 118, arXiv: 2106.11683 [hep-ex].

[33] ATLAS Collaboration, *Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 1085, arXiv: 2007.14858 [hep-ex].

[34] R. Frederix, D. Pagani and M. Zaro, *Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions*, JHEP **02** (2018) 031, arXiv: 1711.02116 [hep-ph].

[35] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[36] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: https://cds.cern.ch/record/1291633, Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: https://cds.cern.ch/record/1451888.

[37] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det].

[38] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex].

[39] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187.

[40] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det].

[41] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47, arXiv: 1909.00761 [hep-ex].

[42] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, JINST **15** (2020) P09015, arXiv: 2004.13447 [hep-ex].

[43] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det].

[44] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[45] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph].

[46] S. Frixione, P. Nason and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph].

[47] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146.

[48] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph].

[49] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph].

[50] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: 1009.2450 [hep-ph].

[51] R. Frederix, E. Re and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, JHEP **09** (2012) 130, arXiv: 1207.5391 [hep-ph].

[52] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09** (2009) 111, arXiv: 0907.4076 [hep-ph], Erratum: JHEP **02** (2010) 011.

[53] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: 1410.3012 [hep-ph].

[54] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419.

[55] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph].

[56] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152.

[57] S. Frixione, E. Laenen, P. Motylinski and B. R. Webber, *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*, JHEP **04** (2007) 081, arXiv: hep-ph/0702198.

[58] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015, arXiv: 1212.3460 [hep-ph].

[59] M. Bähr et al., *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639, arXiv: 0803.0883 [hep-ph].

[60] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv: 1512.01178 [hep-ph].

[61] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. C **75** (2015) 204, arXiv: 1412.3989 [hep-ph].

[62] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph].

[63] C. Gütschow, J. M. Lindert and M. Schönherr, *Multi-jet merged top-pair production including electroweak corrections*, Eur. Phys. J. C **78** (2018) 317, arXiv: 1803.00950 [hep-ph].

[64] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph].

[65] S. Höche, F. Krauss, M. Schönherr and F. Siegert,
A critical appraisal of NLO+PS matching methods, JHEP **09** (2012) 049,
arXiv: 1111.1220 [hep-ph].

[66] S. Höche, F. Krauss, M. Schönherr and F. Siegert,
QCD matrix elements + parton showers. The NLO case, JHEP **04** (2013) 027,
arXiv: 1207.5030 [hep-ph].

[67] S. Catani, F. Krauss, B. R. Webber and R. Kuhn, *QCD Matrix Elements + Parton Showers*,
JHEP **11** (2001) 063, arXiv: hep-ph/0109231.

[68] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*,
JHEP **05** (2009) 053, arXiv: 0903.1219 [hep-ph].

[69] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*,
Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph].

[70] A. Denner, S. Dittmaier and L. Hofer,
Collier: A fortran-based complex one-loop library in extended regularizations,
Comput. Phys. Commun. **212** (2017) 220, arXiv: 1604.06792 [hep-ph].

[71] F. Buccioni et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866, arXiv: 1907.13071 [hep-ph].

[72] D. de Florian et al.,
Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector, (2016),
arXiv: 1610.07922 [hep-ph].

[73] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
ATL-PHYS-PUB-2016-020, 2016, URL: https://cds.cern.ch/record/2216168.

[74] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber,
Single-top hadroproduction in association with a W boson, JHEP **07** (2008) 029,
arXiv: 0805.3067 [hep-ph].

[75] M. L. Ciccolini, S. Dittmaier and M. Krämer,
Electroweak radiative corrections to associated WH and ZH production at hadron colliders,
Phys. Rev. D **68** (2003) 073003, arXiv: hep-ph/0306234 [hep-ph].

[76] O. Brein, A. Djouadi and R. Harlander,
NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders,
Phys. Lett. B **579** (2004) 149, arXiv: hep-ph/0307206.

[77] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke,
Top-Quark Mediated Effects in Hadronic Higgs-Strahlung, Eur. Phys. J. C **72** (2012) 1868,
arXiv: 1111.0761 [hep-ph].

[78] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke,
Gluon-induced Higgs-strahlung at next-to-leading order QCD, JHEP **02** (2013) 078,
arXiv: 1211.5015 [hep-ph].

[79] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*,
Comput. Phys. Commun. **195** (2015) 161, arXiv: 1412.5390 [hep-ph].

[80] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo – Higgs Strahlung at hadron colliders*,
Comput. Phys. Commun. **184** (2013) 998, arXiv: 1210.5347 [hep-ph].

[81] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke,
Soft gluon resummation for gluon-induced Higgs Strahlung, JHEP **11** (2014) 082,
arXiv: 1410.0217 [hep-ph].

[82] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*,
Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

[83] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*,
ATL-PHYS-PUB-2016-017, 2016, URL: https://cds.cern.ch/record/2206965.

[84] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*,
ATL-PHYS-PUB-2015-026, 2015, URL: https://cds.cern.ch/record/2037717.

[85] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006,
arXiv: 1908.00005 [hep-ex].

[86] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **81** (2021) 578,
arXiv: 2012.00578 [hep-ex].

[87] ATLAS Collaboration,
Jet reconstruction and performance using particle flow with the ATLAS Detector,
Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex].

[88] M. Cacciari, G. P. Salam and G. Soyez, *The anti-*k_t *jet clustering algorithm*, JHEP **04** (2008) 063,
arXiv: 0802.1189 [hep-ph].

[89] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896,
arXiv: 1111.6097 [hep-ph].

[90] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 689,
arXiv: 2007.02645 [hep-ex].

[91] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*,
ATLAS-CONF-2014-018, 2014, URL: https://cds.cern.ch/record/1700870.

[92] ATLAS Collaboration,
Selection of jets produced in 13 *TeV proton–proton collisions with the ATLAS detector*,
ATLAS-CONF-2015-029, 2015, URL: https://cds.cern.ch/record/2037702.

[93] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017,
URL: https://cds.cern.ch/record/2273281.

[94] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017,
URL: https://cds.cern.ch/record/2255226.

[95] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with* $t\bar{t}$ *events in pp collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79** (2019) 970,
arXiv: 1907.05120 [hep-ex].

[96] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **78** (2018) 903,
arXiv: 1802.08168 [hep-ex].

[97] ATLAS Collaboration, *Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **99** (2019) 072009, arXiv: 1901.03584 [hep-ex].

[98] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson, *Parameterized neural networks for high-energy physics*, Eur. Phys. J. C **76** (2016) 235, arXiv: 1601.07913 [hep-ex].

[99] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016), URL: http://dx.doi.org/10.1145/2939672.2939785.

[100] J. D. Bjorken and S. J. Brodsky, *Statistical Model for Electron-Positron Annihilation into Hadrons*, Phys. Rev. D **1** (1970) 1416, URL: https://link.aps.org/doi/10.1103/PhysRevD.1.1416.

[101] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: https://cds.cern.ch/record/2677054.

[102] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017.

[103] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 581, arXiv: 1510.03823 [hep-ex].

[104] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph].

[105] F. F. Cordero, M. Kraus and L. Reina, *Top-quark pair production in association with a $W^{\pm}$ gauge boson in the POWHEG-BOX*, Phys. Rev. D **103** (2021) 094014, URL: https://link.aps.org/doi/10.1103/PhysRevD.103.094014.

[106] ATLAS Collaboration, *Measurement of the production cross-section of a single top quark in association with a Z boson in proton–proton collisions at 13 TeV with the ATLAS detector*, Phys. Lett. B **780** (2018) 557, arXiv: 1710.03659 [hep-ex].

[107] F. Demartin, B. Maier, F. Maltoni, K. Mawatari and M. Zaro, *tWH associated production at the LHC*, Eur. Phys. J. C **77** (2017) 34, arXiv: 1607.05862 [hep-ph].

[108] 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 [hep-ex].

[109] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, arXiv: physics/0306116 [physics.data-an].

[110] W. Verkerke and D. Kirkby, *RooFit Users Manual*, URL: http://roofit.sourceforge.net.

[111] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, arXiv: 1007.1727 [physics.data-an], Erratum: Eur. Phys. J. C **73** (2013) 2501.

[112] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instrum. Meth. A **434** (1999) 435, arXiv: hep-ex/9902006.

[113] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693.