ATLAS CONF Note

ATLAS-CONF-2024-001

16th March 2024

Search for heavy neutral Higgs bosons decaying to a top quark pair in 140 fb^{-1} of proton–proton collision data at $\sqrt{s}$ = 13 TeV

The ATLAS Collaboration

A search for heavy pseudo-scalar (A) and scalar (H) Higgs bosons decaying into a top quark pair ($t\bar{t}$) has been performed with 140 fb^{-1} of proton–proton collision data collected by the ATLAS experiment at the Large Hadron Collider at a center-of-mass energy of $\sqrt{s}$ = 13 TeV. Interference effects between the signal process and Standard Model (SM) $t\bar{t}$ production are taken into account. Final states with exactly one or exactly two electrons or muons are considered. No significant deviation from the SM prediction is observed. The results of the search are interpreted in the context of a Two-Higgs-Doublet Model (2HDM) of type II in the alignment limit with mass-degenerate pseudo-scalar and scalar Higgs bosons ($m_A = m_H$) and the minimal supersymmetric standard model scenario (hMSSM). Ratios of the two vacuum expectation values, $\tan\beta$, of smaller than 3.49 (3.16) are observed to be excluded for $m_A = m_H$ = 400 GeV in the 2HDM (hMSSM). Masses up to 1240 GeV (1210 GeV) are observed (expected) to be excluded for the lowest tested $\tan\beta$ value of 0.4 in the 2HDM. In the hMSSM, masses up to 950 GeV (830 GeV) are observed (expected) to be excluded for $\tan\beta$ = 1.0. In addition, generic exclusion limits are derived separately for single scalar and pseudo-scalar states for different choices of their mass and total width.

ATLAS-CONF-2024-001
17 March 2024

© 2024 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

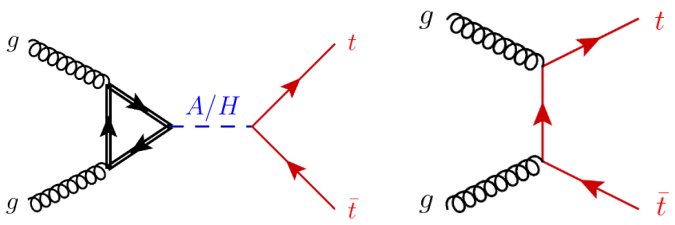
1 Introduction

New massive scalar or pseudo-scalar states with strong couplings to the top quark are predicted in numerous extensions of the Standard Model (SM) of Particle Physics. These include models with an extended Higgs sector, such as Two-Higgs-Doublet Models (2HDMs) [1–4], which postulate the existence of a second complex Higgs doublet. 2HDMs are motivated, for example, by supersymmetry [5–10] and axion models [11]. Further examples of models with an extended Higgs sector include models predicting a new electroweak singlet [12], a combination of singlet and doublet fields [13, 14], or Three-Higgs-Doublet Models (3HDMs) [15]. Additional scalar or pseudo-scalar states could also provide a portal to dark matter, acting as a mediator between the SM and dark sector [16–18].

In this paper, a search for massive scalar (H) and pseudo-scalar (A) states decaying into a top–antitop quark pair ($t\bar{t}$) is presented. The significant interference between the signal process and the dominant, irreducible background from SM $t\bar{t}$ production is taken into account. The search is conducted on a sample of proton–proton (pp) collision data with an integrated luminosity of 140 fb^{-1} at a center-of-mass energy $\sqrt{s}$ = 13 TeV, collected with the ATLAS detector [19]. The results of this search are interpreted in a range of representative benchmark models, in particular 2HDM-based models.

A 2HDM model is assumed with a CP-conserving potential with a softly broken Z_2 symmetry [20]. After electroweak symmetry breaking the model contains five Higgs bosons: a lighter CP-even boson, h, a heavier CP-even boson, H, a CP-odd boson, A, and two charged bosons, $H^{\pm}$. It is assumed that the 125 GeV Higgs boson discovered by ATLAS and CMS [21, 22] corresponds to the lighter CP-even state, h. The 2HDM coupling structure is chosen to be of type-II [20]. In this case, the parameter governing the fermionic decays of the neutral scalar and pseudo-scalar states is the ratio of the two vacuum expectation values of the two Higgs doublets, $\tan\beta$. According to precision analyses of the Higgs-like particle by ATLAS and CMS, the parameters of the 2HDM must be close to the alignment limit ($\cos(\beta-\alpha)=0$), where α denotes the mixing angle between the two doublets. A prominent realisation of such a 2HDM is the Minimal Supersymmetric SM (MSSM). Type-II 2HDMs are also predicted in models of dark matter (DM), such as the 2HDM+a [17], in which an additional pseudo-scalar a, mediating the interaction between the SM- and the DM-sector, is introduced, in addition to the five Higgs bosons of the 2HDM Higgs sector. For large $\tan\beta$ stringent limits on the A/H masses exist, the most stringent using decays $A/H \to \tau\tau$ [23, 24]. For small values of $\tan\beta$ the decay $A/H \to t\bar{t}$ dominates. The present limits are much weaker than for large $\tan\beta$ and the most stringent exclusion are derived from searches targeting heavy scalars decaying to a pair of Higgs bosons, $H \to hh$ [25, 26], from searches for decays of charged Higgs bosons to a top and a bottom quark, $H^{\pm} \to tb$ [27, 28], and from searches targeting the $t\bar{t}$ associated production of A/H with subsequent decay to $t\bar{t}$, leading to a 4-top ($t\bar{t}t\bar{t}$) final state [29]. The dominant production mode for a heavy H or A is, like for the SM Higgs, via gluon-gluon fusion, which proceeds dominantly via a top-quark loop. Due to the top quarks in the loop, the matrix element acquires a complex phase which leads to the fact that the matrix element at the mass pole is no longer purely imaginary. This, in turn, leads to a complicated interference pattern with the irreducible background $gg \to t\bar{t}$. Because of this interference pattern the peak at the A, H mass can disappear completely and is replaced by a broad peak at lower masses and a dip around the resonance mass. In Figure 1, the leading-order Feynman graph for the signal process and for the interfering part of the background are shown.

A search for $A, H \to t\bar{t}$ taking the interference into account was first presented by ATLAS on 20.3 fb^{-1}of 8 TeV data [30] and later by CMS on 36 fb^{-1}of 13 TeV data [31]. Neither of these two searches observed a significant deviation from the SM expectation. In this paper, a search for $gg \to A/H \to t\bar{t}$ on the full dataset collected with the ATLAS detector during LHC Run 2, amounting to 140 fb^{-1}of pp collision data



(a) $gg \to A/H \to t\bar{t}$ (b) SM $t\bar{t}$

Figure 1: Diagram for (a) the resonant production of a pseudo-scalar or scalar from gluon–gluon (gg) initial states via a fermion loop with subsequent decay to $t\bar{t}$ at leading order and (b) the interfering diagram for gg induced $t\bar{t}$ production via the strong force at tree level.

at $\sqrt{s} = 13$ TeV, is presented. The search targets events in which one top-quark decays leptonically and the other hadronically (1-lepton channel), as well as events in which both top quarks decay leptonically (2-lepton channel). Leptonic top-quark decays are classified as all top-quark decays resulting in an electron (e) or muon (μ) in the final state, either directly from the W-boson decay or via a W boson that decays into a leptonically decaying τ-lepton. In the 1-lepton channel, separate analysis strategies targeting resolved and merged hadronic top-quark decays are used. In the latter case, the merged top-quark decay is reconstructed using large jets with variable radius parameter that are re-clustered from calibrated small-R jets. The dominant background in both channels arises from SM $t\bar{t}$ production, which is estimated using simulated events that are corrected to high-precision predictions calculated at next-to-next-to-leading order (NNLO) in QCD and NLO in the electroweak (EW) interaction [32].

This paper is structured as follows: In Section 2, the benchmark models used for the interpretation of the results of this search are introduced. In Sections 3 and 4, the ATLAS detector as well as the data and simulated event samples are described. The event selection, categorisation, reconstruction of observables, the estimation of background processes and systematic uncertainties are described in Sections 6 to 9. The statistical model for the interference analysis is discussed in Section 10. Finally, the results are presented and discussed in Section 11.

2 Theoretical framework

In this paper, a CP-conserving type-II 2HDM with a softly broken Z_2 symmetry is considered as a benchmark model. The alignment and decoupling limit are assumed, i.e. $m_h = 125$ GeV, $v = 246$ GeV, and $\cos(\beta - \alpha) = 0$, where v denotes the electroweak vacuum expectation value. The parameter m_{12} of the Z_2 breaking term of the potential is taken to be $m_{12}^2 = m_A^2 \tan\beta/(1 + \tan\beta^2)$. In this model, the production cross-sections and widths of A and H, as well as the signal shape, are uniquely determined by $\tan\beta$ and the masses m_A and m_H. In the alignment limit, the scalar h behaves SM-like and the scalar H does not couple to gauge bosons. The coupling relative to the SM Higgs coupling for the pseudo-scalar A is multiplied by $\tan\beta$ for down type quarks and by $1/\tan\beta$ for up-type quarks. For the scalar H, the coupling to up-type quarks receives in addition a minus sign [33]. In this paper, the scalar and pseudo-scalar states are assumed to be mass degenerate, i.e. $m_H = m_A$.

As a second benchmark model, the Higgs sector of the hMSSM [34] is considered, which constitutes a special instance of a type-II 2HDM. In this specific Minimal Supersymmetric extension of the Standard

Model (MSSM) [35–39], the lighter scalar h is identified with the Higgs boson ($m_h = 125$ GeV) discovered in 2012. This choice fixes the dominant radiative corrections that enter the MSSM [40]. As a result, the hMSSM can be fully described by only two free parameters, m_A and $\tan\beta$. In particular, in this case, the mass of the heavier scalar H depends on m_A. For $m_A \approx 2m_t$, the H is around 40 GeV heavier than the A, while the mass differences reduces to $\Delta m \sim 10$ GeV at $m_A \approx 1$ TeV. The couplings of the A and the H are to a good approximation the same as in the alignment limit.

A 2HDM-based model of dark matter, referred to as 2HDM+a, is also considered in this publication. In addition to the extra Higgs bosons introduced by the 2HDM, the 2HDM+a includes a fermionic DM particle χ and a pseudo-scalar (CP-odd) mediator a with Yukawa-like couplings to both the SM fermions and the Dirac DM particle χ. The mediator mixes with the pseudo-scalar A of the 2HDM sector with mixing angle θ. The 2HDM+a is promoted by the LHC Dark Matter Working Group as a simple, ultra-violet-complete (UV-complete), gauge-invariant, and renormalisable benchmark model of DM [41]. A range of different benchmark scenarios are recommended to and explored by the LHC experiments [41]. In all cases, the scalar and pseudo-scalar states are assumed to be mass-degenerate: $m_H = m_A$.

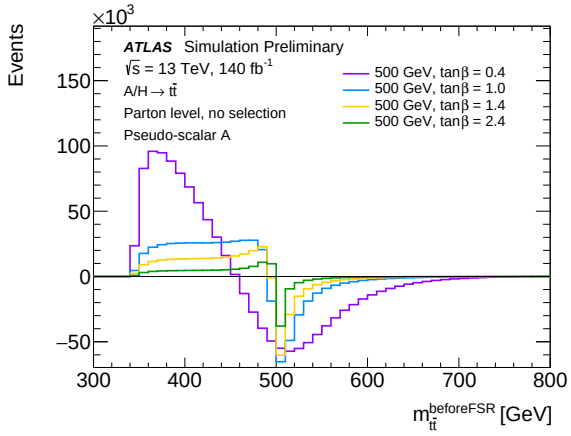
Finally, a simplified, generic scenario is considered in which only the interference pattern of either a scalar or a pseudo-scalar are assumed to appear in the spectra of sensitive variables. The full spectrum in the presence of signal-background interference can be described in the terms of the contributions from the pure (resonant) signal S, the interference term I, and the total background B, which is dominated by SM $t\bar{t}$ production, $B_{t\bar{t}}$. The interference pattern $S + I$, which describes the deviation from the background-only hypothesis in the presence of a signal process with signal-background interference, is obtained by subtracting the SM $t\bar{t}$ background from the inclusive process $S + I + B_{t\bar{t}}$, see Section 4 for details. In the simplified, generic scenario, the coupling to $t\bar{t}$, predicted for a (pseudo-)scalar of a given mass and width, is scaled by the coupling modifier $g_{A/Ht\bar{t}}$, which is a free parameter in this scenario, along with the mass $m_{A/H}$ and the width $\Gamma_{A/H}$. This benchmark scenario differs from the other, 2HDM-based, benchmark models considered in this publication, in which $\Gamma_{A/H}$ is a function of $m_{A/H}$ and $\tan\beta$. In a type-II 2HDM, the coupling modifier is inversely proportional to $\tan\beta$: $g_{A/Ht\bar{t}} = 1/\tan\beta$. The interference pattern in this generic scenario is derived from the pure-signal ($S \equiv S(m_{A/H}, \Gamma_{A/H})$) and the signal-plus-interference ($S + I \equiv (S + I)(m_{A/H}, \Gamma_{A/H})$) templates, obtained for a given choice of $m_{A/H}$ and $\Gamma_{A/H}$, as follows:

$$\begin{aligned}(S + I)(g^2_{A/Ht\bar{t}}) &= g^4_{A/Ht\bar{t}} \cdot S + g^2_{A/Ht\bar{t}} \cdot I \\ &= (g^4_{A/Ht\bar{t}} - g^2_{A/Ht\bar{t}}) \cdot S + g^2_{A/Ht\bar{t}} \cdot (S + I). \end{aligned} \quad (1)$$

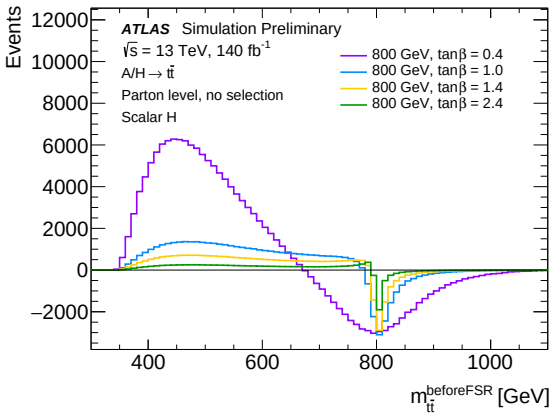
This scenario allows for the results of the search to be re-interpreted in the context of models beyond the 2HDM-based ones discussed in this paper by providing constraints on a wider set of interference patterns obtained from the well-defined S and $S + I$ templates for a given mass and width via variations of the coupling modifier according to Equation 1. This includes models which, unlike 2HDM-based models, predict the existence of a single new pseudo-scalar or single scalar, such as models predicting a single heavy axion-like particles coupling to the top quark [42].

In Figure 2, two sets of interference patterns for the production and decay to $t\bar{t}$ of a single pseudo-scalar with mass $m_A = 500$ GeV and a single scalar with mass $m_H = 800$ GeV for different values of $\tan\beta$ in a type-II 2HDM in the alignment limit are shown. The interference patterns are shown as a function of the $t\bar{t}$ invariant mass, $m_{t\bar{t}}$, calculated at parton level before the emission of final-state radiation (FSR). The SM $t\bar{t}$ background contribution has been subtracted in both cases to yield the signal-plus-interference ($S + I$) distribution only. For all scenarios, a deficit of events compared to the SM expectation is predicted in the $m_{t\bar{t}}$ region around the signal mass, while an excess of events is predicted for lower values of $m_{t\bar{t}}$.

Higher values of $\tan\beta$ correspond to smaller relative widths $\Gamma_{A/H}/m_{A/H}$ and hence a narrower peak-dip structure. In addition, the interference patterns obtained for a single pseudo-scalar and scalar, respectively, with width $\Gamma_{A/H}/m_{A/H} = 10\%$ for the generic scenario are shown in Figure 3 for various values of the coupling modifier $g_{A/Ht\bar{t}}$. For small values of the coupling modifier, the interference pattern exhibits a peak-dip structure. For coupling values $g_{A/Ht\bar{t}} > 1$, the resonant contribution S, which scales like $g^4_{A/Ht\bar{t}}$, dominates over the interference contribution I, which scales like $g^2_{A/Ht\bar{t}}$. Hence for large values of the coupling modifier, the interference pattern can exhibit a peak-peak structure.

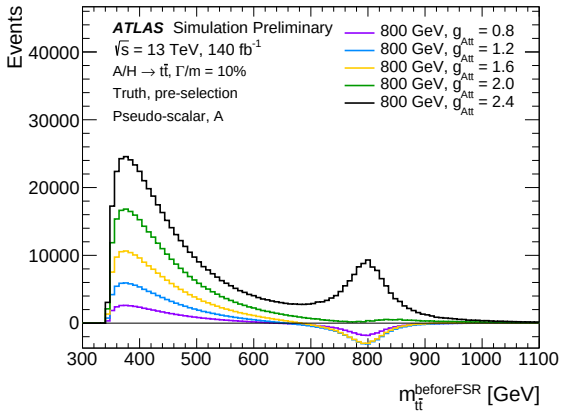


(a) $m_A = 500$ GeV

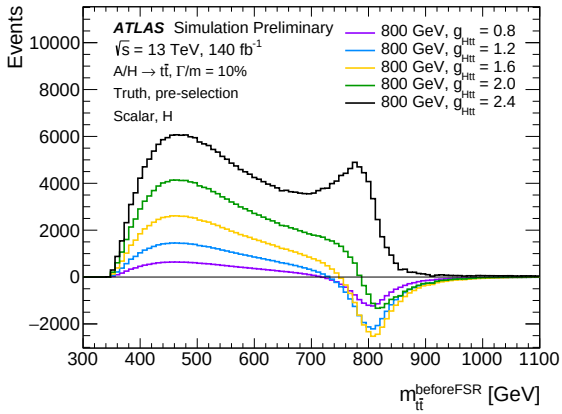


(b) $m_H = 800$ GeV

Figure 2: Signal-plus-interference distributions in $m_{t\bar{t}}$ at parton level before FSR for (a) a single pseudo-scalar A with mass $m_A = 500$ GeV and (b) a single scalar H with mass $m_H = 800$ GeV for various values of $\tan\beta$ in a type-II 2HDM in the alignment limit.



(a) $m_A = 800$ GeV, $\Gamma_A/m_A = 10\%$



(b) $m_H = 800$ GeV, $\Gamma_H/m_H = 10\%$

Figure 3: Signal-plus-interference distributions in $m_{t\bar{t}}$ at parton level before FSR for (a) a single pseudo-scalar A with mass $m_A = 800$ GeV and for (b) a single scalar H with mass $m_H = 800$ GeV, both with relative width $\Gamma_{A/H}/m_{A/H} = 10\%$. The distributions are shown for different values of the coupling modifier $g_{A/Ht\bar{t}}$ in the generic benchmark scenario (see text).

3 ATLAS detector

The ATLAS detector [19] is a multipurpose particle detector with a forward–backward symmetric cylindrical geometry and a near 4π coverage in solid angle.[1] It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadron calorimeters, and a muon spectrometer. The inner tracking detector (ID) consists of silicon pixel and silicon microstrip detectors covering the pseudorapidity range $|\eta| < 2.5$, and a surrounding transition radiation tracking detector that enhances electron identification in the range $|\eta| < 2.0$. A new inner pixel layer, the insertable B-layer [43, 44], was added at a mean radius of 3.3 cm during the period between Run 1 and Run 2 of the LHC. Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity in the region $|\eta| < 3.2$. A steel/scintillator-tile hadron calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The endcap and forward regions ($1.5 < |\eta| < 4.9$) of the hadron calorimeter are made of LAr active layers with either copper or tungsten as the absorber material. The muon spectrometer (MS) surrounds the calorimeters and is based on three large superconducting air-core toroidal magnets with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 Tm across most of the detector. Three layers of high-precision tracking chambers provide coverage in the range $|\eta| < 2.7$, while dedicated fast chambers allow triggering in the region $|\eta| < 2.4$. The ATLAS trigger system consists of a hardware-based level-1 trigger followed by a software-based high-level trigger [45]. The level-1 trigger uses a subset of the detector information to accept events at a rate below 100 kHz, while the software-based trigger reduces the accepted event rate to 1 kHz on average depending on the data-taking conditions. An extensive software suite [46] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

4 Data and simulated event samples

In this search, data from pp collisions at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 140 fb^{-1}, collected in the years 2015 to 2018 with the ATLAS detector, are analysed. The uncertainty in the combined 2015–2018 integrated luminosity is 0.83% [47], obtained using the LUCID-2 detector [48] for the primary luminosity measurements. All detector subsystems were required to be operational during data taking. The average number of interactions per bunch crossing (pile-up) was $\langle\mu\rangle = 33.7$ with some variation over the years.

Candidate events are selected using single-muon and single-electron triggers. These triggers require a muon (electron) with transverse momentum p_{T} (transverse energy E_{T}) above a certain threshold and passing certain data quality and, for some triggers, lepton isolation requirements. The triggers with the lowest p_{T} (E_{T}) threshold and without prescaling require $p_{\mathrm{T}} > 26$ GeV for muons (electrons) in 2016–2018 and include a lepton isolation requirement that is not applied for triggers with higher thresholds. The trigger efficiency is mostly constant in the transverse momentum for leptons with $p_{\mathrm{T}} > 28$ GeV.

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the centre of the detector. The positive x-axis is defined by the direction from the interaction point to the centre of the LHC ring, with the positive y-axis pointing upwards, while the beam direction defines the z-axis. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity η is defined in terms of the polar angle θ by $\eta = -\ln\tan(\theta/2)$, while the rapidity y is defined as $y = 0.5\ln[(E + p_z)/(E - p_z)]$, where E denotes the energy and p_z the component of the momentum along the beam direction. The angular distance ΔR is defined as $\sqrt{(\Delta y)^2 + (\Delta\phi)^2}$.

Monte Carlo (MC) simulated event samples are used to model the signal and background processes, including the interference effects between the signal and SM $t\bar{t}$ production. The ATLAS simulation infrastructure [49] was used to simulate the detector and its response. Signal and nominal background samples as well as several alternative background samples used to assess systematic uncertainties were produced with a detailed GEANT4 [50] detector simulation. A faster simulation based on a parameterisation of the calorimeter response and GEANT4 for the other detector systems [49] was used for the remaining alternative background samples. Pile-up effects were modeled by overlaying minimum-bias events simulated using the soft QCD processes of PYTHIA8.186 [51] with the NNPDF2.3LO set of parton distribution functions (PDFs) [52] and the A3 [53] set of tuned parameters (tune). The pile-up profiles match the ones of each dataset between 2015 and 2018. The same offline reconstruction methods used for data were applied to the simulated samples. Corrections were applied to the simulated events in order to match the selection efficiencies, energy and mass scales and resolutions of reconstructed simulated particles to those measured in data control samples. An overview of the generator choices for the signal and background processes can be found in Table 1, with further details given in the following.

The dominant and irreducible background process is SM $t\bar{t}$ production. Smaller background components arise from single vector-boson (W,Z) production in association with hadronic jets, referred to as W/Z+jets, single-top-quark, and multijets production. In the 2-lepton channel, another small background component arises from processes with at least one fake or non-prompt lepton that satisfies the lepton identification and isolation criteria applied in this search. These are mostly $t\bar{t}$ production with one top quark decaying hadronically and the other leptonically, as well as single-top, and W+jets production. This background component is referred to as "Fakes" in the following. Minor backgrounds from $t\bar{t}$ +V and diboson (WW, ZZ, WZ) production are also considered. All background components, with the exception of multijets production in the 1-lepton channel (Section 8.2), are modelled using MC simulation. Data-driven corrections are applied to the MC-based estimates of the W+jets background in the 1-lepton channel (Section 8.1) and the Z+jets background in the 2-lepton channel (Section 8.3). The modelling of the Fakes background is also validated in data (Section 8.4).

SM $t\bar{t}$ production was generated at next-to-leading order (NLO) accuracy in QCD using POWHEG BOX v2 [54–58] with the NNPDF3.0NLO [59] PDF set and the h_{damp} parameter set to 1.5 m_{top} [60].[2] The functional form of the renormalisation and factorisation scales was set to the default scale $\sqrt{m_{\text{top}}^2 + p_{\text{T}}^2}$. The top-quark mass was set to m_{top} =172.5 GeV. The top quarks were decayed in POWHEG BOX, thus preserving their spin correlations. The matrix-element (ME) generator is interfaced with PYTHIA 8.230 to model the parton shower (PS), hadronisation, and the underlying event with parameters set according to the A14 tune [61] and using the NNPDF2.3LO PDF set. The decays of bottom and charm hadrons were modelled with EVTGEN 1.6.0 [62].

Alternative SM $t\bar{t}$ samples obtained with different generator choices and settings are used to estimate systematic uncertainties related to the modeling of this main background component. The details can be found in Section 9.1.

The top-quark kinematics in the nominal and all alternative SM $t\bar{t}$ samples were corrected to more accurate differential predictions calculated at NNLO-QCD+NLO-EW accuracy for a top-quark mass value m_{top} =173.3 GeV [32]. The corrections were applied as event-by-event weights. The weights were obtained via an iterative recursive reweighting procedure, referred to as *NNLO reweighting* in the following, in which the calculated and generated binned differential distributions of the top and anti-top quark p_{T} as well as that of

[2] The h_{damp} parameter controls the transverse momentum, p_{T}, of the first additional emission beyond the leading-order Feynman diagram in the parton shower and therefore regulates the high-p_{T} emission against which the $t\bar{t}$ system recoils.

the $t\bar{t}$ invariant mass, $m_{t\bar{t}}$, were compared. All variables were obtained at the parton level before final-state radiation. In each step, weights were obtained as the ratio of the bin contents of the calculated parton-level distribution and the corresponding distribution obtained from the given MC sample, already corrected by the previous steps of the reweighting procedure. The weights were then applied to the generated sample. In the case of the top and anti-top quark p_{T} distributions, in order to avoid introducing an artificial asymmetry between the top and anti-top kinematic distributions, the weight applied to the sample was the geometric average of the two weights derived from comparing separately the top and anti-top quark p_{T} distributions. Corrections were first derived based on $m_{t\bar{t}}$, then on the $p_{\mathrm{T}}(t)$ and $p_{\mathrm{T}}(\bar{t})$ distributions, and then again on the $p_{\mathrm{T}}(t)$ and $p_{\mathrm{T}}(\bar{t})$ distributions. The procedure was iterated three times to achieve a good agreement between the reweighted and calculated target distributions in all three variables.

All reweighted $t\bar{t}$ samples were additionally normalised to the cross-section prediction at NNLO in QCD including the resummation of next-to-next-to-leading logarithmic (NNLL) soft-gluon terms calculated using Top++ 2.0 [63–69]. For pp collisions at a centre-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$, this cross-section corresponds to $\sigma(t\bar{t})_{\mathrm{NNLO+NNLL}} = 814 \pm 76\,\mathrm{fb}$ using a top-quark mass of $m_{\mathrm{top}} = 173.3\,\mathrm{GeV}$. The uncertainties in the cross-section due to the PDF and α_{s} were calculated using the PDF4LHC15 prescription [70] with the MSTW2008nnlo [71, 72], CT10nnlo [73, 74] and NNPDF2.3lo PDF sets in the five-flavour scheme, and were added in quadrature to the effect of the scale uncertainty.

Single-top production in the Wt-channel [75], which constitutes the main single-top contribution in this analysis, is generated with Powheg Box v2 and using the NNPDF3.0nlo PDF set. The overlap between $t\bar{t}$ and Wt production is treated within the diagram removal (DR) scheme [76] and the renormalisation and factorisation scales are set to $H_T/2$, where H_T denotes the scalar sum of the transverse momenta of all final-state particles in the event. The ME generator is interfaced with Pythia 8.307.1 with parameters set according to the A14 tune and using the NNPDF2.3lo PDF set. EvtGen 1.7.0 is used to decay bottom and charm hadrons. Single-top production in the s-channel [77], like SM $t\bar{t}$ production, were generated with Powheg Box v2 at NLO in QCD using the five-flavour scheme and the NNPDF3.0nlo PDF set. Single-top production in the t-channel was generated with Powheg Box v1 [78]. This generator uses the four-flavour scheme for the NLO matrix element calculations together with the four-flavour PDF set NNPDF3.04f. For this process, the top-quark decays were simulated using MadSpin [79], preserving all spin correlations. For both processes, the ME generator was interfaced with Pythia 8.230 with parameters set according to the A14 tune and using the NNPDF2.3lo PDF set. EvtGen 1.6.0 was used to decay bottom and charm hadrons. The respective samples were normalised to the theoretical cross-sections for Wt-channel [80], calculated at NNLO+NNLL accuracy, s-channel [81], and $t-$ channel[82] production, both calculated at NLO accuracy.

The background from $t\bar{t}V$ ($V = W, Z$) production was generated at NLO accuracy in QCD with MadGraph5_aMC@NLO 2.3.3 with the NNPDF3.0nlo PDF set, interfaced with Pythia 8.210 using the A14 set of tuned parameters and the NNPDF2.3lo PDF set for parton showering, hadronisation, and underlying event. EvtGen 1.2.0 was used for the decay of bottom and charm hadrons. The samples were normalised to the theoretical cross-section calculated at NLO in QCD [83]. The background from $t\bar{t}+h$ production was generated at NLO accuracy in QCD with Powheg Box v1 with the NNPDF3.0nlo PDF set, interfaced with Pythia 8.230 using the A14 set of tuned parameters and the NNPDF2.3lo PDF set for parton showering, hadronisation, and underlying event. EvtGen 1.6.0 was used to decay bottom and charm hadrons. The samples are normalised to the theoretical cross-section calculated at NLO QCD [84] and NLO EW accuracies.

Production of a single W boson in association with hadronic jets (W+jets) was simulated with Sherpa 2.2.11 [85], while Z+jets production was simulated with Sherpa 2.2.11 (Sherpa 2.2.1) in the 1-lepton (2-lepton)

Table 1: List of ME generators and the order of the strong coupling constant in the perturbative calculation, PDF sets, shower generator and tune for the different signal and background processes.

Process	ME generator	ME order	PDF set	PS and hadronisation	UE tune
Signal	MadGraph5_aMC@NLO 2.6.7	LO	NNPDF3.0nlo	Pythia [8.244]	A14
$t\bar{t}$	Powheg Box v2	NLO, rew. to NNLO + NLO EW	NNPDF3.0nlo	Pythia 8.230	A14
Single top	Powheg Box v2	NLO	NNPDF3.0nlo	Pythia 8.230/8.235	A14
Diboson	Sherpa 2.2.1/2.2.2	MEPS@NLO	NNPDF3.0nnlo	Sherpa	internal
W+jets	Sherpa 2.2.11	MEPS@NLO	NNPDF3.0nnlo	Sherpa	internal
Z+jets	Sherpa 2.2.1	MEPS@NLO	NNPDF3.0nnlo	Sherpa	internal
$t\bar{t}+V$	MadGraph5_aMC@NLO 2.3.3	NLO	NNPDF3.0nlo	Pythia 8.210	A14

channel. NLO MEs were used for up to two partons for all samples, and LO MEs for up to five (four) partons for the Sherpa 2.2.11 (Sherpa 2.2.1) samples. The MEs were calculated with the Comix [86] and OpenLoops [87–89] libraries. They were matched with the Sherpa parton shower [90] using the MEPS@NLO prescription [91–94] using the set of tuned parameters developed by the Sherpa authors. The NNPDF3.0nnlo set of PDFs was used for all V+jets samples. The W+jets and Z+jets samples were normalised to the theoretical cross-section calculated at NLO accuracy in QCD [95]. An additional data-driven normalisation correction was derived for the W+jets background in the 1-lepton channel.

Diboson events with fully leptonic and semileptonic decays were simulated using Sherpa 2.2.1 and Sherpa 2.2.2 [96], respectively. The simulation includes off-shell effects and Higgs boson contributions, where appropriate. Events were generated using matrix elements at NLO accuracy in QCD for up to one additional parton and at LO accuracy for up to three additional parton emissions. Samples for the loop-induced processes $gg \to VV$ were generated using LO-accurate matrix elements for up to one additional parton emission. The matrix element calculations were matched and merged with the Sherpa parton shower based on Catani–Seymour dipole factorisation [86, 90] using the MEPS@NLO prescription. The virtual QCD corrections were provided by the OpenLoops library. The NNPDF3.0nnlo set of PDFs was used, along with the dedicated set of tuned parton-shower parameters developed by the Sherpa authors. The cross-sections from the generator were used for sample normalisation.

The signal process $gg \to A/H \to t\bar{t}$ and its interference with the background from SM $t\bar{t}$ production were simulated with the MadGraph [83] v2.6.7 generator with the model of Ref. [97], which implements A/H production through loop-induced gluon–gluon fusion with loop contributions from top and bottom quarks at leading order (LO) in QCD. Additional, non-resonant contributions at the same order [98] or NLO contributions [99] were not considered in the simulation. The latter are partially accounted for via the multiplicative correction factors described below. The parton luminosities were modeled using the NNPDF3.0nlo PDF set. The factorisation and renormalisation scales were set to $\frac{1}{2}\sum_{i=1}^{N}\sqrt{m_i^2 + p_{T,i}^2}$ and the top-quark mass was set to 173.3 GeV to be consistent with the top-quark mass value that is used in the NNLO-QCD+NLO-EW predictions to which the generated SM $t\bar{t}$ sample is corrected. The widths of the (pseudo-)scalar states for a given signal hypothesis was calculated with 2HDMC [100] v1.8.0 and used as an input to the ME generation. The top-quark decays were simulated using MadSpin to preserve all spin correlations. Pythia 8.244 was used to model the PS, hadronisation, and the underlying event with parameters set according to the A14 tune and using the NNPDF2.3lo PDF set. The decays of bottom and charm hadrons were modelled with EvtGen 1.6.0. All signal samples are produced with the full detector simulation.

For the statistical interpretation, distributions of the $t\bar{t}$ invariant mass and related variables in data are compared to a combination of the expected distributions from all background processes B, the pure signal

process S, and the signal-plus-interference component $S+I$ for a given signal hypothesis, see Equation 6 in Section 10 for the parameterisation of the likelihood function in terms of the S and $S+I$ components. The $S+I$ contributions therefore need to be obtained from the inclusive process $S+I+B_{t\bar{t}}$ that MadGraph produces by default. This is achieved by modifying the MadGraph software to remove the pure (LO) SM $t\bar{t}$ process from the inclusive $S+I+B_{t\bar{t}}$ process at the ME level on an event-by-event basis. The modified MadGraph code was used and validated in the context of Ref. [30]. The $S+I$ events obtained with the modified software can have positive or negative weights and the overall binned distribution typically exhibits several bins with negative yields in the dip region.

Event samples for both the S and $S+I$ components for different signal hypotheses (i.e. different values of $(m_{A/H}, \tan\beta)$ or $(m_{A/H}, \Gamma_{A/H})$ in the generic model) were obtained from a set of signal samples S after the detector simulation by applying an event-by-event reweighting. This reweighting substantially reduces the computing time required to model all tested signal hypotheses. The weight is the ratio of the MadGraph matrix elements, calculated from the four-momenta of the incoming gluons and outgoing top quarks of the generated event with the new and the old values of $(m_{A/H}, \tan\beta)$, respectively.

Two sets of input samples were generated for signal processes involving scalar and pseudo-scalar states, respectively, and each was used to obtain S and $S+I$ distributions for signal hypotheses involving the same CP-state. For the input samples, $\tan\beta = 0.4$ and $m_{A/H}$ ranging from 400 GeV to 1400 GeV were chosen to obtain a good coverage of the $m_{t\bar{t}}$ spectrum and related kinematic distributions. The reweighting procedure was validated by comparing the particle-level $m_{t\bar{t}}$ distributions obtained via reweighting to the equivalent distributions generated directly with MadGraph. The reweighted and generated distributions were found to agree within statistical uncertainties. All S and $S+I$ samples were therefore obtained through reweighting from the above two sets of input samples. Only signal hypotheses with $\tan\beta \geq 0.4$ were considered to ensure the perturbativity of the top-quark Yukawa coupling [101].

Multiplicative correction factors K_S were used to correct the generated signal (S) cross-section to the value calculated at partial NNLO precision in QCD with SusHi v1.7.0 [102–107]. The values of K_S range from $3-4$ for low values of $m_{A/H}$ and $\tan\beta$ to around $1-2$ for high values of $m_{A/H}$ and $\tan\beta$. A multiplicative correction factor K_I is also applied to the interference term I. It is defined as $k_I = \sqrt{k_S \cdot k_B^{\mathrm{LO}}}$ [108], where $k_B^{\mathrm{LO}} = 2.07$ is the ratio of the $t\bar{t}$ cross-section calculated at NNLO+NNLL precision and the cross-section for LO $t\bar{t}$ production with MadGraph using settings consistent with those for S and $S+I$ production. These correction factors K_S and K_I are applied to the S and I templates, respectively, obtained via the event-by-event reweighting, where the I template is obtained by subtracting the S template from the $S+I$ template. The correction factors are applied for all benchmark models with the exception of the 2HDM+a due to the fact that no correction factors K_S are available for the process $gg \to a/A \to t\bar{t}$, which depends on the mixing of the two pseudo-scalar particles a and A.

5 Event reconstruction

Common event-quality criteria and object definitions are applied for both analysis channels, including standard data-quality requirements to select data events with the detector in good operating condition [109]. In addition, in each analysis channel, dedicated event selection criteria, which are specific to the objects and kinematics of interest in those final states, are applied as described in Section 6.

Events are required to have at least one reconstructed pp interaction vertex with a minimum of two associated particle trajectories (*tracks*) with transverse momenta $p_{\mathrm{T}} > 0.5$ GeV. The *primary vertex* is

defined as the vertex with the highest sum of squared transverse momenta of associated tracks [110]. A set of baseline quality criteria are applied to reject events with non-collision backgrounds or detector noise [111]. Two levels of object identification requirements are defined for charged leptons and jets: baseline and signal. Baseline leptons and jets are selected with looser identification criteria, and are used in computing the missing transverse momentum as well as in resolving possible reconstruction ambiguities. Signal leptons and jets are a subset of the baseline objects, with tighter quality requirements which are used to define the search regions. Isolation criteria are used to discriminate between signal leptons and leptons arising from semileptonic heavy-flavour decays or jets misidentified as leptons.

Jets are reconstructed from particle-flow objects [112, 113] using the anti-k_t algorithm [114, 115] with a radius parameter $R = 0.4$. The particle-flow algorithm combines information about ID tracks and energy deposits in the calorimeters to form the input for jet reconstruction. An energy calibration is applied to both the input calorimeter clusters [116] and the final reconstructed jets [113]. Additionally, a pile-up subtraction procedure [117] is applied along with a global sequential calibration to account for flavour dependencies. To suppress jets arising from pile-up, a jet-vertex-tagging technique using a multivariate likelihood [118] is applied to jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$, ensuring that selected jets are matched to the primary vertex. Baseline jets are selected by requiring $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 4.5$. Signal jets are selected via the tighter requirements $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.5$.

The selected and calibrated jets are used as inputs for jet reclustering [119] using the anti-k_t algorithm with a variable radius parameter [120, 121]. These reclustered jets are referred to as "large-VR jets", and are used as proxies of the hadronically decaying top quark in the 1-lepton channel. The effective radius of these jets is inversely proportional to the jet p_{T}, according to the relation $R_{\mathrm{eff}} \simeq \rho/p_{\mathrm{T}}^{\mathrm{jet}}$. The parameter ρ is chosen to be 600 GeV, a value found to be optimal for the reconstruction of boosted hadronically decaying top quarks [121]. The maximum and minimum radius of the large-VR jets is set to 1.5 and 0.4, respectively. The calibration corrections and uncertainties for the reclustered large-VR jets are inherited from the input jets [119]. A trimming procedure [120] is applied to reclustered large-VR jets which removes all the associated small-R jets that have a p_{T} below 5% of the p_{T} of the reclustered jet to suppress gluon radiation and mitigate pile-up effects. The large-VR jets are required to have $p_{\mathrm{T}} > 200$ GeV, $|\eta| < 2.0$, a jet mass $m > 100$ GeV, and at least two constituent jets.

Small-R jets in the range $|\eta| < 2.5$ are identified as containing a b-hadron (henceforth called b-tagged) using the 'DL1r' algorithm [122]. This algorithm is based on a multivariate classification technique with a DNN combining information from the impact parameters of tracks and topological properties of secondary and tertiary decay vertices reconstructed from the tracks associated to the jet. The b-tagged jets are selected in this analysis using a working point corresponding to an efficiency of 77% for identifying true b-jets in simulated SM $t\bar{t}$ events. This working point corresponds to a rejection factor of 6 for charm and of 134 for light-flavour 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-, c-, 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.

Muon candidates are reconstructed from matching tracks in the ID and MS, refined through a global fit which uses the hits from both subdetectors [123]. Baseline muons must have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$, and satisfy a set of medium identification criteria. Additionally, the longitudinal impact parameter is required to satisfy $|z_0 \sin\theta| < 0.5$ mm. Signal muons are required to have $p_{\mathrm{T}} > 28$ GeV and additionally satisfy the following requirement on the transverse impact parameter d_0 and its uncertainty σ_{d_0}: $|d_0/\sigma(d_0)| < 3$. Signal muons are required to be isolated [124] using the requirement that the sum of the transverse momenta of the tracks in a variable-size cone around the muon direction, excluding the

muon track be less than 6% of the transverse momentum of the muon. The track isolation cone size is given by the minimum of $R = 10\ \text{GeV}/p_T^\mu$ and $R = 0.3$, where p_T^μ is the muon p_T. Thus, the cone radius increases with decreasing p_T^μ up to a maximum of 0.3.

Electron candidates are reconstructed from energy deposits in the electromagnetic calorimeter matched to a charged-particle track in the ID [124]. The track is required to be matched to the primary vertex by imposing the requirement $|\Delta z_0 \sin\theta| < 0.5$ mm. Electron candidates are required to be within $|\eta| < 2.47$, excluding the transition region between the barrel and endcap calorimeters ($1.37 < |\eta| < 1.52$). Baseline electrons are required to satisfy $p_\text{T} > 10$ GeV and fulfil loose identification criteria, using a likelihood-based discriminant that combines information about tracks in the ID and energy deposits in the calorimeter system [124]. The number of hits in the innermost pixel layer is used to discriminate between electrons and converted photons. Signal electrons are required to also satisfy $p_\text{T} > 28$ GeV and the tight likelihood identification criteria [124]. Additionally, like signal muons, signal electrons must fulfil $|d_0/\sigma(d_0)| < 5$. The same variable-cone isolation requirement as for muons is imposed on signal electrons, with the exception that the maximum cone radius is set to 0.2.

An overlap removal procedure is applied to resolve the reconstruction ambiguities between electrons, muons, and jets. First, if an electron shares the same ID track with another electron, the electron with the lower transverse momentum is discarded. Electrons sharing the same track with a muon candidate are rejected as they are assumed to be a falsely reconstructed photon from bremsstrahlung. Next, jets are rejected if they lie within $\Delta R = 0.2$ of an electron. Similarly, jets are rejected if they are within $\Delta R = 0.2$ of a muon if the jet has fewer than three associated tracks or the muon is matched to the jet through ghost association [125]. Next, to reduce the background contributions due to muons from heavy-flavour decays inside jets, muons are removed if they are separated from the nearest jet by $\Delta R < 0.04 + 10\ \text{GeV}/p_T^\mu$. For electrons, an *electron-in-jet* subtraction method is used to suppress backgrounds from leptonic heavy-flavour decays inside jets while at the same time maintaining a high reconstruction efficiency for electrons from highly collimated leptonic decays of boosted top quarks, which typically result in an electron close to a b-jet. If an electron is within $\Delta R < 0.4$ of a jet, the electron p_T is subtracted from the jet p_T. If the transverse momentum of the resulting modified jet, p_T (electron-subtracted-jet), fails the requirement $p_\text{T}(\text{jet}) > 25$ GeV, the jet is removed as it can be assumed that the jet was actually caused by the electron. The selected electron is kept in this case. If the modified jet passes the jet p_T cut, the ΔR between the electron and the modified jet is recalculated. If $\Delta R(e, \text{mod. jet}) > 0.2$, then both the electron and jet are kept. Otherwise, it is assumed that the electron resulted from a heavy-flavour decay. In this case, the electron is removed and the original jet is kept.

The missing transverse momentum $\vec{p}_\text{T}^{\,\text{miss}}$, with magnitude E_T^miss, is calculated as the negative vectorial sum of the transverse momenta of all baseline reconstructed objects (electrons, muons, jets and photons [126]) and a soft term. The soft term includes all tracks associated with the primary vertex but not matched to any reconstructed physics object. Tracks not associated with the primary vertex are not considered in the E_T^miss calculation, improving the E_T^miss resolution by suppressing the effect of pile-up [127, 128].

6 Event selection and categorisation

Events are required to have fired one of the single-electron or single-muon triggers. For all events the pre-selection detailed in Section 5 is applied. The further selection depends on the number of leptons in the event.

6.1 Event selection and categorisation in the 1-lepton channel

Events are required to contain exactly one selected electron or muon with a minimum transerse momentum of 28 GeV. This threshold is chosen to ensure that events are selected from the trigger efficiency plateau, avoiding the turn-on region. The electron or muon is also required to be matched, within $\Delta R < 0.15$, to the respective trigger object. Events with a second selected electron or muon with transverse momentum larger than 25 GeV are vetoed to ensure orthogonality of the 1-lepton and 2-lepton channels.

The detector signature of a $t\bar{t}$ decay in the 1-lepton channel involves the presence of a substantial amount of missing transverse momentum from the leptonically decaying W boson. In order to suppress background from strong multijet production that can give rise to (typically smaller) amounts of $E_{\mathrm{T}}^{\mathrm{miss}}$ via leptonic decays in heavy-flavour jets, mis-measured jet energies, etc., the requirement $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV is imposed. Additionally, events are required to fulfil $E_{\mathrm{T}}^{\mathrm{miss}} + m_T^W > 60$ GeV where the transverse mass of the selected lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$, referred to as the W transverse mass, or m_T^W, is defined as:

$$m_T^W = \sqrt{2p_{\mathrm{T}}^{\ell} E_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos\Delta\phi(p_{\mathrm{T}}^{\ell}, E_{\mathrm{T}}^{\mathrm{miss}}))}$$

All events must contain at least one b-tagged small-R jet.

Events are split into two categories, a merged topology where the hadronic top-quark decay is reconstructed as single large-VR jet, and a resolved topology where the decay products of the hadronically decaying top quark are reconstructed as three small-R jets. In order to keep the signal regions orthogonal, the selection criteria for the merged topology are applied first. The criteria for the resolved topology are applied only on the events that fail the merged selection. This preference of the merged over the resolved category for events that would pass the selection requirements for both is based on the superior $m_{t\bar{t}}$ resolution obtained with the merged-category reconstruction compared to that for events in the resolved category (Section 7.1).

Events in the merged category must contain at least one small-R jet with a distance $\Delta R < 2$ to the selected lepton, which is considered as the b-jet candidate from the leptonic top-quark decay. If more than one jet is found within $\Delta R < 2$ from the selected lepton, the b-jet candidate is chosen as the jet with the highest transverse momentum among all b-tagged candidate jets. If none of the candidate jets is b-tagged, the jet with the highest transverse momentum is chosen instead.

In addition, at least one selected large-VR jet must be present in the event with $m > 100$ GeV and a distance $\Delta R > 1.5$ to the lepton and to the selected b-jet candidate from the leptonic top-quark decay.

Only events failing the merged selection are considered for the resolved-topology selection. A fully resolved hadronic top-quark decay is expected to result in three small-R jets, one of which is a b-jet, in addition to a b-jet from the leptonic top-quark decay. Hence, events with a resolved decay topology are required to contain at least four selected small-R jets. The $t\bar{t}$ system is then reconstructed using a χ^2 algorithm, defined in Section 7.1.1, and only well-reconstructed events, defined via the requirement $\log_{10}(\chi^2) < 0.9$, are kept for further analysis. This requirement has a 60% efficiency for SM $t\bar{t}$ events as well as pure-signal events with a signal mass of 500 GeV. All selection criteria are summarised in Table 2.

In Figure 4, the selection efficiency times acceptance, including the branching ratio for the ℓ+jets final state, is shown for the resolved and merged event categories, separately for e+jets and μ+jets events. For reference, the branching ratio for $t\bar{t}$ to e+jets or μ+jets final states is about 17% for each lepton flavour, taking into account leptonic τ-lepton decays [129]. The efficiency times acceptance is dominated by the merged selection for $t\bar{t}$ invariant mass values $m_{t\bar{t}} > 600$ GeV, while the resolved selection covers the region

Table 2: Summary of the event selection criteria for the resolved and merged signal regions in the 1-lepton channel.

Selection	**Criteria**
	Common Selection
Run and event cleaning	All detector components with acceptable conditions
Single lepton trigger	Separate single-electron or single-muon triggers
Exactly one lepton	== 1 e or μ with $p_{\mathrm{T}} > 28$ GeV.
$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$ + W transverse mass	$E_{\mathrm{T}}^{\mathrm{miss}} + m_T^W > 60$ GeV
b-tagging	≥ 1 b-tagged jet
	Merged Selection
Large-VR jet	≥ 1 large-VR jet, $p_{\mathrm{T}} > 200$ GeV
Top tagging (hadronic decay)	Large-VR jet mass consistent with top quark mass: $m > 100$ GeV
Candidate b-jet (leptonic decay)	≥ 1 jet with ΔR (ℓ, R=0.4 jet) < 2.0
	ΔR (lep-b-cand-jet, ℓ) < 2.0
Back-to-back $t\bar{t}$ topology	ΔR (large-VR jet, lep-b-cand-jet) > 1.5
	ΔR (large-VR jet, ℓ) > 1.5
Matching of b-jets and top candidates	≥ 1 top candidate reconstructed using == 1b-tagged jet
	Resolved Selection
At least four jets	≥ 4 jets, $p_{\mathrm{T}} > 25$ GeV
Well-reconstructed $t\bar{t}$ system	$\log_{10}(\chi^2) < 0.9$
Matching of b-jets and top candidates	≥ 1 top candidate reconstructed using == 1b-tagged jet
Veto events passing the merged selection	

down to the kinematic threshold. The distributions correspond to the case of a single pseudo-scalar A. The corresponding results for a single scalar H are very similar.

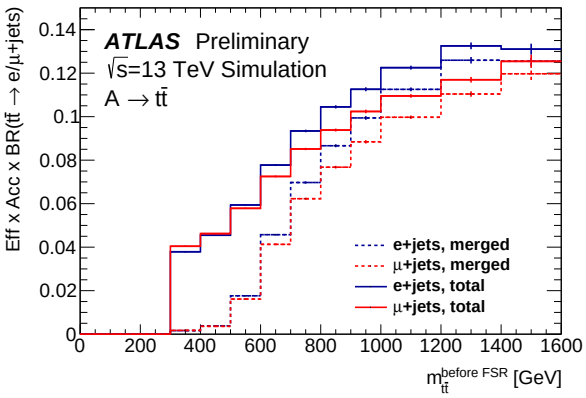


Figure 4: Selection efficiency times acceptance times $t\bar{t}$ branching ratio for the ℓ+jets ($\ell \in e, \mu$) final state (Eff x Acc x BR) as a function of the $t\bar{t}$ invariant mass at the parton level before the emission of FSR. The branching ratio for $t\bar{t}$ to e+jets or μ+jets final states is about 17% for each lepton flavour, taking into account leptonic τ-lepton decays. The distributions are obtained from all generated pure-A samples in the mass range 400 – 1400 GeV. The error bars correspond to the statistical uncertainty on the distributions.

To increase the statistical significance of the analysis, selected events in the resolved region are classified further into separate categories based on whether or not a b-tagged jet can be associated with one or both of the reconstructed top quarks. The matching of b-tagged jets to the leptonic or hadronic top quark is performed by checking whether the small-R jet assigned as b-candidate jet to the leptonic side or one of the small-R jets used for the reconstruction of the hadronic decay (based on the result of the χ^2 minimisation) are b-tagged. The "Resolved $2b$" category contains events in which both top-quark candidates have associated b-tagged jets, while the "Resolved $1b$" category contains events in which only the leptonically or hadronically decaying top-quark candidate has an associated b-tagged jet. The background composition in the resulting four event categories is shown in Figure 5. The highest $t\bar{t}$ purity (94.2%) is found in the "Resolved $2b$" region due to the requirement of two b-tagged jets. In the "Resolved $1b$" and "Merged" signal regions, the relative contribution of SM $t\bar{t}$ production to the total SM background amounts to 74.4% and 84.7%, respectively. Signal regions are also not split according to the flavour of the selected lepton for simplicity and because no increase in sensitivity was found for setups in which e+jets and μ+jets events are sorted into separate categories compared to the setup in which the signal regions are not split according to lepton flavour. The "Resolved $2b$" and "Resolved $1b$" regions are each further split into five equidistant bins of the angular variable $|\cos\theta^*|$, defined in Section 7.1, which provides additional discrimination between the signal process and SM backgrounds in the 1-lepton channel. This yields a total of eleven orthogonal signal regions for the 1-lepton channel that are used in the statistical analysis of the results. Here, θ^* denotes the angle between the momentum of the leptonically decaying top quark in the $t\bar{t}$ centre-of-mass frame and the momentum of the reconstructed $t\bar{t}$ system in the laboratory frame. A flat $\cos\theta^*$ distribution is expected for signal events as the decays of a heavy spin-zero state would result in an isotropic distribution of the resulting top quarks. The main background from SM $t\bar{t}$ production, by contrast, is dominated by t-channel processes, and thus the resulting $\cos\theta^*$ distribution peaks at ± 1. The distributions of $|\cos\theta^*|$ in the "Resolved $2b$" and "Resolved $1b$" regions obtained after the profile-likelihood fit to the data under the background-only hypothesis (Section 10) are shown in Figure 6.

Additional control and validation regions enriched in hadronic jets falsely identified as leptons are defined in the context of the data-driven estimate of the multijet background via the Matrix Method. These are obtained by reverting the $E_{\mathrm{T}}^{\mathrm{miss}}$ and/or m_T^W requirements of the SRs and additionally loosening the lepton ID and isolation requirements. Details are given in Section 8.2.

6.2 Event selection and categorisation in the 2-lepton channel

Candidate events for the 2-lepton channel are required to have exactly two charged leptons (electrons or muons) and at least two reconstructed jets. At least one of these leptons is required to have $p_{\mathrm{T}} > 28$ GeV and to match, within $\Delta R < 0.15$, the lepton with the same flavour reconstructed by the trigger algorithm. Additionally, at least one of the jets is required to be b-tagged. Depending on the flavour of the charged leptons, three channels are defined: ee, $\mu\mu$ and $e\mu$. These *pre-selection* requirements ensure orthogonality of the 1- and 2-lepton channels in this analysis. Events passing the pre-selection requirements are then further classified into a number of signal and control regions.

Events in the signal regions are required to contain two leptons with opposite-sign electric charge (OS). Additionally, in the ee and $\mu\mu$ channels, the dilepton invariant mass, m_{ll}, is required to be greater than 15 GeV and not within the range 81-101 GeV around the Z-boson mass. To further suppress the background from Z+jets production, events in the ee and $\mu\mu$ channels are required to have $E_{\mathrm{T}}^{\mathrm{miss}} > 45$ GeV. Signal-candidate events must also meet the requirement that the invariant masses of the lepton-b-jet pairs, $m_{\ell b}$, are smaller than 150 GeV for at least one of the two possible b-jet to lepton assignments. In this

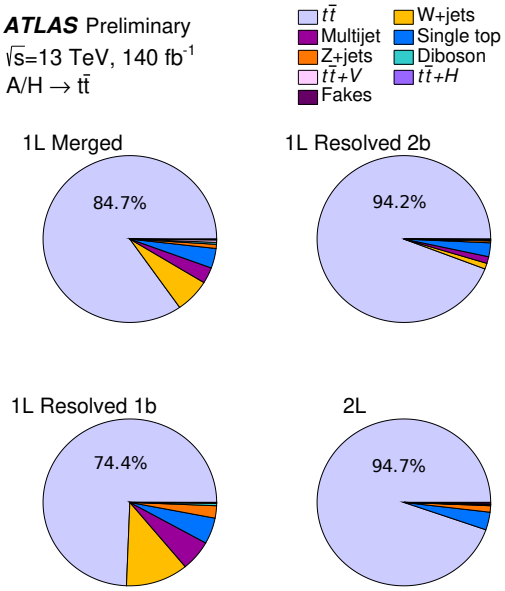
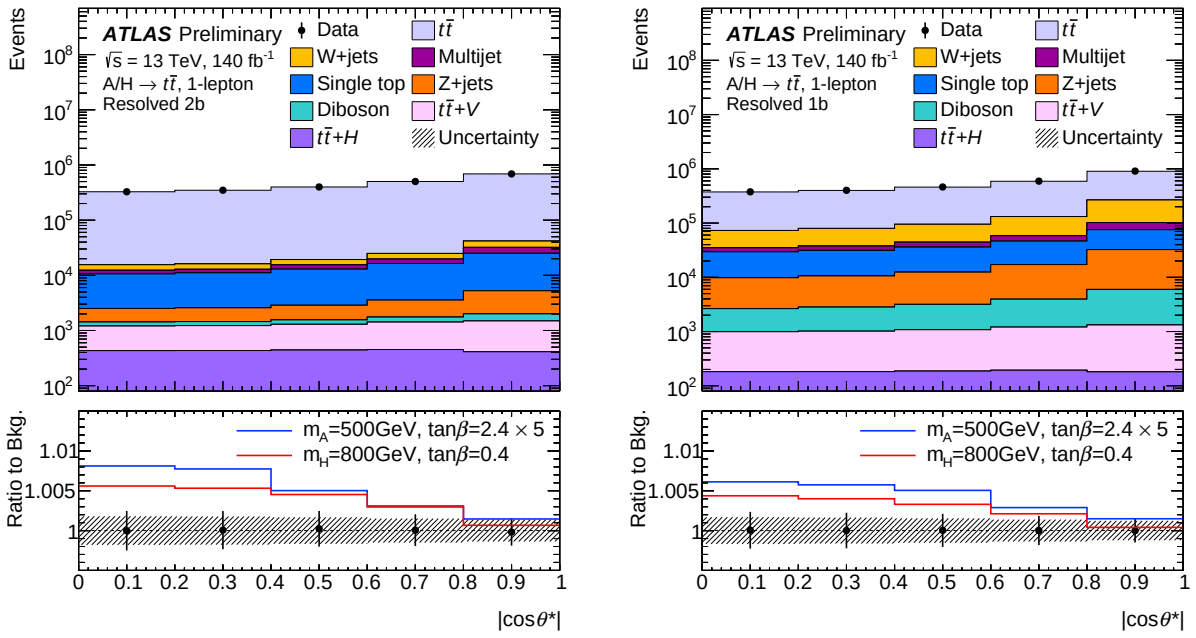


Figure 5: Background composition in the resolved signal regions with two and one b-tag, respectively, and the merged signal region of the 1-lepton channel, as well as the signal region of the 2-lepton channel, all evaluated before the profile likelihood fit to data. The resolved and 2-lepton signal regions are not broken down further into angular bins here for simplicity. The background composition in the different angular bins is similar to that of the respective inclusive regions shown here. Number within each pie charts quantifies the $t\bar{t}$ purity in the respective region.

context, the two b-jets are either taken as the two b-tagged jets with the largest transverse momentum or, if only one jet in the event is b-tagged, the b-tagged jet and jet with the highest transverse momentum among the jets without a b-tag. This requirement is meant to veto events where at least one of the lepton-b-jet pairs does not originate from a top-quark decay. It effectively suppresses events from tW production as well as $t\bar{t}$ events in which at least one of selected b-jets does not originate from a top-quark decay but, for example, from initial-state radiation (ISR). It has a 64% (68%) efficiency for SM $t\bar{t}$ events with both top quarks decaying leptonically (pure-signal events with a signal mass of 1000 GeV).

In analogy to the 1-lepton channel event categories, the signal region of the 2-lepton channel is further split into five equidistant bins in the azimuthal angle between the two leptons $\Delta\phi_{\ell\ell}$ to enhance the sensitivity to spin-0 states. These five orthogonal signal regions are used in the statistical analysis of the results. The post-fit distribution of $\Delta\phi_{\ell\ell}$ in the 2-lepton channel is shown in Figure 7. Each bin corresponds to a signal region.

Two orthogonal regions are used to correct and validate the MC-based modelling of smaller background components. A control region CR_Z enriched in Z+jets events is defined by requiring events to pass the common pre-selection requirements and additionally contain two same-flavour leptons with opposite charge and with an invariant mass consistent with leptons arising from a Z-boson decay, $81 < m_{ll} < 101$ GeV. A second, orthogonal control region CR_f, enriched in events with one prompt lepton and one lepton arising from a semileptonic decays of a hadron inside a jet or a jet mis-identified as a lepton (fake lepton), is obtained by inverting the opposite-sign requirement on the electric charges of the selected leptons and instead selecting same-sign $e\mu$ or $\mu\mu$ events. Events with a same-sign electron pair ee are not used to control and validate the fake-lepton background due to the large contributions from processes with two



(a) Resolved $2b$

(b) Resolved $1b$

Figure 6: Distributions of the reconstructed $|\cos(\theta^*)|$ in the "Resolved $2b$" (left) and "Resolved $1b$" (right) signal regions for the SM backgrounds. Each bin corresponds to a signal region. The expected relative deviation from the background prediction in the presence of an interference pattern, $(S+I+B)/B$, for two representative signal hypotheses is also shown in the ratio panel. All distributions and the uncertainty bands are obtained after the profile likelihood fit to the data under the background-only hypothesis.

prompt electrons where one of the two electrons has its charge mis-identified. Instead, the modelling of fake electrons is assessed in $e\mu$ events after verifying in $\mu\mu$ events that fake muons are well modelled (Section 8.4). To further reduce the contamination from events with two prompt leptons, only events in which the transverse W-boson masses for either the leading or sub-leading lepton is smaller than 100 GeV are considered. The event selection requirements for the signal region are summarised in Table 3. The background composition in this region is also shown in Figure 5. The relative contribution of SM $t\bar{t}$ production to the total SM background amounts to 94.7% in this signal region.

7 Reconstruction of observables

The most relevant variable to separate signal from background is the invariant mass of the $t\bar{t}$ system, $m_{t\bar{t}}$, which needs to be reconstructed accurately. Additionally, angular variables sensitive to the $t\bar{t}$ spin correlations are used in both the 1- and 2-lepton channels to increase the sensitivity of the analysis. These provide additional discrimination power between the signal process, in which the $t\bar{t}$ system is produced in a spin-zero state, and the background from SM $t\bar{t}$ production, which involves $t\bar{t}$ systems in different spin and angular momentum states.

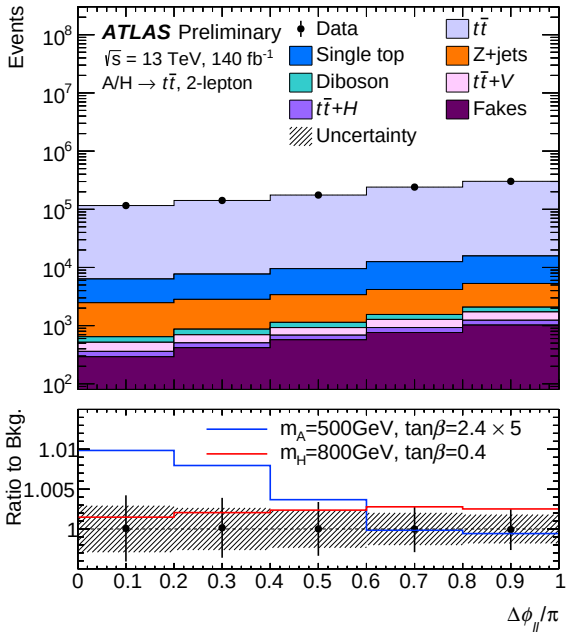


Figure 7: Distributions of the reconstructed $\Delta\phi_{\ell\ell}$ after the signal selection of the 2-lepton channel for the SM backgrounds. Each bin corresponds to a signal region. The expected relative deviation from the background prediction in the presence of an interference pattern, $(S + I + B)/B$, for two representative signal hypotheses is also shown in the ratio panel. All distributions and the uncertainty bands are obtained after the profile likelihood fit to the data under the background-only hypothesis.

Table 3: Summary of the event selection criteria for the signal regions in the 2-lepton channel.

Selection	**Criteria**
Common Selection	
Run and event cleaning	All detector components with acceptable conditions
Single lepton trigger	Separate single-electron or single-muon triggers
Exactly two leptons	2 (ee, $\mu\mu$, $e\mu$) with $p_{\mathrm{T}} > 25$ GeV. Leading one with $p_{\mathrm{T}} > 28$ GeV.
At least two jets	≥ 2 jets
b-tagging	≥ 1 b-tagged jet
Signal Selection	
Opposite-sign leptons	e^+e^-, $\mu^+\mu^-$, $e^+\mu^-$, $e^-\mu^+$
$E_{\mathrm{T}}^{\mathrm{miss}}$	$E_{\mathrm{T}}^{\mathrm{miss}} > 45$ GeV (ee and $\mu\mu$ channels only)
Dilepton invariant mass	$m_{ll} > 15$ GeV
Dilepton invariant mass	$m_{ll} < 81$ GeV or > 101 GeV (ee and $\mu\mu$ channels only)
Lepton-plus-b-jet invariant mass	$m_{lb} < 150$ GeV

7.1 Observables in the 1-lepton channel

The main discriminating variable in the 1-lepton channel is the $t\bar{t}$ invariant mass. In addition to the $t\bar{t}$ invariant mass, the angular variable $\cos\theta^*$ is calculated to define the ten orthogonal signal regions for

the Resolved regions, as discussed in Section 6. The reconstruction of both $m_{t\bar{t}}$ and $\cos\theta^*$ requires the correct identification and reconstruction of all $t\bar{t}$ decay products, including the four-momentum of the neutrino from the leptonic decay of one of the two top-quarks. In both the resolved and the merged event topologies, the transverse momentum of the neutrino is taken to be the transverse missing momentum in the event, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. The four-momentum component longitudinal to the beam axis p_z^ν is calculated from the kinematic constraint that the squared sum of the neutrino and charged lepton four-momenta must yield the squared mass of the W-boson [130]. If the resulting equation has exactly one real solution, it is taken as the neutrino longitudinal momentum. If it has no real solutions, the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ vector is rescaled and rotated in the transverse plane by the minimal amount required to obtain exactly one real solution [131]. This procedure is based on the assumption that the lack of a real solution is caused by a mismeasurement of the transverse missing energy. If two real solutions are found, in the resolved-topology selection, the choice is made by means of a χ^2 algorithm (7.1.1), while in the merged-topology selection the solution with the smaller absolute value of p_z^ν is chosen [131].

7.1.1 Resolved topology

A χ^2 minimisation approach is used to select the four jets from the $t\bar{t}$ decay from all selected small-R jets and assign them to the leptonically- and hadronically- decaying top quarks. It is defined as follows:

$$\chi^2 = \left[\frac{m_{jj} - m_{W_h}}{\sigma_{W_h}}\right]^2 + \left[\frac{(m_{jjb} - m_{jj}) - m_{t_h - W_h}}{\sigma_{t_h - W_h}}\right]^2 + \left[\frac{m_{jl\nu} - m_{t_l}}{\sigma_{t_l}}\right]^2 + \left[\frac{(p_{T,jjb} - p_{T,jl\nu}) - (p_{T,t_h} - p_{T,t_l})}{\sigma_{\mathrm{diff}\,p_T}}\right]^2 . \quad (2)$$

The first term is a constraint requiring the mass of a pair of jets m_{jj} to be close to the W-boson mass. The second term constrains the mass of a three-jet system m_{jjb} to be close to the mass of the hadronically decaying top quark but since m_{jj} and m_{jjb} are heavily correlated, the mass of the hadronically-decaying W-boson is subtracted to decouple this term from the previous one. The third term is used to constrain the mass of the leptonically decaying top quark. The last term constrains the magnitudes of the transverse momenta of the two top quarks to be similar, as expected for $pp \to t\bar{t}$ production at LO. In the χ^2 definition above, t_h and t_l refer to the hadronically and leptonically decaying top quarks. The values of the χ^2 central-value parameters m_{W_h}, $m_{t_h - W_h}$, m_{t_l}, and $p_{T,t_h} - p_{T,t_l}$, and the values of the width parameters σ_{W_h}, $\sigma_{t_h - W_h}$, σ_{t_l}, and $\sigma_{\mathrm{diff}\,p_T}$ are obtained from Gaussian fits to the distributions of relevant reconstructed variables, using signal MC events for which the lepton, the reconstructed neutrino and the three jets are matched to the lepton, neutrino and quarks from the hard scattering process. All possible jet permutations are tested, and the one with the lowest χ^2 is used; if there are two solutions for the neutrino longitudinal momentum, the one that yields the lowest χ^2 is kept. Any b-tagged jet can only be assigned to either of the b-quarks produced by the decay of a top-quark, reducing the number of permutations to test.

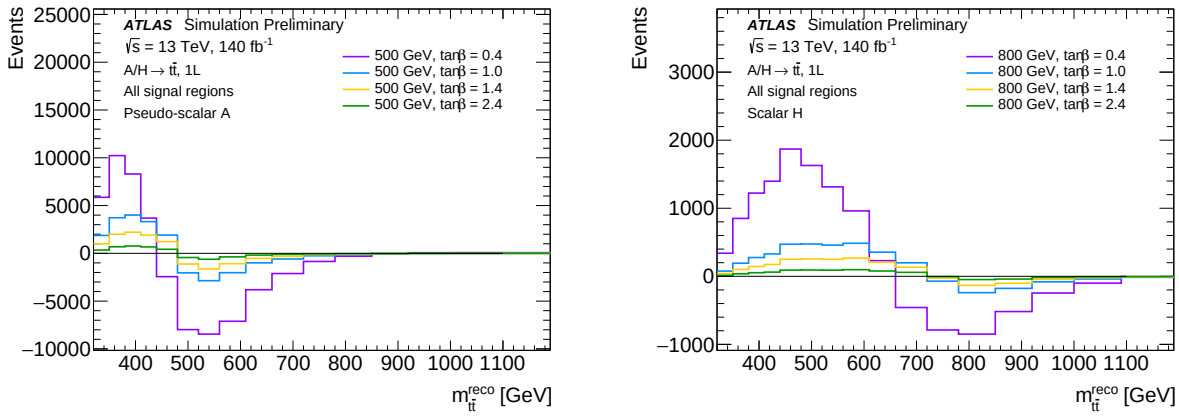
The estimate of the reconstructed $t\bar{t}$ invariant mass by the χ^2 algorithm is improved via in-situ calibrations that rely on the fact that the masses of the top quark and W boson are known to high accuracy. First the momenta of the two jets assigned to the hadronic W decay are scaled by a correction factor $\alpha = (80.4\ \mathrm{GeV})/m_{jj}$ so that their corrected invariant mass is equal to the W-boson mass. Then the momenta of the third jet assigned to the hadronic top-quark decay is scaled by a factor β that is based on the constraints that the total invariant mass of this jet and the two in-situ calibrated jets must yield the top-quark mass. The scale factor β is obtained as the positive solution to the quadratic equation $a\beta^2 + b\beta + c = 0$, where $a = m_b^2$,

$b = m_{jjb} - \alpha m_{jj} - m_b^2$, $c = 173.3\ \text{GeV} - \alpha m_{jj}$, and m_b is the reconstructed mass of the b-candidate jet. The scale factors α and β obtained with this method are in the ranges 0.90–1.00 and 0.88–1.10, respectively, and the jet energy corrections thus within ranges consistent with the uncertainty on the jet energy scale and the jet energy resolution for the jet collection used in this search [113]. No such scaling is applied to the jet assigned to the leptonically decaying top quark to avoid an over-correction of the jet momenta beyond the range compatible with the jet energy resolution and the uncertainty on the jet energy scale due to the poorer $E_{\text{T}}^{\text{miss}}$ resolution. This kinematic scaling improves the $m_{t\bar{t}}$ resolution[3] for the resolved topology by about 12%. The $m_{t\bar{t}}$ resolution for events in the resolved category varies between 12% for $m_{t\bar{t}} = 400$ GeV and 10% for $m_{t\bar{t}} = 1000$ GeV.

7.1.2 Merged topology

For the merged topology, the hadronically decaying top quark can be straightforwardly identified with the selected large-VR jet that passes the requirements in Section 5. If there is more than one large-VR jet passing these requirements, the one with the highest transverse momentum is taken as proxy for the hadronically decaying top quark. The leptonically decaying top quark is reconstructed from the selected jet identified with the b-jet from the leptonic top-quark decay (Section 6.1), the selected lepton, and the reconstructed neutrino (Section 6.1). The $m_{t\bar{t}}$ resolution for events in the merged category is around 10% for $m_{t\bar{t}} > 600$ GeV.

In Figure 8, the $S+I$ distributions in the reconstructed $t\bar{t}$ invariant mass are shown for a single pseudo-scalar A with mass 500 GeV and scalar with mass 800 GeV and different values of $\tan\beta$. The corresponding parton-level distributions can be found in Figure 2. The peak-dip structure of the signal-plus-interference pattern is clearly visible in the $m_{t\bar{t}}$ spectrum.



(a) $m_A = 500$ GeV (b) $m_H = 800$ GeV

Figure 8: Signal-plus-interference distributions in $m_{t\bar{t}}$ after the signal selection of the 1-lepton channel for (a) a single pseudo-scalar A with mass $m_A = 500$ GeV and (b) a single scalar H with mass $m_H = 800$ GeV for various values of $\tan\beta$. Events from all resolved and the merged signal region are included.

[3] The experimental resolution of the reconstructed $t\bar{t}$ invariant mass, $m_{t\bar{t}}$, is defined as the width of a Gaussian fit to the distribution $(m_{t\bar{t}} - m_{t\bar{t}}^{\text{afterFSR}})/m_{t\bar{t}}^{\text{afterFSR}}$ in pure-S MC samples, where $m_{t\bar{t}}^{\text{afterFSR}}$ denotes the combined invariant mass of the six $t\bar{t}$ decay products at the parton level after the emission of FSR.

7.2 Observables in the 2-lepton channel

The invariant mass of the $t\bar{t}$ system cannot be unambiguously reconstructed due to the presence of the two neutrinos from the two leptonically decaying top quarks. Therefore, instead of the $t\bar{t}$ invariant mass, the invariant mass of the two selected b-jets and the two charged leptons, $m_{\ell\ell bb}$ is used as a discriminating variable. In addition, the azimuthal angle between the two leptons, $\Delta\phi_{\ell\ell}$ is calculated to define the five orthogonal signal regions of the 2-lepton channel, as discussed in Section 6.

8 Estimation of background contributions from data

8.1 W+jets background in the 1-lepton channel

Scale factors derived from data are applied before the final fit to correct the normalisation of the W+jets background in the Sherpa MC simulation samples for possible mis-modelling of the cross-section times acceptance. The scale factors are determined by comparing the measured W-boson charge asymmetry in data [132] with that predicted by the simulation. The total number of W+jets events in data in a given signal region, $N_{W^+} + N_{W^-}$, is given by

$$N_{W^+} + N_{W^-} = \frac{r_{\mathrm{MC}} + 1}{r_{\mathrm{MC}} - 1}(D_{\mathrm{corr}+} - D_{\mathrm{corr}-}), \tag{3}$$

where r_{MC} denotes the ratio of the number of W+jets events with a positively charged lepton to that with a negatively charged lepton obtained from the Sherpa MC simulation and $D_{\mathrm{corr}+(-)}$ refers to the number of observed data events with a positively (negatively) charged lepton in the same signal region. This method relies on the fact that the charge asymmetry obtained in MC simulation is in excellent agreement with the value measured in data [133]. Contributions to $D_{\mathrm{corr}+(-)}$ from charge-asymmetric processes such as single-top, WZ, and $t\bar{t} + W$ production are estimated from MC simulation and are subtracted from the data samples. Contributions from charge-symmetric processes such as $t\bar{t}$ production cancel out in the difference on the right-hand side of Eq. 3. The final scale factor, C_A, for a given signal region is then calculated as the ratio of $N_{W^+} + N_{W^-}$ evaluated from data to that predicted from the Sherpa MC simulation. The scale factors are evaluated separately for the e+jets and μ+jets events in the Resolved $2b$, Resolved $1b$, and Merged regions, respectively, without splitting the resolved regions into angular bins. Their values agree within their statistical uncertainties and with the scale factors obtained inclusively for all e+jets and μ+jets events across all signal regions. The latter, $C_A = 1.125 \pm 0.031$, is used to correct the normalisation of the generated W+jets samples. An additional conservative 20% uncertainty is assigned on the W+jets normalisation after the C_A correction to account for potential residual mis-modellings of the relative contributions from W-boson production in association with heavy-flavour jets [131].

8.2 Multijet background in the 1-lepton channel

The multijet background in events satisfying the resolved- or merged-selection criteria consists of events with a non-prompt lepton or a jet misreconstructed as a lepton that satisfies the lepton identification and isolation criteria applied in this search. These are referred to as "tight" lepton requirements in the following. In the μ+jets channel, muons arising from semileptonic decays of hadrons inside jets constitute the main source of this background. In the e+jets channel, additional multijet background arises from

events containing jets with a large electromagnetic component, for example from $\pi \rightarrow \gamma\gamma$ decays, or photons mis-identified as isolated electrons. The normalisation, $m_{t\bar{t}}^{\text{reco}}$ shape, as well as statistical and systematic uncertainties associated with the multijet background are estimated from data using the *matrix method*. The matrix method used in this search is based on the one used in previous ATLAS $t\bar{t}$ resonance searches and measurements [130, 131].

The matrix method relies on an alternative, looser lepton definition that is based on a set of loose identification criteria [124] without any additional isolation requirements applied. The number of multijet events in a given signal region can then be estimated by solving a set of two equations describing the composition of events in two regions: the signal region with its tight lepton requirements and a region defined by applying the same criteria as in the signal region, except that the loose lepton requirements are applied. The number of events with leptons satisfying the loose lepton identification criteria, N_L, is defined as

$$N_{\text{L}} = N_{\text{prompt}} + N_{\text{multijet}}, \tag{4}$$

where N_{prompt} and N_{multijet} denote the number of events with prompt leptons and events with leptons from other sources, respectively, satisfying the loose identification criteria. The number of events with leptons satisfying the tight lepton identification criteria applied in the signal regions, N_T, can be written as

$$N_{\text{T}} = \epsilon \times N_{\text{prompt}} + f \times N_{\text{multijet}}, \tag{5}$$

where the *real rate* ϵ is an estimator for the probability of a prompt lepton passing the loose lepton identification criteria to also pass the tight critieria. Similarly, *fake rate* f is an estimator for the probability that a lepton from other sources that passes the loose criteria also passes the tight criteria. The fake efficiency f is estimated from a control region, defined by applying the same selection criteria as for the resolved signal region, but with the missing transverse momentum and transverse mass requirements inverted. Contributions from processes leading to prompt leptons are subtracted from the data in this multijet enriched control region using MC simulation. The real efficiency ϵ is determined using SM $t\bar{t}$ MC samples. Solving Equations 4 and 5 for N_{prompt} and N_{multijet} yields the multijet contribution in the given signal region. The multijet estimate is performed separately for the e+jets and μ+jets channels.

Good modelling of the shape of kinematic distributions, in particular $m_{t\bar{t}}^{\text{reco}}$, is achieved by parameterising the real and fake rates as functions of the transverse momentum of the lepton and a calorimeter-based isolation variable. In addition, the rates are derived separately for the two cases $\Delta R(\ell, j) > 0.4$ and $\Delta R(\ell, j) \leq 0.4$, where j denotes the nearest selected jet. The fake rates for electrons vary from 5% to 86%, with the largest values occurring at high electron p_{T}, with low nearby calorimeter activity. This behaviour is explained by the track-based lepton isolation criterion that uses a p_{T}-dependent cone and leads to a looser isolation requirement at higher p_{T}. The fake rates for muons vary from 9% to 84%, with the largest values occurring at low muon p_{T}, with low nearby calorimeter activity, which is typical for soft muons arising from semileptonic decays of hadrons inside jets.

The modelling is validated in separate dedicated validation regions, where only one of the $E_{\text{T}}^{\text{miss}}$ or $E_{\text{T}}^{\text{miss}} + m_T^W$ requirements is inverted. These validation regions contain a background composition that is more similar to that in the signal region while still having an enhanced multijet contribution. A conservative 50% normalisation uncertainty is assigned to the multijet background based on the modelling performance in the validation regions.

8.3 Z+jets background in the 2-lepton channel

Data-driven corrections are applied to the SHERPA MC simulation samples for the Z+jets background in the 2-lepton channel before the final fit to correct the overall normalisation of this background component as well as the modelling of relevant kinematic variables. The corrections are derived in the Z+jets enriched CR_Z where a poor modelling of background prediction obtained with the SHERPA MC simulated samples is observed for two kinematic variables in particular: the transverse momentum of the di-lepton system, $p_T^{\ell\ell}$, and the invariant mass $m_{\ell\ell bb}$. In both cases, the mis-modelling exhibits a linear trend, with good agreement observed for small values of these variables and a linear increase in the difference between data and MC prediction towards higher values. A reweighting procedure in $m_{\ell\ell bb}$ is used to correct the SHERPA MC samples for the Z+jets background to the data in CR_Z. Each simulated event is assigned a multiplicative correction weight based on its $m_{\ell\ell bb}$:

$$w = a \cdot (1 - b \cdot m_{\ell\ell bb}).$$

The correction factors a and b are derived from a maximum likelihood fit of the $m_{\ell\ell bb}$ distributions, taking into account only the statistical uncertainty on the MC predictions. The following values are obtained for the correction factors:

$$a = 1.193 \pm 0.003, \quad b = -1.7 \pm 0.1 \cdot 10^{-4}\ \text{GeV}^{-1}.$$

Good agreement between the data and the SM expectation is observed for all relevant kinematic variables after applying the reweighting correction. A systematic uncertainty is assigned to the reweighting procedure based on a comparison of the nominal reweighting with an equivalent reweighting approach based on $p_T^{\ell\ell}$. It amounts to a 3% variation of the correction factor a. In the final fit, an additional conservative 30% normalisation uncertainty is applied to the Z+jets background to cover any residual mismodellings in the production of heavy-flavour jets [134].

8.4 Fakes background in the 2-lepton channel

The Fakes background is estimated using MC simulation and the modelling is validated in the $\mu\mu$ and $e\mu$ channels of the fake-enriched control region CR_f (Section 6.1). First, the modelling of background events arising from fake muons is studied in various representative kinematic variables in the $\mu\mu$ channel, such as the leading and sub-leading muon transverse momenta, the di-muon invariant mass or the W-boson transverse masses for the leading and sub-leading muons. The MC-based predictions are found to be in good agreement with the data within a conservative 30% normalisation uncertainty on the MC-based fake-muon prediction. Second, the modelling of background events with fake electrons is studied in the $e\mu$ channel, using an equivalent set of kinematic variables. Again, good agreement between the MC-based background prediction and the data is found within a conservative 30% normalisation uncertainty.

9 Systematic uncertainties

The modelling of signal and all background events is affected by various experimental uncertainties related to the reconstruction, identification, and calibration of object and event properties. In addition, uncertainties on the theoretical modelling of the simulated backgrounds, most importantly SM $t\bar{t}$ production,

as well as on the data-driven background estimates and corrections are taken into account. Some of the uncertainties affect both the shape and the normalisation of the $m_{t\bar{t}}$ and $m_{\ell\ell bb}$ spectra, while others affect the normalisation only. The sources of the largest systematic uncertainties in the analysis are related to the modelling of the SM $t\bar{t}$ background processes, followed by uncertainties on the jet energy scale and resolution.

9.1 Modelling uncertainties

A range of modelling uncertainties is estimated for the dominant and irreducible background from SM $t\bar{t}$ production. These uncertainties can largely be grouped into three categories:

1. Uncertainties on the higher-order prediction to which the NLO MC sample is reweighted, including uncertainties on the reweighting method itself;
2. Uncertainties obtained from a comparison to alternative SM $t\bar{t}$ MC samples;
3. Uncertainties on the NLO+PS prediction obtained from the nominal SM $t\bar{t}$ MC sample without a corresponding uncertainty on the higher-order prediction.

Category 1 comprises uncertainties related to scale variations in the fixed-order calculations, to the choice of PDF set, and to uncertainties on the EW contribution, which is not present in the NLO+PS prediction. The scale uncertainties are estimated separately for the (anti-)top p_{T} and $m_{t\bar{t}}$ distributions by varying independently the renormalisation and factorisation scales up and down by a factor of two. The PDF uncertainty is estimated as the envelope of the intra-PDF uncertainties of the LUXQED PDF set [135], which is used to obtain the NNLO-QCD+NLO-EW predictions (Section 4). The uncertainty on the EW contribution is estimated by comparing the nominal spectra, obtained with the LUXQED PDF set, to spectra obtained with the NNPDF3.0QED PDF set. This variation has been found to significantly alter the EW part of the prediction, as the two PDF sets rely on a different treatment of the photon PDF. An additional uncertainty is derived to describe possible differences between the parton-level top-quark definitions in the theory and NLO MC predictions, which are due to the fact that the latter includes any number of real emissions from the final-state top quarks, while the NNLO prediction includes at most two real emissions. The resulting uncertainty on the higher-order prediction is estimated by comparing the nominal SM $t\bar{t}$ spectra to those obtained by reweighting an alternative NLO+PS MC sample created with the same generators and settings as the nominal NLO sample but without decaying the top quarks and forbidding PS emissions after the first one, thus obtaining a parton-level NLO+PS prediction coherent with the NNLO prediction. An uncertainty on the reweighting method itself is derived by comparing the nominal reweighted SM $t\bar{t}$ sample to a sample obtained through an alternative reweighting that is first applied twice on the (anti-)top quark p_{T} and then in $m_{t\bar{t}}$.

Uncertainties in Category 2 are estimated by comparing the nominal SM $t\bar{t}$ prediction to alternative predictions obtained from alternative SM $t\bar{t}$ MC samples All alternative MC samples are reweighted to the same higher-order predictions as the nominal Powheg v2 +Pythia 8.230 MC sample. This is done since none of these systematic uncertainties on the NLO+PS prediction are meant to affect the parton-level variables used for the reweighting, but have a non-negligible effect on the final observables due to different correlations between these kinematical variables, as well as other properties of the event that affect acceptance and shapes of the kinematic variables of interest. The uncertainties in Category 2 are therefore referred to as *residual uncertainties*. The uncertainty due to the choice of the PS and hadronisation model is estimated by comparing the nominal predictions to those obtained from a sample

generated at NLO in QCD with POWHEG BOX v2 with the same PDF set and h_{damp} value as for the nominal sample but interfaced with HERWIG 7.713 [136, 137], using the H7UE set of tuned parameters [137] and the MMHT2014LO PDF set [138]. The uncertainty related to the ME-PS matching is estimated by comparing the predictions obtained with the nominal sample to those obtained from an alternative sample generated with the same generator settings as the nominal sample but setting the POWHEG:PTHARD parameter in PYTHIA to 1 instead of 0 [139]. This parameter regulates the definition of the vetoed region of the showering to avoid holes or overlaps in the phase space filled by POWHEG and PYTHIA. This estimate of the uncertainty follows the description in Ref. [140] and replaces the comparison with an alternative sample generated with MADGRAPH5_AMC@NLO that was used in previous ATLAS searches in $t\bar{t}$ final states [30, 130, 131]. An additional uncertainty, referred to as *lineshape* uncertainty, related to the treatment of $t\bar{t}$ spin correlations in different generators is estimated by comparing the predictions obtained with the nominal sample to those obtained with an alternative sample generated with the same setup as the nominal sample but using MADSPIN to decay the top quarks. An uncertainty related to the choice of the h_{damp} parameter is estimated by comparing the predictions of the nominal sample to those obtained with an alternative sample with the h_{damp} parameter increased by a factor of 1.5 compared to its nominal value. In all four cases, the resulting one-sided uncertainties are symmetrised to obtain effective "up" and "down" variations. Variations in the ISR are estimated by varying the factorisation and renormalisation scales independently up and down by a factor of two. Similarly, the uncertainty related to final-state radiation (FSR) is assessed by varying the renormalisation scale for final-state parton-shower emissions up and down by a factor of two. In both cases, the variations are obtained from internal weights of the NLO+PS generator and the resulting alternative samples are reweighted to the higher-order predictions.

Category 3 comprises uncertainties on the NLO+PS prediction without a corresponding uncertainty on the higher-order predictions. These uncertainties are kept without "reducing" them, i.e. without reweighting them separately to the higher-order predictions. Instead, the nominal reweighting is applied on each of the alternative MC predictions. The main uncertainty of Category 3 is that related to the variations of the renormalisation scale in the ISR parton shower, obtained via the corresponding systematic variation in the A14 tune, and accessed through internal generator weights. The uncertainty on the $m_{t\bar{t}}$ and $m_{\ell\ell bb}$ spectra arising from the uncertainty in the top quark mass is evaluated by comparing the spectra obtained using the nominal sample to those generated with top quark masses of 170.0 and 175.0 GeV, and multiplying the difference by 0.7 to approximate a one standard deviation uncertainty, corresponding to the ±0.76 GeV uncertainty on the top-quark mass world average [141]. This uncertainty is of Category 3 to avoid reducing its impact by reweighting the alternative samples to the same higher-order prediction estimated for the central m_{top} value of 173.3 GeV.

Additionally, the uncertainty on the SM $t\bar{t}$ cross-section is applied as a pure normalisation uncertainty. It is calculated by summing in quadrature the uncertainties related to scale, PDF+α_s, and m_{top} variations with respect to the nominal value (Section 4). This yields a $^{+5.6\%}_{-6.1\%}$ variation.

The main uncertainty on the modelling of single-top quark production, arises from a comparison of the diagram removal and the alternative diagram subtraction scheme for the treatment of interference effects and overlap between SM $t\bar{t}$ and tW production. It is estimated by comparing the nominal tW samples, generated with the diagram removal scheme, to a set of alternative samples obtained with the diagram subtraction scheme [60, 76] using the same generator settings as for the nominal single-top samples. As in the case of the SM $t\bar{t}$ background, an uncertainty related to the choice of the PS and hadronisation model is estimated for the tW background component by comparing the nominal prediction to that obtained with the same ME generator settings but interfaced with HERWIG 7.713, using the H7UE set of tuned parameters and the MMHT2014LO PDF set. The uncertainty related to the ME-PS matching is estimated by comparing the

predictions obtained with the nominal sample to those obtained from an alternative sample generated with the same generator settings as the nominal sample but setting the POWHEG:pTHARD parameter in PYTHIA to 1 instead of 0. Modelling uncertainties related to the choice of the renormalisation and factorisation scales, the levels of ISR and FSR, and the choice of PDF set have also been evaluated but are found to be negligible. The theoretical uncertainties on the cross-sections for tW-, t-, and s-channel production (Section 4) are applied as pure normalisation uncertainties on the respective components. They are estimated by summing in quadrature the scale, PDF, and α_s uncertainties on the nominal cross-sections and amount to ±5.4% for tW production, $^{+4.3\%}_{-3.7\%}$ for t-channel production, and $^{+4.4\%}_{-4.1\%}$ for s-channel production.

Both shape and normalisation uncertainties are taken into account for the background from W+jets production in the 1-lepton channel. In addition to the ±20% normalisation uncertainty covering any possible remaining mis-modellings of the W+jets flavour fractions after the data-driven C_A correction factors are applied (Section 8.1), uncertainties related to the choice of scales and PDF set are taken into account for the MC modelling of this background component. The latter affect both the shape and normalisation of the W+jets background spectra but are found to be negligible and pruned from the final likelihood fit. In the 2-lepton channel, no separate uncertainty is assigned to the background from W+jets production as it is included in the fakes background component, the uncertainty on which is described below.

The main systematic uncertainty on the Z+jets background in the 2-lepton channel is the uncertainty related to the data-driven reweighting described in Section 8.3. It affects both the shape and normalisation of this background component. An additional conservative 30% normalisation uncertainty is assigned to cover any residual mismodellings in the production of heavy-flavour jets (Section 8.3). In the 1-lepton channel, where Z+jets production constitutes an even smaller background component, a conservative 30% normalisation uncertainty is applied, which covers both the ±5% uncertainty on the Z+jets production cross-section and acceptance uncertainties related to possible mismodellings of higher jet multiplicities.

In the case of the remaining small backgrounds, only normalisation uncertainties are applied. A conservative 50% normalisation uncertainty is assigned to the data-driven estimate of the multijet background in the 1-lepton channel (Section 8.2). The normalisation uncertainty on the fakes background in the 2-lepton channel amounts to 30% (Section 8.4). For the backgrounds from $t\bar{t}Z$, $t\bar{t}W$, and $t\bar{t}h$ production, the uncertainties on the respective higher-order cross-sections (Section 4) are taken into account. They amount to $^{+10.4\%}_{-12.0\%}$, $^{+13.3\%}_{-12.0\%}$, and $^{+6.8\%}_{-9.8\%}$, respectively. A conservative 50% normalisation uncertainty is applied on the diboson background to take into account any possible mismodelling in the production of additional jets [142] and heavy-flavour jets [143].

Modelling uncertainties are also taken into account for the signal (S) and signal-plus-interference ($S + I$) spectra. Uncertainties due to the choices of renormalisation and factorisation scales are estimated by varying the two scales independently up and down by a factors of 2.0 and 0.5. Uncertainties related to the choice of PDF set and α_s are derived based on the PDF4LHC15 prescription. The uncertainty related to the uncertainty on the top-quark mass is evaluated by comparing the S or $S + I$ spectra obtained for $m_t = 173.3$ GeV to spectra obtained for values of m_t varied by ±0.76 GeV from the nominal value. No uncertainty is applied on the signal MC reweighting as the agreement between generated test samples and reweighted samples has been found to be within statistical uncertainties.

9.2 Experimental uncertainties

The dominant experimental uncertainties affecting the $m_{t\bar{t}}$ and $m_{\ell\ell bb}$ spectra are the uncertainties on the jet energy scale (JES) and resolution (JER). These are evaluated from a combination of simulations, test-beam data, and in situ measurements. Additional contributions from jet flavour composition, punch-through, single-particle response, calorimeter response to different jet flavours and pile-up are taken into account. The jet flavour uncertainties, which were among the dominant JES uncertainties in the previous analysis [30], were significantly reduced compared to those given in Ref. [113]. This was achieved by re-calibrating to remove the dependence on the MC hadronisation model [144], and considering uncertainties on quark, gluon, charm and bottom quarks derived from comparisons between models which agree with previous measurements, and also constraining these for the flavour mixture where the JES is measured in data with *in situ* techniques. The JER uncertainties are obtained with an in situ measurement of the jet response in di-jet events [113]. Uncertainties related to the jet mass scale (JMS), which are propagated to the large-VR jet uncertainties, are derived using the R_{trk} method [145], which compares the ratio of the mass of jets reconstructed from calorimeter clusters to that of jets reconstructed from tracks in data and MC simulation. An uncertainty is assigned on the efficiency of the JVT requirement on jets [118].

Further experimental uncertainties on the correction factors for the b-tagging efficiencies [146], as well as the rates of mis-tagging c-[147], and light-flavour jets [148] are applied to the simulated event samples by looking at dedicated flavour-enriched samples in data. An additional b-tagging uncertainty is applied for high-momentum jets ($p_{\text{T}} > 400$ GeV) to account for uncertainties on the modelling of the track reconstruction in high-p_{T} environments. It is calculated from simulated events by considering variations on the quantities affecting the b-tagging performance such as the impact parameter resolution, percentage of poorly measured tracks, description of the detector material, and track multiplicity per jet. The dominant effect on the uncertainty when extrapolating to high jet p_{T} is related to the different tagging efficiency when smearing the track impact parameters based on the resolution measured in data and simulation.

Smaller experimental uncertainties are related to the efficiencies of the lepton identification, isolation, and reconstruction, as well as the lepton energy scale and resolution [149, 150] and to the scale and resolution of the track soft-term in the $E_{\text{T}}^{\text{miss}}$ calculation [128]. Variations in the reweighting applied to simulated samples to match the expected mean number of interactions observed in each bunch crossing in data are included. They cover the uncertainty on the ratio between the predicted and measured inelastic cross-section. A constant 0.83% normalisation uncertainty is applied to all signal and background samples, except multijet and W+jets, which are estimated from data. It accounts for the uncertainty on the integrated luminosity (Section 4).

9.3 Uncertainty correlation scheme

All experimental uncertainties are treated as fully correlated across samples and the signal regions of the 1- and 2-lepton channels in the final profile-likelihood fit (Section 10). The modelling uncertainties are treated as uncorrelated between all signal and background samples, with the exception of the uncertainty on the top-quark mass m_{top}, which is treated as correlated between signal, signal-plus-interference, and SM $t\bar{t}$ samples. Additionally, the uncertainties on the SM $t\bar{t}$ background related to the modelling of the parton shower and hadronisation, the PS-ME matching, and the choice of the h_{damp} parameter are treated as uncorrelated between all 11+5 signal regions of the 1- and 2-lepton channels. These uncertainties are obtained by comparing the nominal to an alternative SM $t\bar{t}$ sample in each case ("two-point systematics"), an approach known to yield conservative uncertainties with non-negligible constraints. Additionally, the uncertainties

on the SM $t\bar{t}$ background related to the (inclusive) $t\bar{t}$ cross-section, the choice of renormalisation and factorisation scales, and the levels of ISR/FSR cannot be assumed to be fully correlated across different kinematic regimes and are therefore treated as uncorrelated between the resolved and merged signal regions of the 1-lepton channel and between the 1- and 2-lepton channels, although they are treated as correlated across the different angular bins in $\cos\theta^*$ and $\Delta\phi_{\ell\ell}$. All other modelling NPs, such as the one related to the top-quark mass uncertainty, are treated as correlated across all signal regions.

9.4 Uncertainty impact

The relative importance of the different categories of systematic uncertainties is quantified via their post-fit impact on the observed signal strength for two representative signal hypotheses. The $t\bar{t}$ modelling systematics are found to have the largest impact for the majority of the tested signal hypotheses. The dominant uncertainties on this irreducble backgrounds are those related to the NNLO estimate (reweighting, scales, PDFs), the ME-PS matching (pTHARD, h_{damp}), and modelling of ISR and FSR. The largest experimental systematics are those related to the JES and JER of the small-R jets, followed by b-tagging related systematics.

10 Statistical data analysis

The agreement between the data and the SM prediction (null hypothesis) as well as different signal hypotheses is quantified using a profile-likelihood fit [151] of the expected distributions in $m_{t\bar{t}}$ (1-lepton channel) and $m_{\ell\ell bb}$ (2-lepton channel) and the observed ones. The fit is performed simultaneously in the eleven signal regions of the 1-lepton and the five signal regions of the 2-lepton channels. The statistical and systematic uncertainties are taken into account as nuisance parameters (NPs) in the fit with a correlation scheme as described in Section 9.3.

The shape of the binned $m_{t\bar{t}}$ or $m_{\ell\ell bb}$ distributions in the presence of a signal interfering with the background is parameterised in terms of the signal strength μ:

$$\mu\ S + \sqrt{\mu}\ I + B = (\mu - \sqrt{\mu})\ S + \sqrt{\mu}\ (S + I) + B. \tag{6}$$

The terms S and $S + I$ on the right-hand side of Eq. 6 denote the $m_{t\bar{t}}$ or $m_{\ell\ell bb}$ distributions obtained from the S and $S + I$ samples for a given signal hypothesis, respectively, while B denotes the corresponding distributions for the total background expected under the SM-only hypothesis. The fitted variable is $\sqrt{\mu}$ and the case $\mu = 1$ ($\mu = 0$) corresponds to the signal hypothesis under consideration (the background-only hypothesis). In some cases, in which interference effects are negligible, Eq. 6 reduces to the common parameterisation $\mu\ S + B$ and then μ can be interpreted as the ratio of the observed to the predicted value of the (pure-)signal cross-section times the branching fraction to $t\bar{t}$. It should be noted that this approach relies on the assumption that, for a given signal hypothesis, the $m_{t\bar{t}}$ ($m_{\ell\ell bb}$) shapes for S and $S + I$ in Eq. 6, and hence the width of the interference pattern, do not change with μ.

The likelihood used in this analysis is given by [152]:

$$\mathcal{L}(\text{data}|\sqrt{\mu}, \boldsymbol{\theta}_\mu) = \prod_{c=1}^{N_{\text{cats}}} \mathcal{L}_c(\text{data}|\sqrt{\mu}, \boldsymbol{\theta}_{\sqrt{\mu}}) \prod_{k=1}^{N_{\text{cons}}} \mathcal{F}(\tilde{\theta}_{\sqrt{\mu},k}|\theta_{\sqrt{\mu},k}) \tag{7}$$

where $\boldsymbol{\theta}$ is the vector of nuisance parameters (NPs), N_{cats} is the number of categories, N_{cons} is the number of constrained NPs, $\tilde{\theta}_k$ is the global observable corresponding to θ_k, c is the index for the categories, k is the index for the constrained NPs, and $\mathcal{F}$ denotes a Poisson, a Gaussian or a Log-normal distribution depending on the type of uncertainty.

The agreement of the data with a set of signal hypotheses (search stage) is quantified using the CL_s frequentist formalism [153] with the profile likelihood ratio test statistic (q_0), defined as [151]:

$$q_0 = \frac{\mathcal{L}(0, \hat{\hat{\boldsymbol{\theta}}}_0)}{\mathcal{L}(\hat{\sqrt{\mu}}, \hat{\boldsymbol{\theta}}_{\hat{\sqrt{\mu}}})}, \tag{8}$$

where the numerator is evaluated for the set of NPs $\hat{\hat{\boldsymbol{\theta}}}_0$ that maximise $\mathcal{L}$ for $\sqrt{\mu} = 0$, and the denominator is evaluated for the values $\hat{\sqrt{\mu}}$ and $\hat{\boldsymbol{\theta}}_{\hat{\sqrt{\mu}}}$ that jointly maximise the likelihood. A best-fit value $\hat{\sqrt{\mu}} \neq 1$ indicates agreement with an interference pattern that is different from the one predicted by the signal hypothesis under consideration, i.e. the signal hypothesis for which the S and $S + I$ templates in Equation 6 are obtained. The search stage thus does not only quantify the compatibility of the data with one specific signal hypothesis, e.g. the unique interference pattern obtained for a single scalar of mass m_H and relative width Γ_H/m_H ($\sqrt{\mu} = 1$). It also tests the compatibility of the data with the broader set of interference patterns obtained from the S and $S + I$ templates for the original ($\sqrt{\mu} = 1$) hypothesis by varying the value of $\sqrt{\mu}$ according to Equation 7. A best-fit value of $\hat{\sqrt{\mu}} \neq 1$ indicates better compatibility of the data with an interference pattern that is different from the one obtained under the original hypothesis, for example a peak-peak rather than a peak-dip structure, see e.g. Figure 3. In this context, it is worth noting that $g^2_{A/Ht\bar{t}}$ is equivalent to $\sqrt{\mu}$ according to Equation 1. Variations of $\sqrt{\mu}$ thus yield a set of interference patterns like the ones shown in Figure 3 for variations of $g^2_{A/Ht\bar{t}}$ for two represenative parameter points in the generic benchmark scenario.

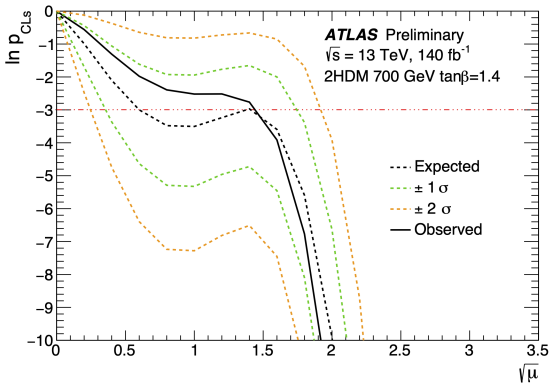
In the absence of any significant deviation from the background-only hypothesis (see Section 11), the level at which a given signal hypothesis is excluded by the data is also quantified with the CL_s frequentist formalism (exclusion stage) but using a different test statistic compared to the one used at the search stage. It is defined as the simple likelihood ratio of the two values of the likelihood obtained for $\sqrt{\mu} = 1$ (signal hypothesis) and $\sqrt{\mu} = 0$ (background-only hypothesis):

$$q_{1,0} = -2\ln\frac{\mathcal{L}(1, \hat{\hat{\boldsymbol{\theta}}}_1)}{\mathcal{L}(0, \hat{\hat{\boldsymbol{\theta}}}_0)}. \tag{9}$$

The asymptotic formula of this test statistic, which provides an analytic approximation of its sampling distribution [154, 155], has been taken from Ref. [151]. Its implementation in the statistical framework and general validity of the asymptotic approximation in the context of this analysis has been verified using toy experiments. The use of this test statistic for the calculation of exclusion regions is motivated by the fact that at the exclusion stage, the goal is to quantify the rejection of the specific benchmark scenario under consideration, which corresponds to $\sqrt{\mu} = 1$. Values of $\sqrt{\mu} \neq 1$ do not yield interference patterns compatible with the signal model under consideration because, as stated above, the width of the interference pattern, which is determined by the S and $S + I$ templates, does not change in the fit with the signal strength (or, equivalently, the coupling modifier $g_{A/Ht\bar{t}}$), while in a realistic model, the width is a function of the coupling to $t\bar{t}$. In a type-II 2HDM, for example, the latter is determined by the value of $\tan\beta$ (Section 2) The choice of test statistic for the exclusion stage thus differs from that for the search stage, where the aim is to test whether the SM hypothesis should be rejected in favour of an alternative signal hypothesis that

agrees better with the data. In this case, it is sensible to consider the wider range of interference patterns obtained for values $\sqrt{\mu} \neq 1$.

A signal hypothesis is excluded at 95% confidence level (CL) if the CL_s value, p_{CL_s}, for $\sqrt{\mu} = 1$ is smaller than 0.05. In this context, it should be noted that the quadratic dependence of the likelihood function on $\sqrt{\mu}$ can lead to a non-monotonic behaviour of p_{CL_s} as a function of $\sqrt{\mu}$, as shown in Figure 9 for one of the signal hypotheses considered in this paper. This means that for some signal hypotheses, the p_{CL_s} function may cross the value of $p\mathrm{CL}_s = 0.05$ multiple times, leading to disjoint exclusion intervals. In these cases, no unique upper limit on the signal strength can be defined. Additionally, the crossing point(s) for the median and the $N\sigma$ bands may vary significantly for the same signal hypothesis. Furthermore, the behaviour of the CL_s scans varies across signal hypotheses, with some scans exhibiting only unique crossing points or even a monotonous behaviour, and others exhibiting several crossing points in either the median and/or $N\sigma$ ($N \in \{\pm 1, \pm 2\}$) bands.



(a)

Figure 9: Distribution of the logarithm of the CL_s value, $\ln(p_{\mathrm{CL}_s})$ as a function of $\sqrt{\mu}$ for the signal hypothesis with $m_A = m_H = 700$ GeV and $\tan\beta = 1.4$ showing non-monotonic behaviour for the observed and expected median, $\pm 1\sigma$, and $\pm 2\sigma$ bands. The red horizontal line corresponds to $p_{\mathrm{CL}_s} = 0.05$. Values of $\sqrt{\mu}$ with $p_{\mathrm{CL}_s} < 0.05$ are excluded at 95% CL.

This behaviour of p_{CL_s} as a function of $\sqrt{\mu}$ has implications for the calculation of two-dimensional exclusion regions in the parameter spaces of the different benchmark models. Instead of calculating upper limits on $\sqrt{\mu}$ for selected signal hypotheses and then using an interpolation technique to obtain the contours corresponding to $\sqrt{\mu} = 0.05$ to define the parameter regions excluded at 95% CL, a different approach is chosen in this paper. The 95% CL exclusion regions are obtained by calculating the observed (expected) values of p_{CL_s} using the test statistic in Equation 9 for each point in a fine, uniform grid of points in the parameter plane of interest. Each point corresponds to a different signal hypothesis, for which S and $S + I$ templates have been obtained via the reweighting technique described in Section 4. A linear interpolation between the values of $\ln(p_{\mathrm{CL}_s})$ is then used to determine the observed (expected) exclusion contour corresponding to $p_{\mathrm{CL}_s}(\sqrt{\mu} = 1) = 0.05$. This approach avoids the ambiguities in the definition of upper limits on $\sqrt{\mu}$ in the presence of signal-background interference effects that necessitate a quadratic likelihood parameterisation.

The non-linear dependence of the likelihood and hence CL_s value on $\sqrt{\mu}$ also has consequences for the calculation of the $N\sigma$ bands. These are commonly understood as an indication of how frequently each signal hypothesis would be excluded under the background-only hypothesis, with the edges of the $\pm 1\sigma$

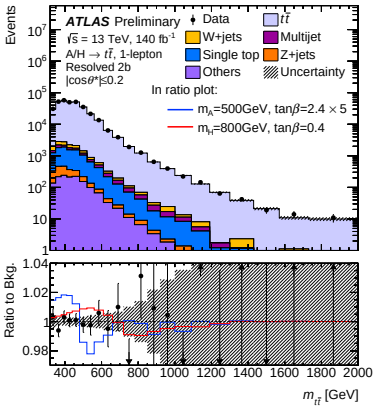
($\pm 2\sigma$) bands marking the range of signal hypotheses that would be excluded under the background-only hypothesis in 68% (95%) of equivalent searches. Common ATLAS statistics tools rely on a slightly different definition of the $N\sigma$ exclusion regions. The latter are obtained from a set of Asimov datasets that are representative of $N\sigma$ fluctuations under the background-only hypothesis, referred to as $N\sigma$ Asimov datasets. The exclusion limit for a given $N\sigma$ Asimov dataset defines the $N\sigma$ exclusion region, referred to as $N\sigma$ *limit contour* in the following. If the likelihood function is quadratic in the POI, as is the case for models including signal-background interference, unphysical crossings of the limit contours with the median exclusion limit can occur. Such crossings contradict the simple Frequentist interpretation of the $N\sigma$ exclusion regions outlined above. Therefore, an alternative, more appropriate way of defining the $N\sigma$ exclusion regions, referred to as *limit bands*, is chosen in the analysis presented in this paper. These limit bands represent the hypotheses that have an exclusion rate between $\Phi(-N)$ and $\Phi(N)$, where $\Phi(X)$ is the standard Gaussian cumulative distribution function and $\Phi(N)$ is the probability of N^{th} normile. This approach to determine the $N\sigma$ exclusion regions more accurately reflects the Frequentist interpretation of the $N\sigma$ exclusion regions and, by construction, avoids unphysical crossings of the $N\sigma$ band edges and the median expected limit. It should be noted in the case of a likelihood that is linear in the POI, the limit contours approximate the limit bands very well.

11 Results

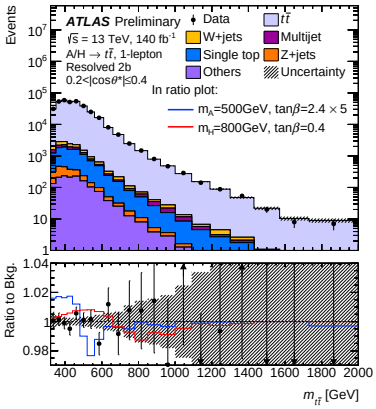
The level of agreement between the observed data and the SM is tested in a fit under the background-only hypothesis ($\mu = 0$) in which only the nuisance parameters are allowed to vary. The $m_{t\bar{t}}$ distributions in the eleven signal regions of the 1-lepton channel after the fit to the full 140 fb^{-1}dataset are shown in Figures 10 and 11. The $m_{\ell\ell bb}$ distributions in the five signal regions of the 2-lepton channel are shown in Figure 12. A comparison of the post- and pre-fit background predictions is included the respective ratio panels. The observed $m_{t\bar{t}}$ and $m_{\ell\ell bb}$ spectra are compatible with the post-fit expected spectra within the (constrained) uncertainty bands in all signal regions.

The agreement between the data and various signal hypotheses is quantified at the search stage. The significance of individual local excesses or deficits is quantified by fitting the observed data with different signal hypotheses ($S + I + B$ model), predicting either a single scalar or pseudo-scalar with masses in the range 400 – 1400 GeV and values of the relative width of 1%, 5%, 10%, 20%, 30% and 40%. No significant interference pattern is found in the data. The most significant deviation from the background-only prediction is obtained for the interference pattern of a pseudo-scalar with $m_A = 800$ GeV, $\Gamma_A/m_A = 10\%$, and a best-fit value of $\sqrt{\mu} = 4.0$, with a local significance of 2.3σ. In the absence of any significant interference patterns in the data compared to the background-only prediction, exclusion regions are derived for the benchmark scenarios described in Section 2.

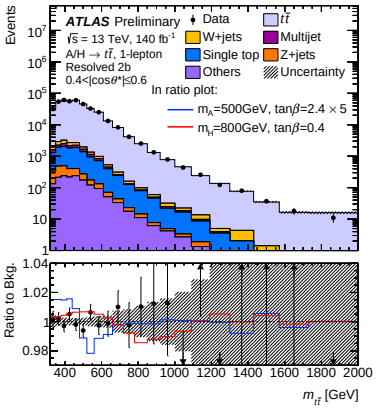
The excluded region at 95% CL in the $m_{A/H}$-$\tan\beta$ plane for the hypothesis of both a pseudo-scalar and a scalar with equal masses ($m_A = m_H$) in a type-II 2HDM with $\cos(\beta - \alpha) = 0$ is shown in Figure 13(a). The exclusion region for the hMSSM is shown in Figure 13(b). In both cases, the interference patterns for both A and H are assumed to be present in the spectra of the fitted variables. Values of $\tan\beta$ smaller than 3.49 (3.52) are observed (expected) to be excluded for $m_A = m_H = 400$ GeV in the 2HDM. In the hMSSM, values of $\tan\beta$ smaller than 3.16 (3.37) are observed (expected) to be excluded for $m_A = 400$ GeV. Masses up to 1240 GeV (1210 GeV) are observed (expected) to be excluded for the lowest tested $\tan\beta$ value of 0.4 in the 2HDM. In the hMSSM, masses up to 950 GeV (830 GeV) are observed (expected) to be excluded for $\tan\beta = 1.0$. In both scenarios, the observed exclusion is stronger than the expected



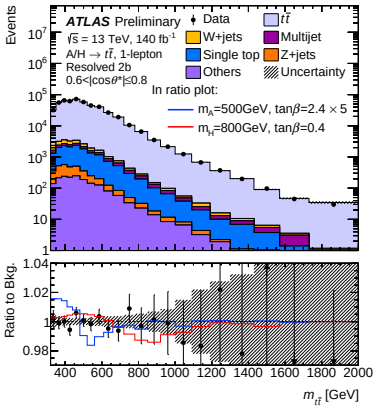
(a) Resolved, $2b$, $|\cos(\theta^*)| < 0.2$



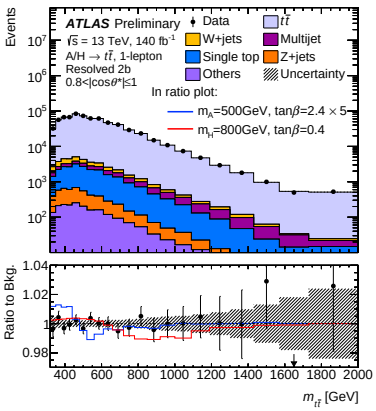
(b) Resolved, $2b$, $0.2 < |\cos(\theta^*)| < 0.4$



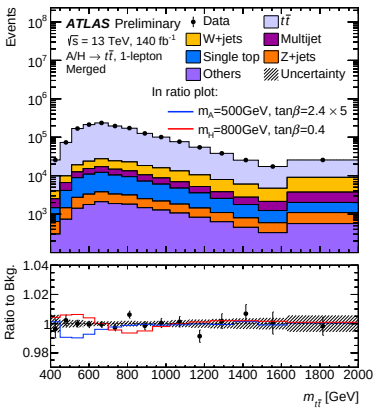
(c) Resolved, $2b$, $0.4 < |\cos(\theta^*)| < 0.6$



(d) Resolved, $2b$, $0.6 < |\cos(\theta^*)| < 0.8$

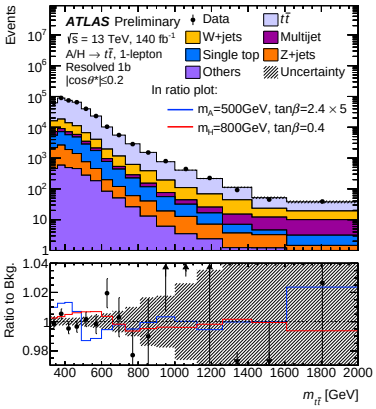


(e) Resolved, $2b$, $0.8 < |\cos(\theta^*)| < 1.0$

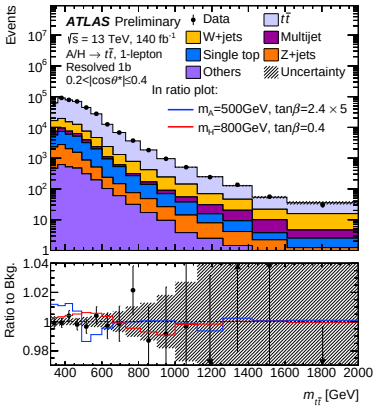


(f) Merged

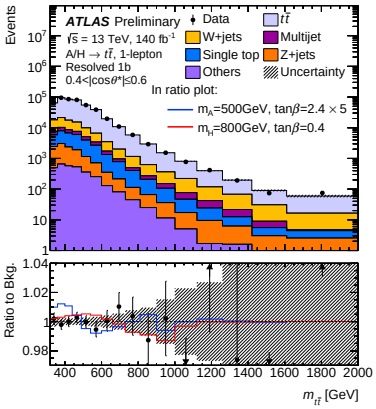
Figure 10: Post-fit distributions of the reconstructed $m_{t\bar{t}}$ in (a-e) the five Resolved $2b$ signal regions and (f) the Merged signal region of the 1-lepton channel. In the lower panel, the ratio of the data and the post-fit prediction is shown (data points). The expected relative deviation from the background prediction in the presence of an interference pattern, $(S + I + B)/B$, for two representative signal hypotheses is also shown in the ratio panel.



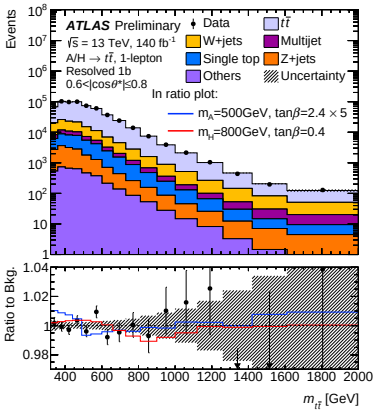
(a) Resolved, $1b$, $|\cos(\theta^*)| < 0.2$



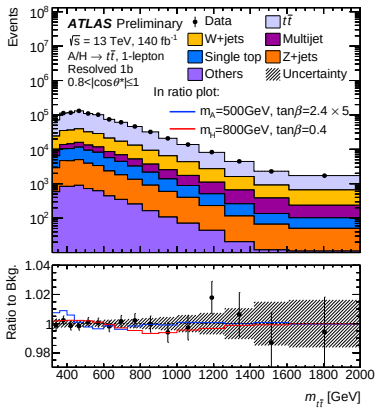
(b) Resolved, $1b$, $0.2 < |\cos(\theta^*)| < 0.4$



(c) Resolved, $1b$, $0.4 < |\cos(\theta^*)| < 0.6$

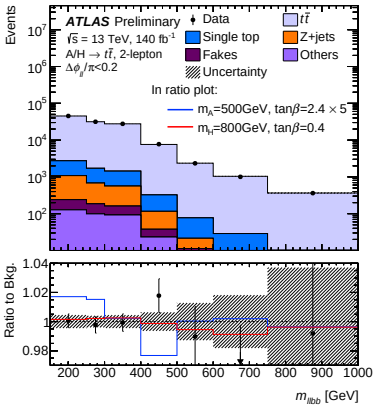


(d) Resolved, $1b$, $0.6 < |\cos(\theta^*)| < 0.8$

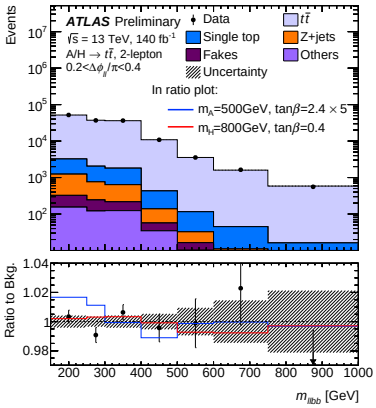


(e) Resolved, $1b$, $0.8 < |\cos(\theta^*)| < 1.0$

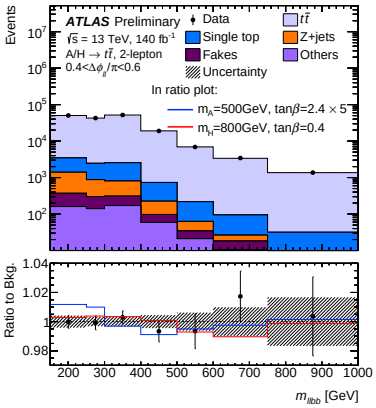
Figure 11: Post-fit distributions of the reconstructed $m_{t\bar{t}}$ in the five Resolved $1b$ signal regions of the 1-lepton channel. In the lower panel, the ratio of the data and the post-fit prediction is shown (data points). The expected relative deviation from the background prediction in the presence of an interference pattern, $(S + I + B)/B$, for two representative signal hypotheses is also shown in the ratio panel.



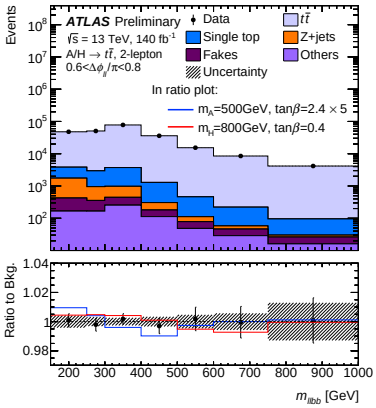
(a) $0.0 < \Delta\phi_{\ell\ell} < 0.2$



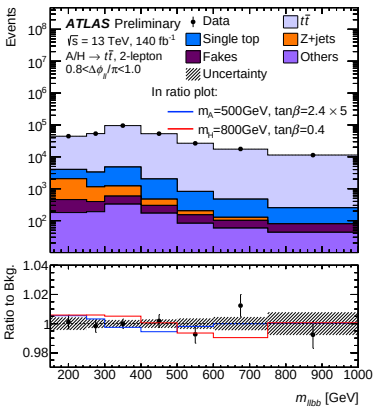
(b) $0.2 < \Delta\phi_{\ell\ell} < 0.4$



(c) $0.4 < \Delta\phi_{\ell\ell} < 0.6$



(d) $0.6 < \Delta\phi_{\ell\ell} < 0.8$



(e) $0.8 < \Delta\phi_{\ell\ell} < 1.0$

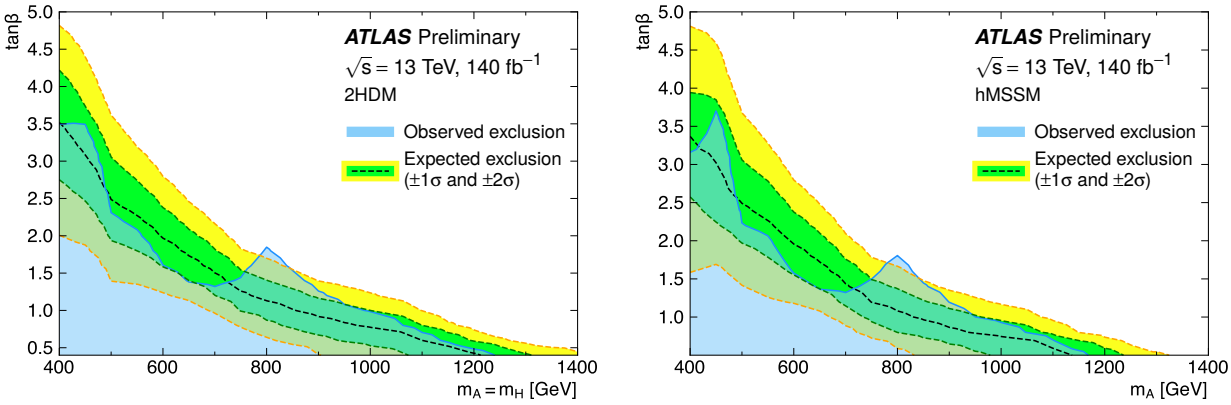
Figure 12: Post-fit distributions of the reconstructed $m_{\ell\ell bb}$ in the five signal regions of the 2-lepton channel. In the lower panel, the ratio of the data and the post-fit prediction is shown (data points). The expected relative deviation from the background prediction in the presence of an interference pattern, $(S + I + B)/B$, for two representative signal hypotheses is also shown in the ratio panel.

exclusion by about 2σ in the mass region $m_A = m_H \approx 850$ GeV. This deviation is consistent with the location of the largest local excess in the data, which is found for $m_{t\bar{t}} \approx 850$ GeV, as mentioned above. In the type-II 2HDM and the hMSSM, a local deficit compared to the SM expectation is predicted, hence the local excess in the data leads to a stronger-than-expected exclusion of these BSM scenarios in this mass region. The sensitivity of the search is dominated by the 1-lepton channel due to its larger branching ratio. The sensitivity improvement from the statistical combination with the 2-lepton channel is largest for low values of $m_{A/H}$. For $m_{A/H} = 400$ GeV, the observed (expected) exclusion range in $\tan\beta$ is 11% (5%) larger for the combined 1- and 2-lepton channels compared to the exclusion obtained with the 1-lepton channel alone.

Constraints are also derived for the 2HDM+a benchmark, specifically for benchmark scenarios 2a and 2b defined in Ref. [156]. In these benchmark scenarios, the value of the mixing angle is chosen such that $\sin\theta = 0.35$ and 0.7, respectively. The masses of the additional neutral and charged Higgs bosons of the 2HDM set to 600 GeV ($m_A = m_H = m_{H^\pm}$), while the mass of the DM particle is set to 1 GeV. Exclusion limits are derived as a function of the mass of the pseudo-scalar mediator, m_a and $\tan\beta$ (Figure 14). They only show a moderate dependence on m_a, which results from a decrease in the production cross-section of the mediator a with increasing m_a combined with an increase of the branching ratio for the decay $A \to t\bar{t}$. The latter is related to the fact that the branching ratio for the decay $A \to ah$ decreases with m_a. Values of $\tan\beta$ up to almost 1.1 (0.9) are excluded in the probed m_a range for the benchmark scenario with $\sin\theta = 0.35$ (0.70). The lower sensitivity in the case of the scenario with the larger $a - A$ mixing parameter $\sin\theta$ is due to the fact that the branching ratio for the decay $A \to t\bar{t}$ is smaller for larger $a - A$ mixing as the mixing increases the branching ratio for the invisible decay $A \to \chi\chi$.

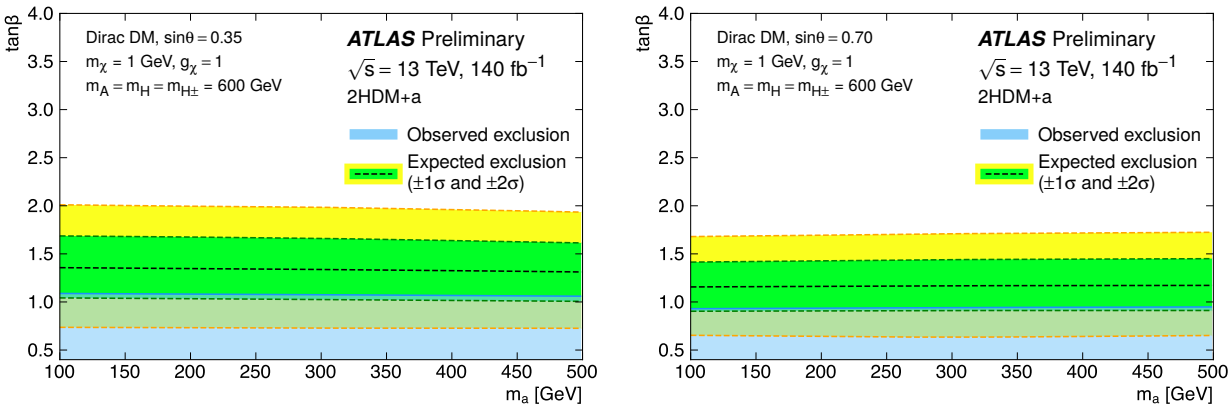
The constraints on the coupling modifiers for the pseudo-scalar A and the scalar H to $t\bar{t}$ separately are shown as a function of the (pseudo-)scalar mass in Figures 15 and 16, respectively. In these cases, only the interference pattern for either A or H are assumed to be present in the spectra of the fitted variables. The constraints are derived for different, fixed values of the relative total width $\Gamma_{A/H}/m_{A/H}$. Given that the partial width $\Gamma(A/H \to t\bar{t})$ is proportional to $g_{A/Ht\bar{t}}$, it can exceed the total width in some regions of the shown phase space. These unphysical regions are marked by hatched lines. The constraints on the coupling modifier $g_{A/Ht\bar{t}}$ for signal hypotheses with a single pseudo-scalar are more stringent than for those with a single scalar of the same width due to the fact that the (pure-signal) cross-sections are generally larger for the pseudo-scalar compared to the scalar case for the same masses and width. Further differences in the exclusion regions arise due to differences in the interference patterns of scalars and pseudo-scalars. In the former case, the peak is generally narrower and located closer to the dip. The observed constraints on the coupling strength are slightly weaker than expected for $m_{A/H} \approx 850$ GeV, especially in the scenarios with $\Gamma_{A/H}/m_{A/H} = 5\%$ and $\Gamma_{A/H}/m_{A/H} = 10\%$, consistent with the small narrow excess of events in data compared to the SM expectation observed in this region. This weaker-than-expected exclusion can be explained by the fact that for $g_{At\bar{t}} > 1$ the pure-signal component S, which scales like $g^4_{A/Ht\bar{t}}$ (Equation 1), is enhanced over the interference component, which scales like $g^2_{A/Ht\bar{t}}$. This means that larger values of $g_{A/Ht\bar{t}}$ correspond to interference patterns with a shallower dip or, for very large couplings, even a peak-peak instead of a peak-dip structure. Such patterns are more compatible with a narrow excess in the data than the interference patterns with a more pronounced dip obtained for smaller values of $g_{At\bar{t}}$. A small, discrete "island" occurs in the observed exclusion contour for the scenario with a single scalar with total width $\Gamma_A = 10\%$ in the region around $m_A = 800$ GeV, for which coupling values around $g_{Ht\bar{t}} \approx 2.0$ are observed to be excluded but slightly larger values of $g_{At\bar{t}}$ are not (Figure 16(b)). This island can be explained by the fact that CL_s value for this choice of m_H and Γ_H exhibits a local minimum around $g_{Ht\bar{t}} \approx 1.2$ when expressed as a function of $g_{Ht\bar{t}}$, similar to the scenario shown in Figure 9, where $\sqrt{\mu}$ can

be read as the equivalent of $g^2_{A/Ht\bar{t}}$ (compare also Equations 1 and 6). This local minimum causes the CL_s value for a narrow range of $g_{A/Ht\bar{t}}$ to fall below the exclusion threshold of 0.05. For the scenarios with $\Gamma_{A/H}/m_{A/H} = 1\%$, no constraints were obtained within the physical parameter region.



(a) (b)

Figure 13: Observed and expected exclusion contours in the $m_{A/H}$-tan β plane for (a) a type-II 2HDM in the alignment limit ($\cos(\beta - \alpha) = 0$) with mass-degenerate pseudo-scalar and scalar states, $m_A = m_H$ and (b) the hMSSM.



(a) (b)

Figure 14: Observed and expected exclusion contours in the m_a-tan β plane for the 2HDM+a with $m_A = m_H = 600$ GeV and (a) $\sin\theta = 0.35$ and (b) $\sin\theta = 0.70$. The model settings correspond to those of benchmark scenarios 2a and 2b defined in Ref. [156], respectively.

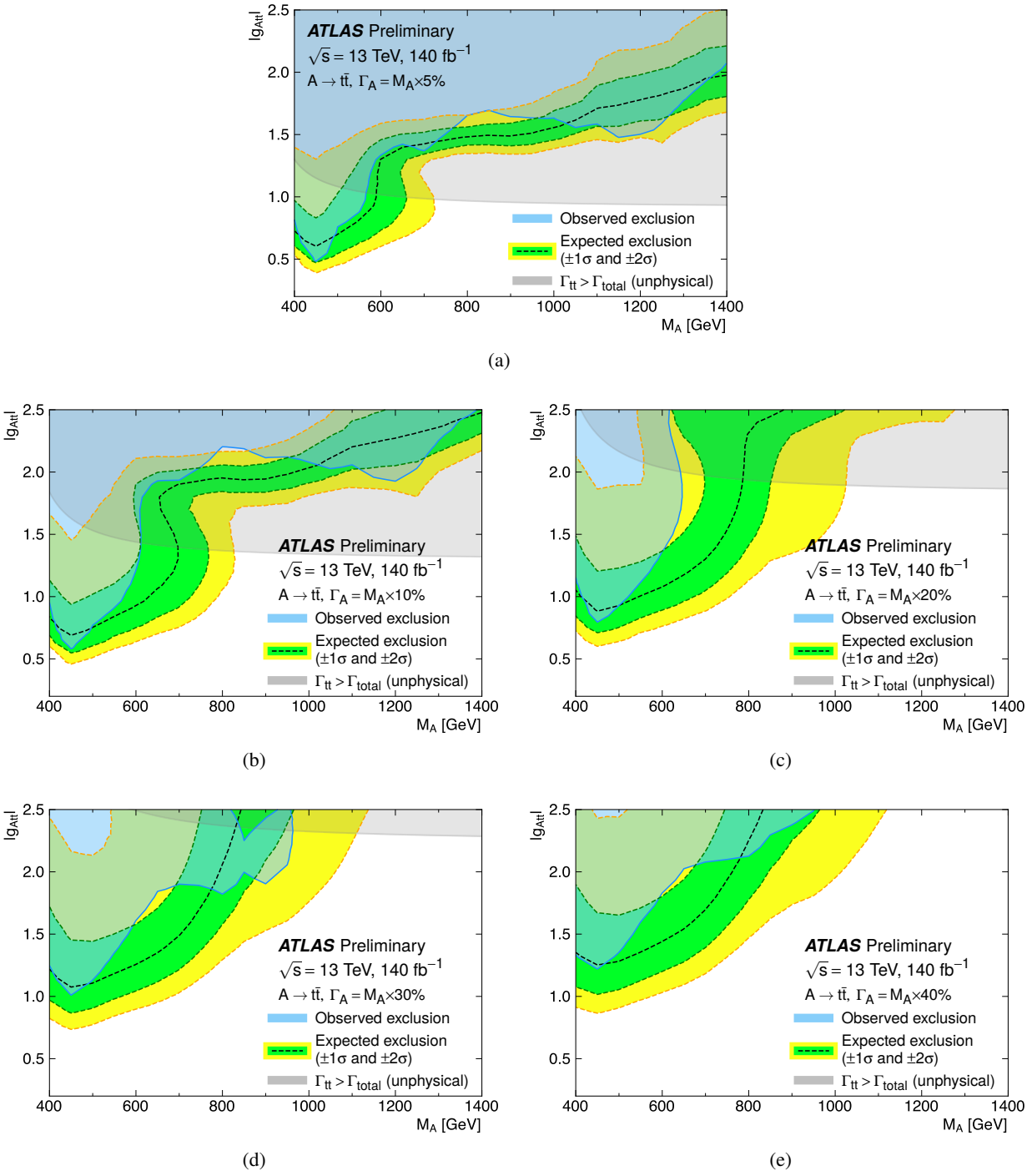
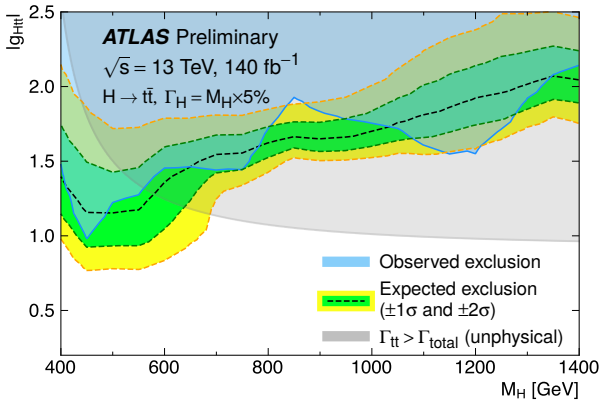
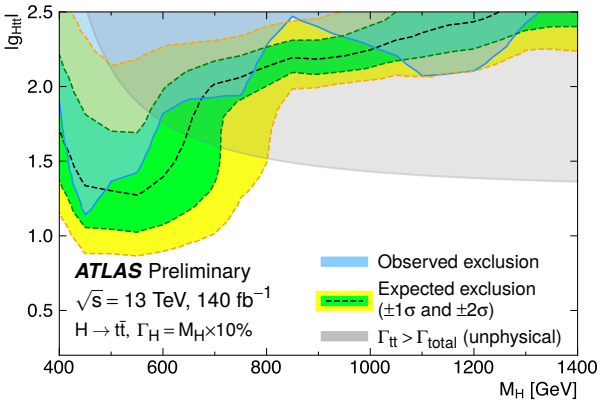


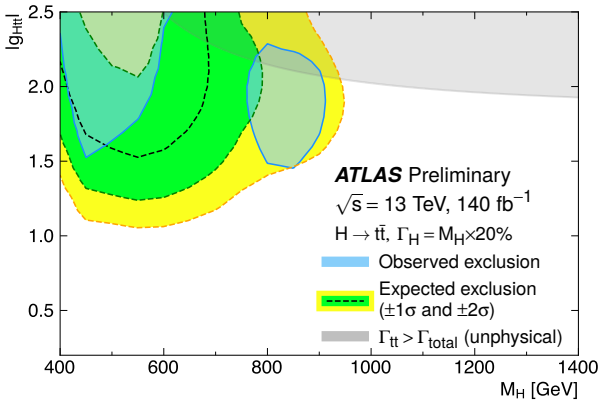
Figure 15: Constraints on the coupling strength modifier $g_{At\bar{t}}$ as a function of m_A for different values of the relative width of the pseudo-scalar A: (a) 5%, (b) 10%, (c) 20%, (d) 30%, (e) 40%. The observed exclusion regions are indicated by the shaded area. The boundary of the expected exclusion region under the background-only hypothesis is marked by the dashed line. The surrounding shaded bands correspond to the ±1 and ±2 standard deviation ($\pm 1\sigma, \pm 2\sigma$) uncertainty. The hatched area indicates the unphysical region of phase space where the partial width $\Gamma(A \to t\bar{t})$ is larger than the total width of A.



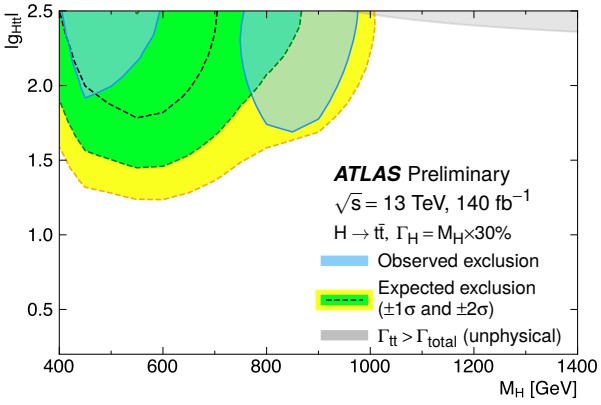
(a)



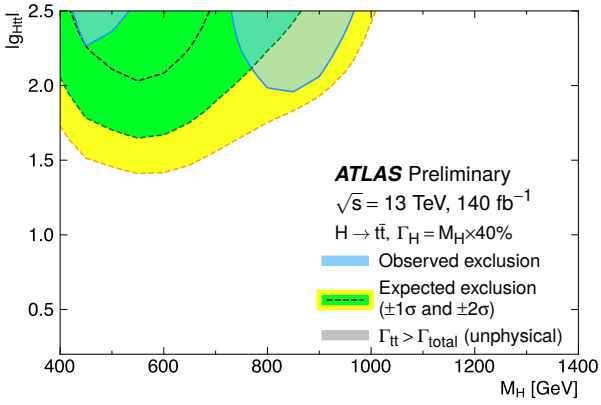
(b)



(c)



(d)



(e)

Figure 16: Constraints on the coupling strength modifier $g_{Ht\bar{t}}$ as a function of m_H for different values of the relative width of the scalar H: (a) 5%, (b) 10%, (c) 20%, (d) 30%, (e) 40%. The observed exclusion regions are indicated by the shaded area. The boundary of the expected exclusion region under the background-only hypothesis is marked by the dashed line. The surrounding shaded bands correspond to the ±1 and ±2 standard deviation ($\pm 1\sigma, \pm 2\sigma$) uncertainty. The hatched area indicates the unphysical region of phase space where the partial width $\Gamma(H \to t\bar{t})$ is larger than the total width of H.

12 Conclusion

A search for massive pseudo-scalar and scalar resonances decaying to $t\bar{t}$ has been conducted on 140 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV recorded by the ATLAS experiment. The non-negligible interference between the signal and the main background from SM $t\bar{t}$ production is taken into account. The search uses semi-leptonic and di-leptonic decays of the $t\bar{t}$ system. For the semi-leptonic decays, separate analysis strategies targeting resolved and merged hadronic top-quark decays are used. The agreement between the data and the SM prediction is quantified in distributions of the $t\bar{t}$ invariant mass (1-lepton final states) and the invariant mass of the two leptons and two b-jets (2-lepton final states) in several orthogonal signal regions, which are combined in a final likelihood fit.

No significant deviation from the SM prediction is observed, and exclusion regions at 95% CL are derived for several representative benchmark models that predict new scalar and pseudo-scalar states decaying to $t\bar{t}$, such as a generic type-II 2HDM, the hMSSM, and the 2HDM+a benchmark for LHC DM searches. Additionally, the search results are interpreted in a more model-independent scenario in which only the interference pattern of a single scalar or pseudo-scalar of a given mass and width is considered and upper limits are derived on the coupling of this particle to $t\bar{t}$. In the 2HDM, values of $\tan\beta$ smaller than 3.0 are excluded for $m_A = m_H = 400$ GeV, and mass values up to 1250 GeV are excluded for the lowest tested $\tan\beta$ value of 0.4. In the hMSSM, values of $\tan\beta$ smaller than 3.1 are excluded for $m_A = 400$ GeV and mass values up to 1000 GeV are excluded for $\tan\beta$ =1.0. The search presented in this paper provides the most stringent constraints on the 2HDM and hMSSM parameter space in the region of high m_A and low $\tan\beta$ to date, surpassing previous constraints from $t\bar{t}$ interference and $t\bar{t}t\bar{t}$ searches on Run-2 data. In the 2HDM+a, values of $\tan\beta$ below 1.2 (1.0) are excluded across the probed range of the mediator mass m_a for a benchmark scenario with low (high) $a - A$ mixing. The search presented in this paper is the first to consider the more complex interference patterns arising in the presence of two pseudo-scalars and a scalar and the first $t\bar{t}$ interference search to set constraints on pseudo-scalar mediators to dark matter.

References

[1] T. D. Lee, *A Theory of Spontaneous T Violation*, Phys. Rev. D **8** (4 1973) 1226, URL: https://link.aps.org/doi/10.1103/PhysRevD.8.1226 (cit. on p. 2).

[2] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Physics Reports **516** (2012) 1, ISSN: 0370-1573, URL: http://dx.doi.org/10.1016/j.physrep.2012.02.002 (cit. on p. 2).

[3] K. Inoue, A. Kakuto, H. Komatsu and S. Takeshita, *Aspects of Grand Unified Models with Softly Broken Supersymmetry*, Prog. Theor. Phys. **68** (1982) 927, [Erratum: Prog.Theor.Phys. 70, 330 (1983)] (cit. on p. 2).

[4] J. F. Gunion and H. E. Haber, *CP-conserving two-Higgs-doublet model: The approach to the decoupling limit*, Physical Review D **67** (2003), ISSN: 1089-4918, URL: http://dx.doi.org/10.1103/PhysRevD.67.075019 (cit. on p. 2).

[5] Y. A. Golfand and E. P. Likhtman, *Extension of the Algebra of Poincare Group Generators and Violation of p Invariance*, JETP Lett. **13** (1971) 323 (cit. on p. 2).

[6] D. V. Volkov and V. P. Akulov, *Is the Neutrino a Goldstone Particle?*, Phys. Lett. B **46** (1973) 109 (cit. on p. 2).

[7] J. Wess and B. Zumino, *Supergauge Transformations in Four-Dimensions*, Nucl. Phys. B **70** (1974) 39, ed. by A. Salam and E. Sezgin (cit. on p. 2).

[8] J. Wess and B. Zumino, *Supergauge Invariant Extension of Quantum Electrodynamics*, Nucl. Phys. B **78** (1974) 1 (cit. on p. 2).

[9] S. Ferrara and B. Zumino, *Supergauge Invariant Yang-Mills Theories*, Nucl. Phys. B **79** (1974) 413 (cit. on p. 2).

[10] A. Salam and J. A. Strathdee, *Supersymmetry and Nonabelian Gauges*, Phys. Lett. B **51** (1974) 353 (cit. on p. 2).

[11] J. E. Kim, *Light Pseudoscalars, Particle Physics and Cosmology*, Phys. Rept. **150** (1987) 1 (cit. on p. 2).

[12] K. Huitu, N. Koivunen, O. Lebedev, S. Mondal and T. Toma, *Probing pseudo-Goldstone dark matter at the LHC*, Phys. Rev. D **100** (2019) 015009, arXiv: 1812.05952 [hep-ph] (cit. on p. 2).

[13] M. Mühlleitner, M. O. P. Sampaio, R. Santos and J. Wittbrodt, *Phenomenological Comparison of Models with Extended Higgs Sectors*, JHEP **08** (2017) 132, arXiv: 1703.07750 [hep-ph] (cit. on p. 2).

[14] G. Moortgat-Pick et al., *Dark Matter phenomenology in 2HDMS*, SciPost Phys. Proc. **12** (2023) 072, arXiv: 2301.08314 [hep-ph] (cit. on p. 2).

[15] A. Dey, V. Keus, S. Moretti and C. Shepherd-Themistocleous, *A smoking gun signature of the 3HDM*, (2023), arXiv: 2310.06593 [hep-ph] (cit. on p. 2).

[16] M. R. Buckley, D. Feld and D. Goncalves, *Scalar Simplified Models for Dark Matter*, Phys. Rev. D **91** (2015) 015017, arXiv: 1410.6497 [hep-ph] (cit. on p. 2).

[17] M. Bauer, U. Haisch and F. Kahlhoefer, *Simplified dark matter models with two Higgs doublets: I. Pseudoscalar mediators*, JHEP **05** (2017) 138, arXiv: 1701.07427 [hep-ph] (cit. on p. 2).

[18] Y. Afik, E. Gozani and Y. Rozen, *Searching for dark matter with $t\bar{t}$ resonance*, Eur. Phys. J. C **79** (2019) 230, arXiv: 1812.06718 [hep-ex] (cit. on p. 2).

[19] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 2, 6).

[20] J. F. Gunion and H. E. Haber, *The CP conserving two Higgs doublet model: The Approach to the decoupling limit*, Phys. Rev. D **67** (2003) 075019, arXiv: hep-ph/0207010 (cit. on p. 2).

[21] 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] (cit. on p. 2).

[22] 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] (cit. on p. 2).

[23] 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] (cit. on p. 2).

[24] CMS Collaboration, *Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton–proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **09** (2018) 007, arXiv: 1803.06553 [hep-ex] (cit. on p. 2).

[25] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **800** (2020) 135103, arXiv: 1906.02025 [hep-ex] (cit. on p. 2).

[26] CMS Collaboration, *Search for a heavy Higgs boson decaying into two lighter Higgs bosons in the $\tau\tau bb$ final state at 13 TeV*, JHEP **11** (2021) 057, arXiv: 2106.10361 [hep-ex] (cit. on p. 2).

[27] ATLAS Collaboration, *Search for charged Higgs bosons decaying into a top quark and a bottom quark at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **06** (2021) 145, arXiv: 2102.10076 [hep-ex] (cit. on p. 2).

[28] CMS Collaboration, *Search for charged Higgs bosons decaying into a top and a bottom quark in the all-jet final state of pp collisions at $\sqrt{s} = 13$ TeV*, JHEP **07** (2020) 126, arXiv: 2001.07763 [hep-ex] (cit. on p. 2).

[29] ATLAS Collaboration, *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*, (2022), arXiv: 2211.01136 [hep-ex] (cit. on p. 2).

[30] 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] (cit. on pp. 2, 10, 25, 27).

[31] 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] (cit. on p. 2).

[32] M. Czakon et al., *Top-pair production at the LHC through NNLO QCD and NLO EW*, JHEP **10** (2017) 186, arXiv: 1705.04105 [hep-ph] (cit. on pp. 3, 7).

[33] ATLAS Collaboration, *Beyond-the-Standard-Model Higgs boson searches at a High-Luminosity LHC with ATLAS*, ATL-PHYS-PUB-2013-016, 2013, URL: https://cds.cern.ch/record/1611190 (cit. on p. 3).

[34] A. Djouadi et al., *The post-Higgs MSSM scenario: Habemus MSSM?*, Eur. Phys. J. C **73** (2013) 2650, arXiv: 1307.5205 [hep-ph] (cit. on p. 3).

[35] P. Fayet, *Supersymmetry and Weak, Electromagnetic and Strong Interactions*, Phys. Lett. B **64** (1976) 159 (cit. on p. 4).

[36] P. Fayet, *Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions*, Phys. Lett. B **69** (1977) 489 (cit. on p. 4).

[37] G. R. Farrar and P. Fayet, *Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry*, Phys. Lett. B **76** (1978) 575 (cit. on p. 4).

[38] P. Fayet, *Relations Between the Masses of the Superpartners of Leptons and Quarks, the Goldstino Couplings and the Neutral Currents*, Phys. Lett. B **84** (1979) 416 (cit. on p. 4).

[39] S. Dimopoulos and H. Georgi, *Softly Broken Supersymmetry and SU(5)*, Nucl. Phys. B **193** (1981) 150 (cit. on p. 4).

[40] A. Djouadi, L. Maiani, A. Polosa, J. Quevillon and V. Riquer, *Fully covering the MSSM Higgs sector at the LHC*, JHEP **06** (2015) 168, arXiv: 1502.05653 [hep-ph] (cit. on p. 4).

[41] T. Abe et al., *LHC Dark Matter Working Group: Next-generation spin-0 dark matter models*, Phys. Dark Univ. **27** (2020) 100351, arXiv: 1810.09420 [hep-ex] (cit. on p. 4).

[42] F. Esser, M. Madigan, V. Sanz and M. Ubiali, *On the coupling of axion-like particles to the top quark*, JHEP **09** (2023) 063, arXiv: 2303.17634 [hep-ph] (cit. on p. 4).

[43] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: https://cds.cern.ch/record/1291633 (cit. on p. 6), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: https://cds.cern.ch/record/1451888.

[44] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 6).

[45] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on p. 6).

[46] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187 (cit. on p. 6).

[47] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, (2022), arXiv: 2212.09379 [hep-ex] (cit. on p. 6).

[48] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017 (cit. on p. 6).

[49] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on p. 7).

[50] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 7).

[51] 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] (cit. on p. 7).

[52] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on p. 7).

[53] 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 (cit. on p. 7).

[54] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on p. 7).

[55] S. Frixione, G. Ridolfi and P. Nason,
A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction,
JHEP **09** (2007) 126, arXiv: 0707.3088 [hep-ph] (cit. on p. 7).

[56] 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] (cit. on p. 7).

[57] 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] (cit. on p. 7).

[58] J. M. Campbell, R. K. Ellis, P. Nason and E. Re,
Top-Pair Production and Decay at NLO Matched with Parton Showers, JHEP **04** (2015) 114,
arXiv: 1412.1828 [hep-ph] (cit. on p. 7).

[59] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040,
arXiv: 1410.8849 [hep-ph] (cit. on p. 7).

[60] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
ATL-PHYS-PUB-2016-020, 2016, URL: https://cds.cern.ch/record/2216168
(cit. on pp. 7, 25).

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

[62] D. J. Lange, *The EvtGen particle decay simulation package*,
Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on p. 7).

[63] M. Beneke, P. Falgari, S. Klein and C. Schwinn,
Hadronic top-quark pair production with NNLL threshold resummation,
Nucl. Phys. B **855** (2012) 695, arXiv: 1109.1536 [hep-ph] (cit. on p. 8).

[64] M. Cacciari, M. Czakon, M. Mangano, A. Mitov and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*,
Phys. Lett. B **710** (2012) 612, arXiv: 1111.5869 [hep-ph] (cit. on p. 8).

[65] P. Bärnreuther, M. Czakon and A. Mitov, *Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \to t\bar{t} + X$*,
Phys. Rev. Lett. **109** (2012) 132001, arXiv: 1204.5201 [hep-ph] (cit. on p. 8).

[66] M. Czakon and A. Mitov,
NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels,
JHEP **12** (2012) 054, arXiv: 1207.0236 [hep-ph] (cit. on p. 8).

[67] M. Czakon and A. Mitov,
NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction,
JHEP **01** (2013) 080, arXiv: 1210.6832 [hep-ph] (cit. on p. 8).

[68] M. Czakon, P. Fiedler and A. Mitov,
Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$,
Phys. Rev. Lett. **110** (2013) 252004, arXiv: 1303.6254 [hep-ph] (cit. on p. 8).

[69] M. Czakon and A. Mitov,
Top++: A program for the calculation of the top-pair cross-section at hadron colliders,
Comput. Phys. Commun. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph] (cit. on p. 8).

[70] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 8).

[71] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189, arXiv: 0901.0002 [hep-ph] (cit. on p. 8).

[72] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. C **64** (2009) 653, arXiv: 0905.3531 [hep-ph] (cit. on p. 8).

[73] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv: 1007.2241 [hep-ph] (cit. on p. 8).

[74] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. D **89** (2014) 033009, arXiv: 1302.6246 [hep-ph] (cit. on p. 8).

[75] 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] (cit. on p. 8).

[76] 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] (cit. on pp. 8, 25).

[77] 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] (cit. on p. 8), Erratum: JHEP **02** (2010) 011.

[78] 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] (cit. on p. 8).

[79] 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] (cit. on p. 8).

[80] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. D **82** (2010) 054018, arXiv: 1005.4451 [hep-ph] (cit. on p. 8).

[81] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028, arXiv: 1001.5034 [hep-ph] (cit. on p. 8).

[82] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011) 091503, arXiv: 1103.2792 [hep-ph] (cit. on p. 8).

[83] 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] (cit. on pp. 8, 9).

[84] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, **2/2017** (2016), arXiv: 1610.07922 [hep-ph] (cit. on p. 8).

[85] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph] (cit. on p. 8).

[86] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 9).

[87] F. Buccioni et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866, arXiv: 1907.13071 [hep-ph] (cit. on p. 9).

[88] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv: 1111.5206 [hep-ph] (cit. on p. 9).

[89] 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] (cit. on p. 9).

[90] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph] (cit. on p. 9).

[91] 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] (cit. on p. 9).

[92] 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] (cit. on p. 9).

[93] S. Catani, F. Krauss, B. R. Webber and R. Kuhn, *QCD Matrix Elements + Parton Showers*, JHEP **11** (2001) 063, arXiv: hep-ph/0109231 (cit. on p. 9).

[94] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, JHEP **05** (2009) 053, arXiv: 0903.1219 [hep-ph] (cit. on p. 9).

[95] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO*, Phys. Rev. Lett. **103** (2009) 082001, arXiv: 0903.2120 [hep-ph] (cit. on p. 9).

[96] ATLAS Collaboration, *Multi-Boson Simulation for* 13 *TeV ATLAS Analyses*, ATL-PHYS-PUB-2017-005, 2017, URL: https://cds.cern.ch/record/2261933 (cit. on p. 9).

[97] D. Buarque Franzosi and C. Zhang, *Bottom and Top loop structure in ggH and ggA*, URL: https://cp3.irmp.ucl.ac.be/projects/madgraph/wiki/Models/ggHFullLoop (cit. on p. 9).

[98] S. Moretti and D. A. Ross, *On the top-antitop invariant mass spectrum at the LHC from a Higgs boson signal perspective*, Phys. Lett. B **712** (2012) 245, arXiv: 1203.3746 [hep-ph] (cit. on p. 9).

[99] D. Buarque Franzosi, E. Vryonidou and C. Zhang, *Scalar production and decay to top quarks including interference effects at NLO in QCD in an EFT approach*, JHEP **10** (2017) 096, arXiv: 1707.06760 [hep-ph] (cit. on p. 9).

[100] D. Eriksson, J. Rathsman and O. Stal, *2HDMC: Two-Higgs-Doublet Model Calculator Physics and Manual*, Comput. Phys. Commun. **181** (2010) 189, arXiv: 0902.0851 [hep-ph] (cit. on p. 9).

[101] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) 1, arXiv: 1106.0034 [hep-ph] (cit. on p. 10).

[102] R. V. Harlander, S. Liebler and H. Mantler, *SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM*, Comput. Phys. Commun. **184** (2013) 1605, arXiv: 1212.3249 [hep-ph] (cit. on p. 10).

[103] R. V. Harlander and W. B. Kilgore, *Next-to-next-to-leading order Higgs production at hadron colliders*, Phys. Rev. Lett. **88** (2002) 201801, arXiv: hep-ph/0201206 (cit. on p. 10).

[104] R. V. Harlander and W. B. Kilgore, *Higgs boson production in bottom quark fusion at next-to-next-to leading order*, Phys. Rev. D **68** (2003) 013001, arXiv: hep-ph/0304035 (cit. on p. 10).

[105] U. Aglietti, R. Bonciani, G. Degrassi and A. Vicini, *Two-loop light fermion contribution to Higgs production and decays*, Phys. Lett. B **595** (2004) 432, arXiv: hep-ph/0404071 (cit. on p. 10).

[106] R. Bonciani, G. Degrassi and A. Vicini, *On the Generalized Harmonic Polylogarithms of One Complex Variable*, Comput. Phys. Commun. **182** (2011) 1253, arXiv: 1007.1891 [hep-ph] (cit. on p. 10).

[107] R. Harlander and P. Kant, *Higgs production and decay: Analytic results at next-to-leading order QCD*, JHEP **12** (2005) 015, arXiv: hep-ph/0509189 (cit. on p. 10).

[108] B. Hespel, F. Maltoni and E. Vryonidou, *Signal background interference effects in heavy scalar production and decay to a top-anti-top pair*, JHEP **10** (2016) 016, arXiv: 1606.04149 [hep-ph] (cit. on p. 10).

[109] P. Laycock et al., *ATLAS data preparation in run 2*, Journal of Physics: Conference Series **898** (2017) 042050, URL: https://dx.doi.org/10.1088/1742-6596/898/4/042050 (cit. on p. 10).

[110] 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 (cit. on p. 11).

[111] 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 (cit. on p. 11).

[112] 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] (cit. on p. 11).

[113] 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] (cit. on pp. 11, 20, 27).

[114] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 11).

[115] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on p. 11).

[116] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 11).

[117] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, arXiv: 1703.09665 [hep-ex] (cit. on p. 11).

[118] 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] (cit. on pp. 11, 27).

[119] ATLAS Collaboration, *Jet reclustering and close-by effects in ATLAS Run 2*, ATLAS-CONF-2017-062, 2017, URL: https://cds.cern.ch/record/2275649 (cit. on p. 11).

[120] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) 084, arXiv: 0912.1342 [hep-ph] (cit. on p. 11).

[121] ATLAS Collaboration, *Boosted Object Tagging with Variable-R Jets in the ATLAS Detector*, ATL-PHYS-PUB-2016-013, 2016, URL: https://cds.cern.ch/record/2199360 (cit. on p. 11).

[122] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, Eur. Phys. J. C **83** (2023) 681, arXiv: 2211.16345 [physics.data-an] (cit. on p. 11).

[123] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 11).

[124] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex] (cit. on pp. 11, 12, 22).

[125] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) 005, arXiv: 0802.1188 [hep-ph] (cit. on p. 12).

[126] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, Eur. Phys. J. C **79** (2019) 205, arXiv: 1810.05087 [hep-ex] (cit. on p. 12).

[127] 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] (cit. on p. 12).

[128] ATLAS Collaboration, *E_T^{miss} performance in the ATLAS detector using 2015–2016 LHC pp collisions*, ATLAS-CONF-2018-023, 2018, URL: https://cds.cern.ch/record/2625233 (cit. on pp. 12, 27).

[129] R. L. Workman et al., *Review of Particle Physics*, PTEP **2022** (2022) 083C01 (cit. on p. 13).

[130] ATLAS Collaboration, *A search for $t\bar{t}$ resonances using lepton-plus-jets events in proton–proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, JHEP **08** (2015) 148, arXiv: 1505.07018 [hep-ex] (cit. on pp. 19, 22, 25).

[131] ATLAS Collaboration, *Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 565, arXiv: 1804.10823 [hep-ex] (cit. on pp. 19, 21, 22, 25).

[132] ATLAS Collaboration, *Measurement of the cross-section and charge asymmetry of W bosons produced in proton–proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, Eur. Phys. J. C **79** (2019) 760, arXiv: 1904.05631 [hep-ex] (cit. on p. 21).

[133] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Lett. B **759** (2016) 601, arXiv: 1603.09222 [hep-ex] (cit. on p. 21).

[134] ATLAS Collaboration, *Measurements of the production cross-section for a Z boson in association with b-jets in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **07** (2020) 044, arXiv: 2003.11960 [hep-ex] (cit. on p. 23).

[135] A. Manohar, P. Nason, G. P. Salam and G. Zanderighi, *How bright is the proton? A precise determination of the photon parton distribution function*, Phys. Rev. Lett. **117** (2016) 242002, arXiv: 1607.04266 [hep-ph] (cit. on p. 24).

[136] M. Bähr et al., *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639, arXiv: 0803.0883 [hep-ph] (cit. on p. 25).

[137] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv: 1512.01178 [hep-ph] (cit. on p. 25).

[138] 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] (cit. on p. 25).

[139] ATLAS Collaboration, *Studies on the improvement of the matching uncertainty definition in top-quark processes simulated with Powheg+Pythia8*, ATL-PHYS-PUB-2023-029, 2023, URL: https://cds.cern.ch/record/2872787 (cit. on p. 25).

[140] S. Höche, S. Mrenna, S. Payne, C. T. Preuss and P. Skands, *A Study of QCD Radiation in VBF Higgs Production with Vincia and Pythia*, SciPost Phys. **12** (2022) 010, URL: https://scipost.org/10.21468/SciPostPhys.12.1.010 (cit. on p. 25).

[141] ATLAS, CDF, CMS, D0 Collaborations, *First combination of Tevatron and LHC measurements of the top-quark mass*, (2014), arXiv: 1403.4427 [hep-ex] (cit. on p. 25).

[142] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, Eur. Phys. J. C **53** (2008) 473, arXiv: 0706.2569 [hep-ph] (cit. on p. 26).

[143] ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. D **97** (2018) 072016, arXiv: 1712.08895 [hep-ex] (cit. on p. 26).

[144] ATLAS Collaboration, *Dependence of the Jet Energy Scale on the Particle Content of Hadronic Jets in the ATLAS Detector Simulation*, ATL-PHYS-PUB-2022-021, 2022, URL: https://cds.cern.ch/record/2808016 (cit. on p. 27).

[145] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector*, Eur. Phys. J. C **79** (2019) 135, arXiv: 1807.09477 [hep-ex] (cit. on p. 27).

[146] 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] (cit. on p. 27).

[147] ATLAS Collaboration, *Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s}$ = 13 TeV collected with the ATLAS detector*, Eur. Phys. J. C **82** (2021) 95, arXiv: 2109.10627 [hep-ex] (cit. on p. 27).

[148] ATLAS Collaboration, *Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139 fb^{-1} of ATLAS proton-proton collision data at $\sqrt{s}$ = 13 TeV*, (2023), arXiv: 2301.06319 [hep-ex] (cit. on p. 27).

[149] 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] (cit. on p. 27).

[150] 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] (cit. on p. 27).

[151] 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] (cit. on pp. 28, 29), Erratum: Eur. Phys. J. C **73** (2013) 2501.

[152] K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep., New York U., 2012, URL: https://cds.cern.ch/record/1456844 (cit. on p. 28).

[153] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693 (cit. on p. 29).

[154] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Annals Math. Statist. **9** (1938) 60 (cit. on p. 29).

[155] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society **54** (1943) 426 (cit. on p. 29).

[156] ATLAS Collaboration, *Combination and summary of ATLAS dark matter searches interpreted in a 2HDM with a pseudo-scalar mediator using 139 fb^{-1} of $\sqrt{s}$ = 13 TeV pp collision data*, (2023), arXiv: 2306.00641 [hep-ex] (cit. on pp. 35, 36).