Observation of quantum entanglement in top-quark pair production using $p\,p$ collisions of $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

We report the highest-energy observation of entanglement so far in top–antitop quark events produced at the Large Hadron Collider, using a proton–proton collision data set with a centre-of-mass energy of $\sqrt{s} = 13$ TeV and an integrated luminosity of 140 fb^{-1}. Spin entanglement is detected from the measurement of a single observable D, inferred by the angle between the charged leptons in their parent top- and antitop-quark rest frames. The observable is measured on a narrow interval around the top-quark–antitop-quark production threshold, where the entanglement detection is expected to be significant. The entanglement observable is measured in a fiducial phase-space with stable particles. The entanglement witness is measured to be $D = -0.547 \pm 0.002$ (stat.) ± 0.021 (syst.) for $340 < m_{t\bar{t}} < 380$ GeV. The large spread in predictions from several mainstream event generators indicates that modelling this property is challenging. The predictions depend in particular on the parton-shower algorithm used. The observed result is more than five standard deviations from a scenario without entanglement and hence constitutes the first observation of entanglement in a pair of quarks, and the observation of entanglement at the highest energy to date.

Contents

1 Introduction

Entanglement is a striking feature of quantum mechanics [1–3], with applications in fields such as metrology, cryptography, quantum information, and quantum computation [4–6]. If two particles are entangled, the state of one particle cannot be described independently from the other. More precisely, an entangled state is one that cannot be written as a convex combination of product states of density matrices [7, 8]. It has been observed in a wide variety of systems, ranging from the microscopic [9–13] to the macroscopic [14–16] scale. However, entanglement remains largely unexplored above low energy scales. Particle colliders, such as the Large Hadron Collider at CERN, probe fundamental particles and their interactions at the highest energies accessible in the laboratory, exceeded only by astrophysical sources. Recently, the heaviest fundamental particle known to exist, the top quark, was proposed as a new laboratory to study quantum entanglement and quantum information [17–24].

The simplest example of entanglement involves a pair of qubits, such as the spins of two spin-1/2 fermions. These fermions, specifically quarks, charged leptons, and neutrinos, are fundamental particles within the Standard Model (SM) of particle physics. Among these fermions, the top quark stands out as a promising

candidate for high-energy spin measurements due to its unique properties: its lifetime ($\sim 10^{-25}$s) is shorter than the timescale for hadronisation ($\sim 10^{-23}$s) and spin decorrelation ($\sim 10^{-21}$s) [25]. Because the top quark decays before it can hadronise, its spin information is transferred to its decay products. This feature is unique to top quarks and provides the opportunity to study a pseudo-bare quark free of the confinement effects inherent to other quarks.

Quarks are most commonly produced by hadron collider experiments as matter–antimatter pairs. A pair of top–antitop quarks ($t\bar{t}$) is a two-qubit system whose spin quantum state is described by the spin density matrix ρ:

$$\rho = \frac{I_4 + \sum_i \left(B_i^+ \sigma^i \otimes I_2 + B_i^- I_2 \otimes \sigma^i \right) + \sum_{i,j} C_{ij} \sigma^i \otimes \sigma^j}{4} \tag{1}$$

where I_n is the $n \times n$ identity matrix, σ^i are the corresponding Pauli matrices, and $i, j = 1, 2, 3$. $B_i^{\pm}$ characterise the spin polarization of each particle and C_{ij} encodes the spin correlation between them. At hadron colliders, $t\bar{t}$ pairs are produced dominantly via the strong interaction and thus have no intrinsic polarisation (i.e. $B_i^{\pm} = 0$), because of parity conservation in quantum chromodynamics [26]. However, their spins are expected to be correlated and this correlation was already observed by both the ATLAS and CMS experiments at the Large Hadron Collider (LHC) [27–31]. Entanglement in top-quark pairs can be observed via their spin correlation.

Due to their short lifetime, top quarks cannot be measured directly in experiments. In the SM they decay almost exclusively to a bottom quark and a W boson, and the W boson subsequently decays to either a pair of lighter quarks or to a charged lepton and a neutrino. In this measurement, only W bosons decaying into leptons are considered since charged leptons, in particular electrons and muons, are readily detected with high precision at collider experiments. Under the excellent assumption in which the spin-analysing power of the charged leptons is unity [32], the directions of these leptons are correlated with their parent top-quark and antitop-quark spins in such a way that the normalised differential cross-section of the process may be written as [33]:

$$\frac{1}{\sigma} \frac{d\sigma}{d\Omega_+ d\Omega_-} = \frac{1 + \mathbf{B}^+ \cdot \hat{\mathbf{q}}_+ - \mathbf{B}^- \cdot \hat{\mathbf{q}}_- - \hat{\mathbf{q}}_+ \cdot \mathbf{C} \cdot \hat{\mathbf{q}}_-}{(4\pi)^2} , \tag{2}$$

where $\hat{\mathbf{q}}_{\pm}$ are the antilepton (lepton) directions in each one of the parent top quark (antitop quark) rest frames and $\Omega_{\pm}$ is the solid angle associated to the antilepton (lepton). The vectors $\mathbf{B}^{\pm}$ determine the top quark (antitop quark) polarisations, while the matrix $\mathbf{C}$ contains their spin correlations. The information of the polarisations and spin correlations of the short-lived top quarks is transferred to the leptons and their values can be extracted from a measurement of angular observables of the decay leptons. With the help of these spin observables, the quantum tomography of the $t\bar{t}$ quantum state can be performed [17], allowing to study two-qubit states at high-energy colliders.

Despite the genuine quantum nature of the SM, the observation of entanglement is challenging, owing to the decoherence induced by the momentum measurement process and the lack of control of internal degrees of freedom in the initial state [20]. Indeed, entanglement can be only detected with the help of a dedicated analysis in a fiducial phase-space, and previous inclusive measurements showed no sign of entanglement [31].

At the LHC, $t\bar{t}$ pairs are produced mainly via gluon-gluon fusion. When they are produced close to their production threshold, that is, when their invariant mass ($m(t\bar{t})$) is close to twice the mass of the top quark ($m(t\bar{t}) \sim 2 \cdot m_t \sim 350$ GeV), they exist nearly in a spin singlet state [34–36], which is therefore expected to be maximally engangled. After averaging over all possible top-quark directions, entanglement

3

only survives at threshold because there the $t\bar{t}$ pairs are produced in a spin singlet, which is rotationally invariant. The rotational invariance of a spin singlet implies that the trace of the correlation matrix $\mathbf{C}$ is a good entanglement witness. It is an observable that can signal the presence of entanglement without any assumption on the particular form of the quantum state, with $\mathrm{tr}[\mathbf{C}] + 1 < 0$ as a sufficient condition for entanglement [17].

It is more convenient to define an entanglement witness using $D = \mathrm{tr}[\mathbf{C}]/3$, which can be experimentally measured as:

$$D = -3 \cdot \langle \cos \varphi \rangle , \tag{3}$$

where $\langle \cos \varphi \rangle$ is the average of the angle between the charged lepton directions in each one of the parent top quark and antitop quark rest frames which can be experimentally measured from an ensemble dataset. The existence of an entangled state is demonstrated if the measurement satisfies $D < -1/3$.

2 The ATLAS detector and event samples

ATLAS is a multipurpose particle detector [37–39] with a forward–backward symmetric cylindrical geometry and a coverage in a solid angle of almost 4π.[1] It consists of an inner-tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadronic calorimeters, and a muon spectrometer. The muon spectrometer surrounds the calorimeters and is based on three large superconducting air-core toroidal magnets with eight coils each. An extensive software suite [40] is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment. The complete dataset of proton–proton (pp) collision events at a centre-of-mass energy of $\sqrt{s} = 13$ TeV collected by the ATLAS experiment during 2015–2018 is used, corresponding to an integrated luminosity of 140 fb^{-1}. The analysis focuses on the data sample recorded using single-electron or single-muon triggers [41].

Reconstructed (detector level) objects, measured in the detector, are defined as follows. Electron candidates are required to pass the "tight" likelihood-based identification requirement as well as calorimeter- and track-based isolation criteria [42], and have pseudorapidity $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$. Muon candidates are required to pass the "medium" identification requirement as well as track-based isolation criteria [43–45], and have $|\eta| < 2.5$. Electrons and muons must have a minimum transverse momentum (p_{T}) of 25–28 GeV, depending on the data-taking period. Showers of particles that arise from the fragmentation of quark- or gluon-initiated particle (jets) [46] are reconstructed from particle-flow objects [47] using the anti-k_t algorithm [48, 49], with a radius parameter $R = 0.4$, a minimum p_{T} threshold of 25 GeV and $|\eta| < 2.5$. A Jet-Vertex-Tagger requirement is applied to jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$ to suppress jets originating from additional interactions in the same or neighboring bunch crossings (pile-up) [50]. Jets are tagged as containing B-hadrons using the DL1r tagger [51] with a b-tagging efficiency of 85%. Missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) [52, 53] is determined from the imbalance in the transverse momenta of all reconstructed objects.

---

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the center of the detector and the z-axis along the beam pipe. The x-axis points from the interaction point to the center of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, where ϕ is the azimuth angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

Particle-level objects are reconstructed using stable particles before detector reconstruction, but after hadronisation. A particle is defined as stable if it has a mean lifetime greater than 30 ps, within the pseudorapidity acceptance of the detector. Objects are selected to correspond to the reconstructed objects used in the detector-level selection. Electrons, muons and neutrinos are required to come from the electroweak decay of a top quark, and are discarded if they arise from the decay of hadrons or a τ lepton. Electrons and muons are then 'dressed' by summing their four-momenta with any prompt photons within $\Delta R = 0.1$, and must have $\Delta R > 0.4$ from a jet. Leptons are also required to have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$ and at least one lepton must have $p_{\mathrm{T}} > 25$ GeV. Jets are built by clustering all stable particles, using the anti-k_t algorithm with a radius parameter of $R = 0.4$, and are tagged as containing B-hadrons if they have at least one ghost-matched B-hadron [54, 55] with $p_{\mathrm{T}} > 5$ GeV. Jets are also required to have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. W bosons are reconstructed by combining the available electron and electron neutrino and muon and muon neutrino. Top quarks are reconstructed by pairing the two leading b-tagged jets, or the b-tagged jet and the highest-p_{T} untagged jet in events with only one b-tag, with the reconstructed W bosons. Both potential jet–lepton combinations are formed and the one which minimizes $|m_t - M(W_1 + b_{1/2})| + |m_t - M(W_2 + b_{2/1})|$ is taken as the correct pairing, where m_t refers to the mass of the top quark, $b_{1/2}$ refers to the two jets selected for the reconstruction, $W_{1/2}$ refers to the reconstructed W bosons and M is the invariant mass of the objects in brackets.

Monte Carlo (MC) event simulations are used to model the expected SM background and the $t\bar{t}$ signal. The production of $t\bar{t}$ events was modelled using the Powheg Box v2 heavy quark (hvq) [56–59] generator at next-to-leading-order (NLO) with the NNPDF3.0nlo [60] parton distribution function (PDF) set and the h_{damp} parameter[2] set to $1.5\,m_t$ [61]. As an alternative, an additional event sample was generated with Powheg Box Res [62, 63], developed to treat decaying resonances within the Powheg Box framework, including also off-shell and non-resonant effects in the matrix element calculation, labelled as $bb4\ell$ in the following. In this event sample, spin correlations are calculated at NLO, and a full NLO accuracy in $t\bar{t}$ production and decays is applied. To model the parton shower, hadronisation, and underlying event, the events of both Powheg Box v2 and Powheg Box Res were interfaced to Pythia 8.230 [64], with parameters set according to the A14 set of tuned parameters [65] and using the NNPDF2.3lo set of PDFs [66]. Similarly, the events of Powheg Box v2 (hvq) were interfaced also to Herwig 7.21 [67, 68], using the Herwig 7.21 default set of tuned parameters. The decays of bottom and charm hadrons were performed by EvtGen 1.6.0 [69]. An important difference between Pythia 8.230 and Herwig 7.21 is that the former uses a p_{T}-ordered shower, while the latter uses an angular-ordered shower (see Appendix D). To summarise, three different models of $t\bar{t}$ production are used for comparison to data: Powheg Box interfaced with Pythia, Powheg Box Res interfaced with Pythia ($bb4\ell$) and Powheg Box interfaced with Herwig.

The SM background processes which contribute to the analysis are the production of a single top quark with a W boson (tW), $t\bar{t} + X$ ($X = H, W, Z$) production, and the production of dileptonic events from a single or two massive gauge bosons (W, Z bosons). The generators for the hard-scatter process, showering and PDFs are listed in Appendix A. All simulated samples are reweighed to reproduce the observed distribution of the average number of collisions per bunch crossing. The detector level objects are the same for data and MC.

---

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

3 Analysis procedure

Only events taken during stable beam conditions, and for which all relevant components of the detector were operational, are considered. A selection of exactly one electron and one muon with opposite-sign electric charge is applied, where at least one of the leptons is matched geometrically to the object that fired the trigger. A minimum of two jets is required, where at least one of them is identified as a b-tagged jet.

The background contribution of events with reconstructed objects that have been misidentified as leptons, referred to as the "fake lepton" background, is estimated using a combination of MC prediction and a data-driven scale factor. This data-driven scale factor is obtained from a control region dominated by fake leptons. It is defined by using the same selection criteria as above, but requiring that the two leptons have the same sign in the electric charge.

In order to measure D, the top quarks must be reconstructed from their measured decay products. In the $t\bar{t}$ dileptonic decay, in addition to charged leptons and jets, there are two neutrinos which are not measured by the detector. Several methods are available to reconstruct the top quarks from the detector level charged leptons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$. The main method used in this work is the Ellipse method [70], which is a geometric approach to analytically calculate the neutrino momenta. Approximately 85% of events are successfully reconstructed by this method. If this method fails, the Neutrino Weighting method [71], which assigns a weight to each possible solution by the compatibility between the neutrino momenta and the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event, after scanning possible values of the pseudo-rapidities of the neutrinos, is used. If both methods fail, a simple pairing of each lepton with its closest b-tagged jet is used. If only one b-tagged jet is present in the event, the leading p_{T} untagged jet is used instead. In all cases, the W boson mass and the top quark mass are used as input parameters [25].

Events that pass the event selection described above are split into three analysis regions, based on $m(t\bar{t})$. The signal region is constructed to be dominated by events close to the production threshold, as this is the dominant region where the top quarks are expected to be entangled. The optimal window for the signal region was determined to be $340 < m(t\bar{t}) < 380$ GeV. Two additional validation regions are defined, in order to verify the method used for the measurement. Firstly, a region is defined close to the limit where entanglement is not expected, and also with significant dilution from mis-reconstructed events from non-entangled regions, requiring $380 < m(t\bar{t}) < 500$ GeV. Secondly, a region in which no signal of entanglement is expected is defined with $m(t\bar{t}) > 500$ GeV. The purity of $t\bar{t}$ events in each one of the regions is about 90%. The distribution of $\cos\varphi$ in the signal region and the reconstructed level D after background subtraction are shown on the left and right panels of Figure 1, respectively. The equivalent distributions in the validation regions are presented in Appendix E.

The detector response and event selection distort the shape of the $\cos\varphi$ distribution. The observed distribution is corrected for these effects with a simple method: a calibration curve connects any value at the detector level to the corresponding particle level value. The data are corrected for detector effects to the fiducial particle level using a calibration curve in the signal and in the validation regions, after the expected background has been subtracted. For this purpose, MC event samples with alternative values of D are obtained by reweighting the events, following the procedure described in Appendix B. The calibration curve corrects the value measured at the detector level D_{reco} to a corresponding value at particle level D_{particle}. To construct the calibration curve, several hypotheses for different values if D, denoted as D'_{particle} with a corresponding D'_{reco}, are created corresponding to changes in the expected value of the entanglement observable. Each pair of D'_{reco} and D'_{particle} values are plotted and then a linear fit is used to interpolate

between the points. Once the curve has been constructed, any value for D_{reco} measured in data can be calibrated to its corresponding D_{particle} value using this curve.
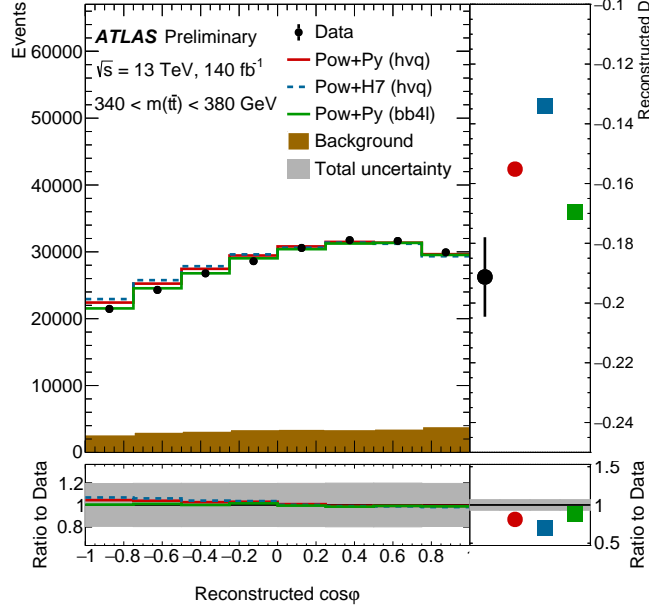

Figure 1: In the left panel, the $\cos\varphi$ observable in the signal region at reconstructed level and in the right, the entanglement witness D at the reconstructed detector-level distribution from different three different MC generators; the POWHEG BOX + PYTHIA and and POWHEG BOX + HERWIG heavy quark models, labelled "Pow+Py (hvq)" and "Pow + H7 (hvq)", respectively, and the POWHEG BOX + PYTHIA bb4ℓ models, labelled "Pow+Py (bb4ℓ)", after backgrounds are subtracted. The uncertainty band includes all sources of uncertainty added in quadrature. The ratio of the predictions with respect to data is shown at the bottom of the figure. The quoted value for D for the bb4ℓ model also includes a subtraction of the single-top-quark background.

4 Uncertainties

Systematic uncertainties include three categories: modelling uncertainties on the $t\bar{t}$ production and decay, modelling uncertainties on the backgrounds, and detector-related uncertainties for both the $t\bar{t}$ signal and the SM backgrounds. Each source of systematic uncertainty can result in a different calibration curve, as it changes the shape of $\cos\varphi$ either at particle level and / or at detector level. For all of the detector-related uncertainties, the particle level quantity is not affected and only detector level values change. For signal modelling uncertainties, the effects at particle level propagate to detector level, resulting in shifts in both. A summary of the different sources of systematic uncertainties and their impact on the result are found in Table 1. The size of the systematic uncertainties also depend upon the value of D. Those shown in Table 1 are for the expected SM value. All of the uncertainties above are not assigned to multijet events, which are determined from data. The systematic uncertainties considered in the analysis are explained in detail in Appendix C.

Systematic source	$\Delta D_{\text{observed}}(D = -0.547)$	ΔD (%)	$\Delta D_{\text{expected}}(D = -0.470)$	ΔD (%)
Signal Modelling	0.017	3.2	0.015	3.2
Electrons	0.002	0.4	0.002	0.4
Muons	0.001	0.1	0.001	0.1
Jets	0.004	0.7	0.004	0.8
b-tagging	0.002	0.4	0.002	0.4
Pile-up	< 0.001	< 0.1	< 0.001	< 0.1
$E_{\text{T}}^{\text{miss}}$	0.002	0.3	0.002	0.4
Backgrounds	0.010	1.8	0.009	1.8
Total Statistical Uncertainty	0.002	0.3	0.002	0.4
Total Systematic Uncertainty	0.021	3.8	0.018	3.9
Total Uncertainty	0.021	3.8	0.018	3.9

Table 1: A summary of the effect of the groups of uncertainty at the expected SM value for $D_{\text{expected}} = -0.470$, corresponding to the Pow+Py modelling, and the observed value $D_{\text{observed}} = -0.547$, both in the signal region. The total systematic uncertainty is calculated as the sum in quadrature of the individual grouped sources of systematic uncertainty.

5 Results

The calibration curve procedure is performed in the signal region and the two validation regions, and the data are corrected to a fiducial phase space at particle level, described in Section 3. All systematic uncertainties are included in the three regions. The observed (expected) results are:

$$D = -0.547 \ \pm 0.002 \ [\text{stat.}] \pm 0.021 \ [\text{syst.}] \ (-0.470 \pm 0.002 \ [\text{stat.}] \pm 0.018 \ [\text{syst.}]) \,,$$

in the signal region of $340 < m(t\bar{t}) < 380$ GeV and:

$$D = -0.222 \ \pm 0.001 \ [\text{stat.}] \pm 0.027 \ [\text{syst.}] \ (-0.258 \pm 0.001 \ [\text{stat.}] \pm 0.026 \ [\text{syst.}]) \,,$$

$$D = -0.098 \ \pm 0.001 \ [\text{stat.}] \pm 0.021 \ [\text{syst.}] \ (-0.103 \pm 0.001 \ [\text{stat.}] \pm 0.021 \ [\text{syst.}]) \,,$$

in the validation regions of $380 < m(t\bar{t}) < 500$ GeV and $m(t\bar{t}) > 500$ GeV, respectively. The calibration curve for the signal region and a summary of the results in all regions are presented in Figure 2.

6 Discussion

The observed values of the entanglement witness D are compared to the entanglement limit. The parton level bound $D = -1/3$ is converted to a particle level bound by folding the limit to particle level, to account for parton shower effects. For Powheg + Pythia this yields -0.322 ± 0.009, where the uncertainty includes all uncertainties on the Powheg + Pythia model, with the exception of the parton shower uncertainty (for more details on these uncertainties, see Appendix C). Similarly, for Powheg + Herwig, with an angular
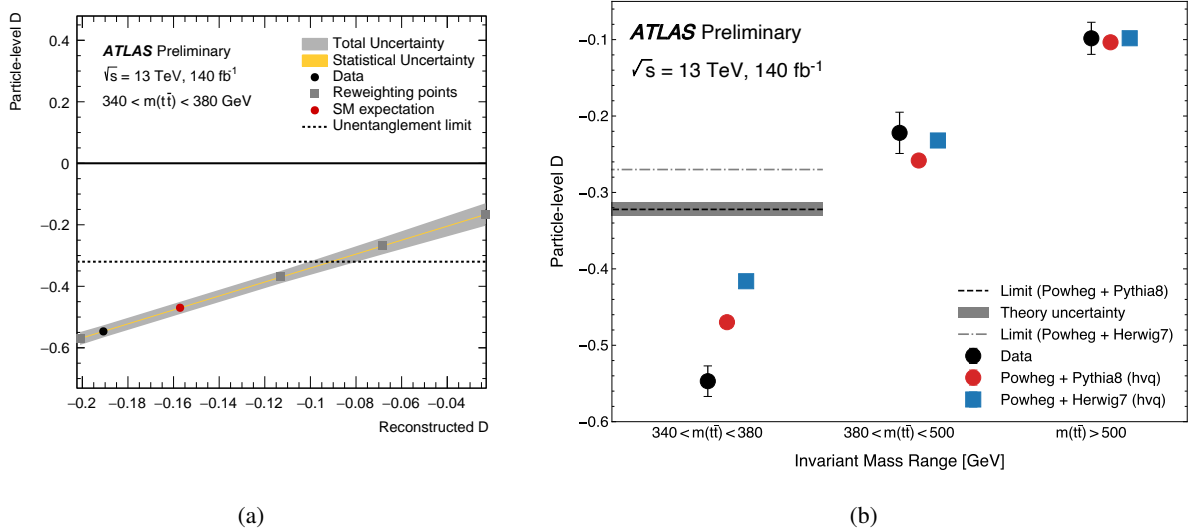
(a)   (b)

Figure 2: (a): Calibration curve for the dependence between the particle-level value of D to the reconstructed value of D, in the signal region. The yellow band represents the statistical uncertainty, while the gray band represents the total uncertainty due to statistical and systematic uncertainties. The total uncertainty is obtained by adding in quadrature the statistical and all the systematic uncertainties. The measured and the expected values are marked with black and red circles, respectively, and the entanglement limit is shown as a dashed line. (b): The particle-level D results for the signal and validation regions compared to various MC models. The entanglement limit shown is a conversion from its parton-level value of $D = -1/3$ to the corresponding value at particle level, and the uncertainties which are considered for the band are described in the text.

ordered parton shower, a value of -0.27 is obtained. No uncertainties are assigned in this case since it is merely used as an alternative model. In both of the validation regions, with no entanglement signal, the measurements are found to agree with the different MC setup predictions within uncertainties. This serves as a consistency check to validate the method used for the measurement. Even if the different models used yield different predictions, the measurements in the validation regions do not allow, at their current precision, to rule out any of the setups that were used.

In the signal region, the observed and expected significance with respect to the entanglement limit is well beyond five standard deviations independently of the MC model used for correcting the entanglement limit to account for the fiducial phase-space of the measurement. This is illustrated in Figure 2 (b), where the limit for the null hypothesis of entanglement is shown. The large discrepancy between the Monte Carlo generators stems from the different algorithms used in PYTHIA and HERWIG for the ordering of the parton shower. It is important to note that close to the threshold, non-relativistic QCD processes alter the top quark pair production [34] and are not accounted for in MC generators. Though it has higher precision, the data after correction cannot be compared directly to $bb4\ell$ as it is not possible to remove the off-shell component in a formally correct way in order to compare directly to the corrected data. However, the effect of using this model was tested in an approximate way and was found to not significantly change the conclusions of the measurement. The observed result in the region with $340 < m(t\bar{t}) < 380$ GeV establishes the formation of an entangled $t\bar{t}$ state. This constitutes the first observation of entanglement in a quark-anti-quark pair.

Apart from the fundamental interest of testing quantum entanglement in a new environment, this measurement of entanglement in top quarks paves the way to use high-energy colliders, such as the LHC, as

a laboratory to study quantum information and other foundational problems in quantum mechanics. From a quantum information perspective, this is particularly interesting due to the genuine relativistic nature of high-energy colliders, the richness of the interactions and symmetries, and the fundamental character of the Standard Model. Furthermore, highly-demanding measurements in conventional setups, such as measuring quantum discord according to its definition or reconstructing the steering ellipsoid, can be naturally implemented at the LHC due to the vast amount of available statistics [22]. From a high-energy physics perspective, borrowing concepts from quantum information theory inspires new approaches and observables that can be used to search new physics beyond the Standard Model [72–75].

Acknowledgements

Appendix

A Simulated Background Event Samples

MC simulations are used to model the expected SM backgrounds. Below the particle generators used to simulate the events are described.

To aid the analysis, simulated Monte Carlo (MC) samples were produced using either the full ATLAS detector simulation [77] based on the Geant4 framework [78] or, for the estimation of some of the systematic uncertainties, a faster simulation with parameterised showers in the calorimeters [79]. The effect of multiple interactions in the same and neighbouring bunch crossings (pile-up) was modelled by overlaying each hard-scattering event with inelastic pp collisions generated with Pythia 8.186 [80] using the NNPDF2.3 set of PDFs [66] and the A3 set of tuned parameters [81]. The EvtGen 1.6.0 program [69] is used for properties of the bottom and charm hadron decays.

Single-top tW associated production was modelled using the Powheg Box v2 [57–59, 82] generator, which provided matrix elements at next-to-leading order (NLO) in the strong coupling constant α_s in the five-flavour scheme with the NNPDF3.0nlo [60] parton distribution function (PDF) set. The functional form of the renormalisation and factorisation scales was set to the default scale, which is equal to the top-quark mass (m_{top} = 172.5 GeV). The diagram removal scheme [83] was employed to handle the interference with $t\bar{t}$ production [61]. The events were interfaced with Pythia 8.230 [64] using the A14 tune [65] and the NNPDF2.3lo PDF set. The decays of bottom and charm hadrons were simulated using the EvtGen 1.6.0 program [69]. The inclusive cross-section was corrected to the theory prediction calculated at NLO in QCD with NNLL soft-gluon corrections [84, 85]. For proton–proton collisions at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV, this cross-section corresponds to $\sigma(tW)_{\text{NLO+NNLL}}$ = 71.7 ± 3.8 pb, using a top-quark mass of m_{top} = 172.5 GeV. The uncertainty in the cross-section due to the PDF was calculated using the MSTW2008nnlo 90% CL [86, 87] PDF set, and was added in quadrature to the effect of the scale uncertainty.

Samples of diboson final states (VV) were simulated with the Sherpa 2.2.2 [88] generator, including off-shell effects and Higgs boson contributions, where appropriate. Fully leptonic final states and semileptonic final states, where one boson decays leptonically and the other hadronically, 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 $gg \rightarrow VV$ were generated using LO-accurate matrix elements for up to one additional parton emission for both the cases of fully leptonic and semileptonic final states. The matrix element calculations were matched and merged with the Sherpa parton shower based on Catani–Seymour dipole factorisation [89, 90] using the MEPS@NLO prescription [91–94]. The virtual QCD corrections were provided by the OpenLoops library [95–97]. The NNPDF3.0nnlo set of PDFs was used [60], along with the dedicated set of tuned parton shower parameters developed by the Sherpa authors.

The production of V+jets was simulated with the Sherpa 2.2.11 [88] generator using NLO matrix elements (ME) for up to two partons, and leading-order (LO) matrix elements for up to five partons calculated with the Comix [89] and OpenLoops 2 [95–98] 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 [60] was used and the samples were normalised to a next-to-next-to-leading-order (NNLO) prediction [99].

The production of $t\bar{t}V$ events was modelled using the MadGraph5_aMC@NLO 2.3.3 [100] generator, which provided matrix elements at NLO in the strong coupling constant α_s with the NNPDF3.0NLO [60] PDF. The functional form of the renormalisation and factorisation scales was set to the default of $0.5 \times \sum_i \sqrt{m_i^2 + p_{\text{T},i}^2}$, where the sum runs over all the particles generated from the matrix element calculation. Top quarks were decayed at LO using MadSpin [101, 102] to preserve spin correlations. The events were interfaced with Pythia 8.210 [64] for the parton shower and hadronisation, using the A14 set of tuned parameters [65] and the NNPDF2.3LO [60] PDF set. The decays of bottom and charm hadrons were simulated using the EvtGen 1.2.0 program [69]. The cross-sections were calculated at NLO QCD and NLO EW accuracy using MadGraph5_aMC@NLO as reported in Ref. [103]. In the case of $t\bar{t}\ell\ell$ the cross-section was scaled by an off-shell correction estimated at one-loop level in α_s.

The production of $t\bar{t}H$ events was modelled using the Powheg Box v2 [56–59, 104] generator, which provided matrix elements at NLO in the strong coupling constant α_s in the five-flavour scheme with the NNPDF3.0NLO [60] PDF set. The functional form of the renormalisation and factorisation scales was set to $\sqrt[3]{m_\text{T}(t) \cdot m_\text{T}(\bar{t}) \cdot m_\text{T}(H)}$. The events were interfaced to Pythia 8.230 [64] using the A14 tune [65] and the NNPDF2.3LO [60] PDF set. The decays of bottom and charm hadrons were performed by EvtGen 1.6.0 [69]. The cross-section was calculated at NLO QCD and NLO EW accuracy using MadGraph5_aMC@NLO as reported in Ref. [103]. The predicted value at $\sqrt{s} = 13$ TeV is 507^{+35}_{-50} fb, where the uncertainties were estimated from variations of α_s as well as the renormalisation and factorisation scales.

B Reweighting the $\cos\varphi$ distribution

In order to construct the calibration curve, templates for alternative scenarios with different degrees of entanglement, and therefore with different values of D, must be extracted. The effects of quantum entanglement are fundamental to the calculations in the MC generators and cannot be easily changed. However, the effects of entanglement can be directly accessed via the observable in the event. Therefore, an event-by-event reweighting based on this observable is used to vary the degree of entanglement. Though the measurement uses reconstructed and particle level objects, the fundamental physics of entanglement must be changed at parton level, where the observable D is well-defined. Therefore, each event is reweighted according to its parton level values of $m(t\bar{t})$ and $\cos\varphi$, as described below.

The entanglement variable D is extracted at parton level from the $\cos\varphi$ distribution by either taking the mean of the distribution $D = -3 \cdot \langle \cos\varphi \rangle$ or the slope of the normalized cross section, $\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}\cos\varphi} = \frac{1}{2}\left(1 - D\,\cos\varphi\right)$. For simplicity, the analysis always uses the mean, however, both methods are equivalent. It means that for the purpose of reweighting, one must change the slope of the $\cos\varphi$ distribution, at parton level. Each event is reweighted according to this change, which in turn changes also the distributions at particle- and detector level accordingly.

The observable D depends on $m(t\bar{t})$. To perform the reweighting, the value of D as a function of $m(t\bar{t})$, $D_\Omega(m(t\bar{t}))$, has to be calculated. This is achieved by fitting a 3rd order polynomial of the form:

$$D_\Omega(m_{t\bar{t}}) = x_0 + x_1 \cdot m_{t\bar{t}}^{-1} + x_2 \cdot m_{t\bar{t}}^{-2} + x_3 \cdot m_{t\bar{t}}^{-3} \,, \tag{4}$$

where x_0, x_1, x_2, x_3 are constants. The parameters of $D_\Omega(m(t\bar{t}))$ are MC dependent and have to be calculated for the nominal sample and for each one of the $t\bar{t}$ theory systematics, as they change the parton level $\cos\varphi$ and therefore also $D_\Omega(m(t\bar{t}))$.

The reweighting method is a simple scaling of the $\cos\varphi$ distribution according to the desired new value of D. This is done by assigning a weight for each event at parton level as:

$$w = \frac{1 - D_\Omega(m_{t\bar{t}}) \cdot X \cdot \cos\varphi}{1 - D_\Omega(m_{t\bar{t}}) \cdot \cos\varphi} \,, \tag{5}$$

with X as the scaling hypothesis of D. If, for example, $X = 1.2$, it means that D is scaled higher with respect to its nominal value by 20%. In order to build the calibration curve, 4 alternative values for D are reweighted: $X = 0.4, 0.6, 0.8, 1.2$, in addition to the nominal value with no reweighting ($X = 1.0$). It is important to note that these values shift D across the entire $m_{t\bar{t}}$ spectrum. Figure 3 shows the parton level distribution of D in the signal region after reweighting.
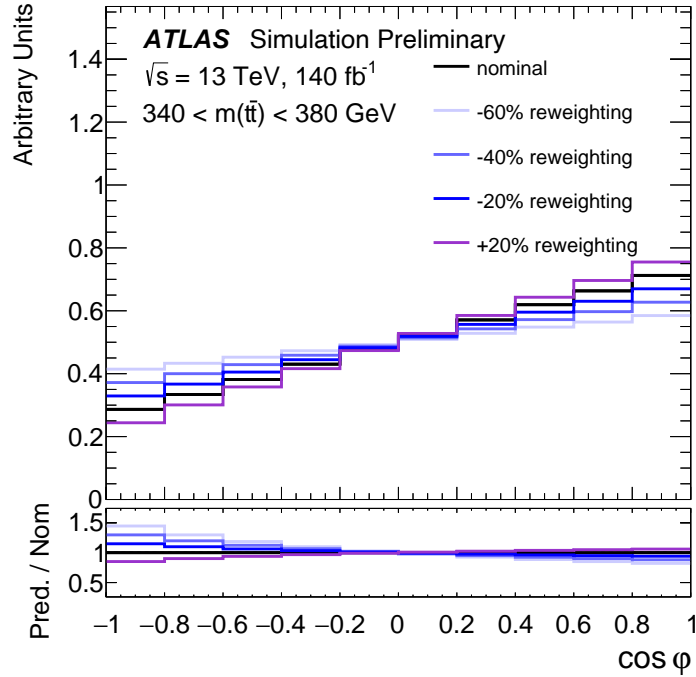

Figure 3: Example of the results of the nominal distribution and the reweighting technique with scaling hypotheses $X = 0.4, 0.6, 0.8, 1.2$ in the $m(t\bar{t}) < 380$ GeV region at parton level. The ratio shows the various reweighting prediction point "Pred." over the nominal "Nom."

C Systematic Uncertainties

The systematic uncertainties can be divided into three separate categories: signal modelling systematic uncertainties, which stem from the theoretical prediction of $t\bar{t}$ production; object systematic uncertainties, which arise due to the uncertainty in the detector response to measure the objects used in the analysis; and background modelling systematic uncertainties, which are related to the theoretical prediction of the SM backgrounds. All systematic uncertainties, grouped by their sources, are described in the following section. In total, nearly 400 sources of systematic uncertainty are considered, and the dominant ones are the signal modelling uncertainties.

For each source of systematic uncertainty, a new calibration curve is created and the expected (or observed) data are corrected, resulting in a shifted corrected result. The systematic uncertainty in most cases is taken as the difference between the nominal expected/observed result and the systematic shifted result. In cases where a systematic shift only effects the background model (e.g. background cross-section uncertainties), the systematically shifted background sample is subtracted from the data instead before the calibration is performed. In cases where the systematic uncertainty is one-sided, the uncertainty is symmetrized by taking the mirror in the opposite direction. In cases where uncertainties may be asymmetric, the larger of the two variations is symmetrized. As they dominate the measurement, the size of the expected systematics for the sources of signal modelling are presented in Table 2.

C.1 Signal modelling systematics

Signal modelling systematic uncertainties are those related to the choice of Powheg Box + Pythia as the nominal MC setup as well as uncertainties on the theoretical calculation itself. These systematics are considered in two forms; alternative generators and weights. For the alternative generator uncertainties, the difference between the calibrated D values is taken as the systematic uncertainty. For the systematic uncertainties that use weights, the difference between the calibrated D values for nominal and the sample shifted by the weights is taken as the uncertainty. These systematics are enumerated as follows:

- **pThard1 Setting**: The choice of the internal parameter of Powheg Box, which regulates the definition of the region of phase-space that is vetoed in the showering when matched to a parton shower, is varied by setting an internal parameter of Powheg Box (pThard = 1) following the prescription described in [105].

- **Top Decay**: The uncertainty on the decay of the top quarks is estimated by comparing the nominal decay in Powheg Box to the decays modeled with MadSpin [101, 102].

- **NNLO reweighting**: The uncertainty due to missing higher order corrections is estimated by reweighting the p_T of the top quarks, the p_T of the $t\bar{t}$ system and the $m(t\bar{t})$ spectra at parton level to match predicted NNLO differential cross-sections [106].

- **Parton Shower**: This uncertainty is estimated by comparing two different Parton Shower algorithms, Pythia 8.230 and Herwig 7.2, interfaced to the same ME generator (Powheg Box).

- **Recoil to Top**: The nominal sample uses a recoil scheme where partons recoil against b-quarks. This recoil scheme changes the modelling of second and subsequent gluon emission from quarks produced by coloured resonance decays, such as the b-quark in a top quark decay, and therefore affects how the momentum is re-arranged between the W boson and the b-quark. An alternative sample is produced in which the recoil is set against the top quark itself for the second and subsequent emissions [107].

- **Scale Uncertainties:** Generator weights are used to estimate the effect of increasing and decreasing the renormalisation (μ_R) and factorisation (μ_F) scales by a factor of 2 in the nominal Powheg setup, including varying both scales simultaneously up and down. The envelope of each of these variations is taken as the final uncertainty.

- **Initial State Radiation**: The uncertainty due to initial-state radiation (ISR) was estimated by choosing the Var3c up/down variants of the A14 tune as described in Ref. [108].

- **Final State Radiation**: The impact of final-state radiation (FSR) was evaluated by varying the renormalisation scale for emissions from the parton shower up or down by a factor two.

14

- **Parton Distribution Function**: The systematic uncertainty due to the choice of PDF is assessed using the PDF4LHC15 30 eigenvector decomposition [109]. The full difference between the nominal PDF4LHC prediction and the variation is taken and symmetrised for each eigenvector. These eigenvector variations exist as generator weights in the nominal PowHEG Box sample. For brevity, the quadrature sum of each of these variations is then quoted in Table 2.

- **h_{damp}**: The h_{damp} parameter is a resummation damping factor and one of the parameters that controls the matching of PowHEG Box matrix elements to the parton shower and thus effectively regulates the high-p_{T} radiation against which the $t\bar{t}$ system recoils. The systematic uncertainty due to the setting of the h_{damp} parameter is assessed by comparing the nominal PowHEG Box + PYTHIA to one with a factor of two higher setting for the h_{damp} parameter. This difference is then symmetrised.

- **Top Quark Mass**: The uncertainty on the top quark mass, m_{top}, is applied by comparing the nominal sample with alternative samples that uses different values of the top mass as an input. The nominal value is 172.5 GeV, and the alternative samples use 172.0 and 173.0 GeV.

Systematic	Relative Size [D = SM (-0.47)]
Top-quark Decay	1.6 %
Parton Distribution Function	1.2 %
Recoil To Top	1.1 %
Final State Radiation	1.1 %
Scale Uncertainties	1.1 %
NNLO Reweighting	1.1 %
pThard1 Setting	0.8 %
Top-quark Mass	0.7 %
Initial State Radiation	0.2 %
Parton Shower	0.2 %
h_{damp} Setting	0.1 %

Table 2: A comparison of the relative size of the uncertainties related to signal modelling at the SM expectation point with respect to D_{particle}. The PDF uncertainties are the quadrature sum of all variations.

C.2 Object systematics

Systematic uncertainties which originate due to the uncertainty in the detector response to measure the objects used in the analysis are estimated.

- **Electron:** The systematic uncertainties considered for electrons mainly arise due to corrections based on the uncertainty in their trigger, reconstruction, identification, and isolation and are measured using tag-and-probe measurements in Z and J/ψ decays [42, 110]. The electron-based systematics have a negligible impact on the final measurement, with a total contribution of about 0.2%.

- **Muon:** The systematic uncertainties considered for muons arise due to the uncertainty on their trigger, identification, resolution, energy scale, and isolation and are measured using tag-and-probe measurements in Z and J/ψ decays [43–45]. The muon-based systematics have a negligible impact on the final measurement, with a total contribution of about 0.3%.

- **Jet:** The systematic uncertainties associated with jets are separated into those related to the Jet Energy Scale (JES), the Jet Energy Resolution (JER), and the Jet Vertex Tagging (JVT). The JES is derived from test-beam data, LHC collision data and simulation [46] and is measured to a precision of between 1 and 4% for jets used in this analysis, with the uncertainty falling at higher p_T. The highest impact of a single source is 0.2%.

- **b-tagging:** The estimation of these uncertainties is described in Ref. [111]. A total of 17 independent systematic variations are taken into account; 9 related to B-hadrons, 4 related to C-hadrons, 4 related to light-jet misidentification, and two high-p_T extrapolation uncertainties. The highest impact of a single systematic variation is 0.4%.

- **E_T^{miss}:** All object-based uncertainties are fully correlated with the E_T^{miss} object construction, however, there are some uncertainties specific to the construction of the E_T^{miss} object based on the soft tracks that are not associated to leptons or jets. These uncertainties are divided into parallel and perpendicular response as well as a scale uncertainty [52]. These have a negligible effect on the measurement.

- **Pile-up:** The effect of multiple interactions in the same and neighboring bunch crossings (pile-up) was modeled by overlaying the simulated hard-scattering events with inelastic pp events. In order to assess the systematic uncertainty due to pile-up, the reweighing to match simulation to data is varied within its uncertainty [50]. The resulting uncertainty has a less than 0.1% effect on the measurement and is therefore not a significant source of systematic uncertainty.

- **Luminosity:** The luminosity uncertainty only changes the total normalization of the signal and background samples. D is calculated from the normalized $\cos\varphi$ distribution and therefore is not affected by varying the total normalization. However, the total expected statistical uncertainty can be affected by the luminosity uncertainty. This analysis uses the latest luminosity estimate of 140.07 ± 1.17 fb^{-1} [112]. The final effect on the measurement is less than 0.1%.

C.3 Background modelling systematics

Background events are a relatively small source of uncertainty in this measurement since the event selection and top quark reconstruction, especially the $m(t\bar{t})$ constraint, tend to suppress them. The uncertainties and their sources are listed in the following.

- **Single top quark:** Two uncertainties are currently considered for the single top background; a cross-section rate uncertainty of 5.3% based on the NNLO cross-section uncertainty [85], and an uncertainty based on the scheme used to remove higher-order diagrams that overlap with the $t\bar{t}$ process. For the latter, the nominal POWHEG BOX + PYTHIA sample, generated with the diagram removal scheme [83], was compared with an alternative sample generated using the diagram subtraction scheme [61, 83]. The cross-section uncertainty has a 0.4% effect on the measurement whereas the diagram scheme has less than 0.1% effect on the measurement.

- **$t\bar{t} + X$:** A single normalisation uncertainty is currently considered for each one of the $t\bar{t} + X$ backgrounds: a cross-section rate uncertainty of $^{+10.4\%}_{-12.0\%}$ for $t\bar{t} + Z$, and $^{+13.3\%}_{-12.0\%}$ for $t\bar{t} + W$. Both are based on the NLO cross-section uncertainty derived from renormalisaton and factorisation scale variations and PDF uncertainties in the ME calculation. These uncertainties have a negligible effect on the measurement, since the contribution of the processes in the signal region is very small.

- **Diboson:** A normalisation uncertainty is considered for the diboson process of $\pm 10\%$ to account for the difference between the NLO precision of the SHERPA generator and precision NNLO QCD + NLO EW correction theoretical cross-sections. This simple k-factor approach is taken rather than a more elaborate prescription, as the diboson process is a small background and the phase space selected by the analysis ($m(t\bar{t}) < 380$ GeV) is unlikely to be sensitive to shape effects in the EW corrections typically observed in high p_T tails. This uncertainty has less than a 0.1% effect on the measurement.

- **$Z \to \tau\tau$:** A conservative cross-section uncertainty of $\pm 20\%$ is applied to the $Z \to \tau\tau$ background in order to account for the uncertainty in the cross-section prediction (which is significantly smaller than this variation) as well as to account for possible mismodelling of the rate of associated heavy flavour production that is typically observed in ee and $\mu\mu$ dileptonic $t\bar{t}$ analyses, which was estimated to be a 5%(3%) effect in previous iterations of this analysis that included the $ee(\mu\mu)$ channels. This assumption is conservative as it is not possible to isolate a pure $Z \to \tau\tau$ control region in which to estimate this effect and therefore additional lepton-flavour related effects that exist in the ee and $\mu\mu$ channels are also included. This uncertainty has a noticeable impact on the final measurement and is the largest background-related uncertainty. The reason that the systematic becomes large, despite this background being relatively small, is that the reconstructed level $\cos\varphi$ is flat and therefore subtracting even a relatively small amount of $Z \to \tau\tau$ background can noticeably effect the mean and therefore the D observable. This uncertainty has an impact of 1.5% effect on the measurement.

- **Fake and non-prompt leptons:** A normalisation uncertainty of $\pm 50\%$ is taken to account for the uncertainty in the total fakes yield in the signal region compared to the same-sign control region in order to assure adequate coverage of the understanding of the rates of these types of events. It is a conservative uncertainty based on the observed data-MC agreement in the same-sign region. The uncertainty has only a 0.1% effect on the final measurement.

C.4 Expected *D* systematic uncertainties

A list of the most significant systematic uncertainties for the expected value of D are presented in Table 3 for the SM point ($D = -0.47$) and for the -20% point ($D = -0.36$) in Table 4, both for the signal region. In general, the majority of systematic uncertainties that are considered are inconsequential to the measurement and the dominant sources of systematic uncertainty arise mostly due to the signal modelling and this is true for the validation regions as well.

D Parton shower and hadronisation effects

The studies documented in the following aim to understand the difference of entanglement- and spin-correlation-related observables that are visible in different parton shower and hadronisation algorithms in more detail. The nominal MC sample is produced with the NLO ME implemented in POWHEG BOX (hvq). The four-momenta produced with POWHEG BOX are then either interfaced with PYTHIA 8.230 or with HERWIG 7.21 for the PS, hadronisation and underlying event model.

The origin of this scrutiny is the observation of a large difference between the POWHEG BOX + PYTHIA and POWHEG BOX + HERWIG predictions used in the paper. Both setups of POWHEG BOX with PYTHIA and with HERWIG predict very different $\cos\varphi$ distributions at detector level and at particle level, while the

Leading Systematics	Relatvie Size [D = SM (−0.47)]
Top-quark decay	1.6 %
$Z \to \tau\tau$ Cross-section	1.5 %
Recoil To Top	1.1 %
Final State Radiation	1.1 %
Scale Uncertainties	1.1 %
NNLO Reweighting	1.1 %
Parton Distribution Function (5)	0.8 %
pThard1 Setting	0.8 %
Top-quark Mass	0.7 %
Single Top Quark Wt Cross-section	0.4 %

Table 3: Summary of the top 10 most important systematic uncertainties for the signal region, in the measurement at the SM expectation.

Leading Systematics	Relative Size [D = −20% of SM (−0.36)]
Parton Shower	2.6 %
Top decay	2.3 %
Final State Radiation	1.8 %
Recoil To Top	1.5 %
$Z \to \tau\tau$ Cross-section	1.4 %
NNLO Reweighting	1.4 %
Scale Uncertainties	1.3 %
Top mass	1.1 %
Parton Distribution Function (5)	0.9 %
pThard1 Setting	0.9 %

Table 4: Summary of the top 10 most important systematic uncertainties for the signal region, in the measurement at the −20% of the SM expectation.

parton level predictions are nearly identical. Therefore, with the current Monte Carlo generators, a parton level measurement would suffer from the ambiguity in $\cos \varphi$. Apart from using a different tuning strategy, there are two main differences between our two nominal setups: the hadronisation model and the shower ordering. While PYTHIA is based on the Lund-string model and uses a p_T-ordered shower [113–115], HERWIG is based on a cluster model and uses as default an angular-ordered shower [116].

A comparison of Monte Carlo simulations with different hadronisation models showed that these have a negligible effect, both inclusively and in a smaller part of phase space which corresponds to the signal region of the analysis, with $340 < m(t\bar{t}) < 380$ GeV. The majority of the differences seem to originate from the different ordering in the parton shower. To illustrate this, particle level comparisons between different generator setups are done with $\cos \varphi$. The particle level distributions for the POWHEG BOX + PYTHIA and POWHEG BOX + HERWIG samples used in the analysis are shown, together with distributions of two different setups of HERWIG 7. In these setups, HERWIG 7 is used both for the production of the $t\bar{t}$ events as well as for the shower, hadronisation and underlying event. The samples are produced at LO, using either a dipole-ordered shower or an angular-ordered shower. The comparisons are shown at particle level

in Figures 4 and 5, inclusively (without a cut on $m(t\bar{t})$) and in the signal region, respectively. A difference of up to 6% is observed when comparing Powheg Box + Pythia to Powheg Box + Herwig. The same behavior is observed when comparing the two different showering orders for Herwig.

The similarities between the samples used in this analysis and the Herwig samples with different showering orders implies that the ordering of the shower is the main cause of the observed differences. These findings lead to the conclusion that performing the measurement at particle level is more attractive, since the overall uncertainties are smaller. Given that neither Powheg Box + Pythia nor Powheg Box + Herwig can be discarded using the validation regions of the analysis, a parton level measurement will suffer from a large ambiguity. The treatment of spin effects in Monte Carlo generators combining the Matrix Element with a Parton Shower requires special attention for future higher-precision quantum information studies at the LHC.
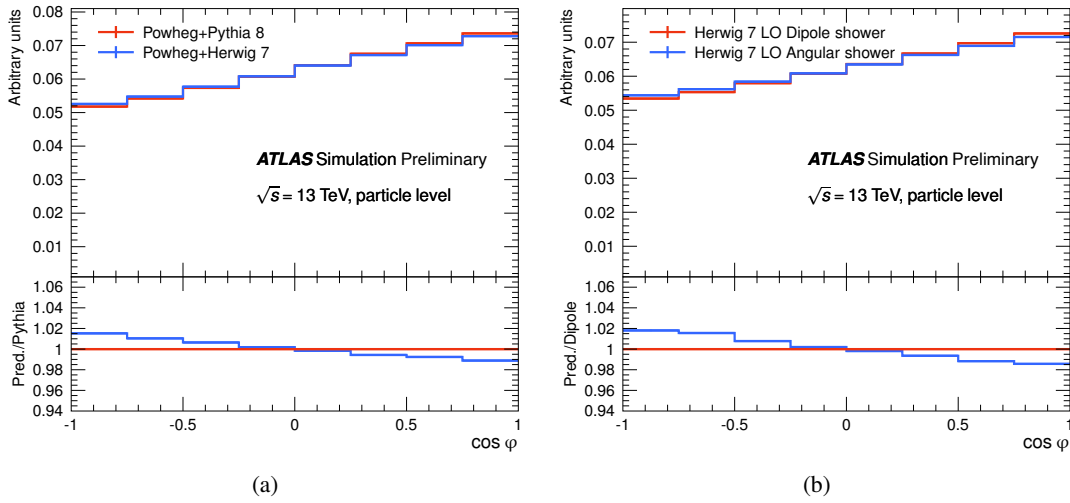

Figure 4: Comparison of $\cos\varphi$ (inclusive) for (a) Powheg Box and (b) Herwig setups at particle level after an inclusive $t\bar{t}$ selection but without any requirements on the invariant mass of the top-quark pair.

E Reconstructed level results in validation regions

The distribution of $\cos\varphi$ in the signal region and the reconstructed level D after background subtraction in the two validation regions are shown on the left and right panels of Figures 6 and 7.
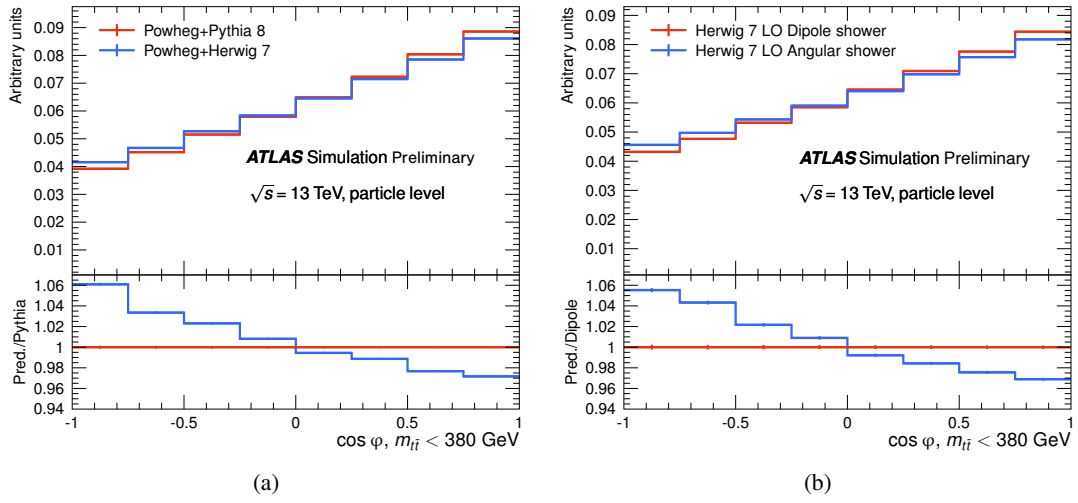
Figure 5: Comparison of $\cos\varphi$ in the entanglement region for (a) Powheg Box and (b) Herwig setups at particle level in the signal region.
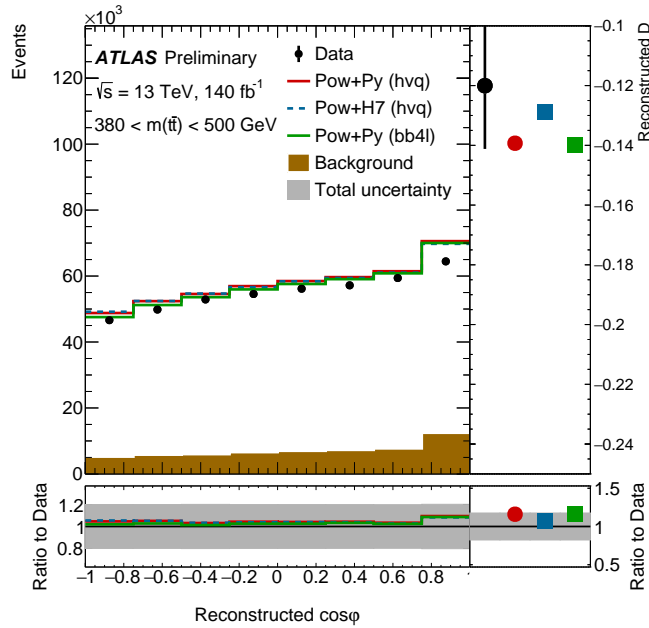

Figure 6: The $\cos\varphi$ observable in validation region 1 at reconstructed level (left) and the D at the reconstructed detector level distribution from different three different MC generators; the Powheg Box + Pythia and and Powheg Box + Herwig heavy quark models, labelled "Pow+Py (hvq)" and "Pow + H7 (hvq)", respectively, and the Powheg Box + Pythia bb4ℓ models, labelled "Pow+Py (bb4ℓ)", after backgrounds are subtracted (right). The uncertainty band includes all sources of uncertainty added in quadrature. The ratio of the predictions with respect to data is shown at the bottom of the figure. The D quoted for the bb4ℓ model also includes a subtraction of the single-top-quark background.
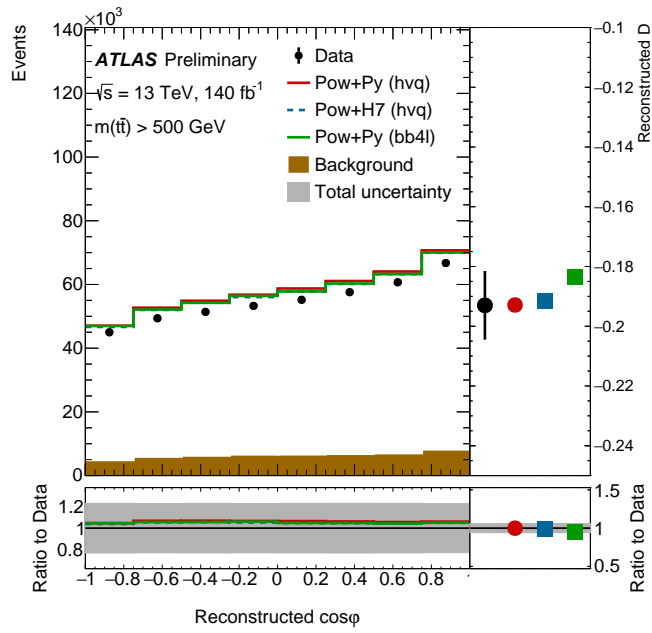
Figure 7: The cos φ observable in the validation region 2 at reconstructed level (left) and the D at the reconstructed detector level distribution from different three different MC generators; the Powheg Box + Pythia and and Powheg Box + Herwig heavy quark models, labelled "Pow+Py (hvq)" and "Pow + H7 (hvq)", respectively, and the Powheg Box + Pythia bb4ℓ models, labelled "Pow+Py (bb4ℓ)", after backgrounds are subtracted (right). The uncertainty band includes all sources of uncertainty added in quadrature. The ratio of the predictions with respect to data is shown at the bottom of the figure. The D quoted for the bb4ℓ model also includes a subtraction of the single-top-quark background.

References

[1] A. Einstein, B. Podolsky and N. Rosen,
     Can quantum mechanical description of physical reality be considered complete?,
     Phys. Rev. **47** (1935) 777 (cit. on p. 2).

[2] E. Schrodinger, *Discussion of Probability Relations between Separated Systems*,
     Pro. Cambridge Phi. Soc. **31** (1935) 555 (cit. on p. 2).

[3] J. S. Bell, *On the Einstein-Podolsky-Rosen paradox*, Physics Physique Fizika **1** (1964) 195
     (cit. on p. 2).

[4] C. H. Bennett and D. P. DiVincenzo, *Quantum information and computation*,
     Nature **404** (2000) 247 (cit. on p. 2).

[5] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*,
     Cambridge Series on Information and the Natural Sciences, Cambridge University Press, 2000,
     ISBN: 9780521635035, URL: https://books.google.es/books?id=65FqEKQOfP8C
     (cit. on p. 2).

[6] C. D. Marciniak et al., *Optimal metrology with programmable quantum sensors*,
     Nature **603** (2022) 604 (cit. on p. 2).

[7] R. Horodecki, P. Horodecki, M. Horodecki and K. Horodecki, *Quantum entanglement*,
     Rev. Mod. Phys. **81** (2009) 865, arXiv: quant-ph/0702225 (cit. on p. 2).

[8] H. Casini and M. Huerta, *Lectures on entanglement in quantum field theory*,
     PoS **TASI2021** (2023) 002, arXiv: 2201.13310 [hep-th] (cit. on p. 2).

[9] A. Aspect, P. Grangier and G. Roger, *Experimental Realization of Einstein-Podolsky-Rosen-Bohm
     Gedankenexperiment: A New Violation of Bell's Inequalities*, Phys. Rev. Lett. **49** (2 1982) 91,
     URL: https://link.aps.org/doi/10.1103/PhysRevLett.49.91 (cit. on p. 2).

[10] E. Hagley et al., *Generation of Einstein-Podolsky-Rosen Pairs of Atoms*,
     Phys. Rev. Lett. **79** (1 1997) 1,
     URL: https://link.aps.org/doi/10.1103/PhysRevLett.79.1 (cit. on p. 2).

[11] M. Steffen et al.,
     Measurement of the entanglement of two superconducting qubits via state tomography,
     Science **313** (2006) 1423 (cit. on p. 2).

[12] W. Pfaff et al., *Demonstration of entanglement-by-measurement of solid-state qubits*,
     Nature Physics **9** (2013) 29 (cit. on p. 2).

[13] A. Go et al.,
     Measurement of EPR-type flavour entanglement in Upsilon(4S) —> B0 anti-B0 decays,
     Phys. Rev. Lett. **99** (2007) 131802, arXiv: quant-ph/0702267 (cit. on p. 2).

[14] B. Julsgaard, A. Kozhekin and E. S. Polzik,
     Experimental long-lived entanglement of two macroscopic objects, Nature **413** (2001) 400
     (cit. on p. 2).

[15] K. Lee et al., *Entangling Macroscopic Diamonds at Room Temperature*,
     Science (New York, N.Y.) **334** (2011) 1253 (cit. on p. 2).

[16] C. Ockeloen-Korppi et al., *Stabilized entanglement of massive mechanical oscillators*,
     Nature **556** (2018) 478 (cit. on p. 2).

[17] Y. Afik and J. R. M. de Nova, *Entanglement and quantum tomography with top quarks at the LHC*, Eur. Phys. J. Plus **136** (2021) 907, arXiv: `2003.02280 [quant-ph]` (cit. on pp. 2–4).

[18] M. Fabbrichesi, R. Floreanini and G. Panizzo, *Testing Bell Inequalities at the LHC with Top-Quark Pairs*, Phys. Rev. Lett. **127** (2021) 161801, arXiv: `2102.11883 [hep-ph]` (cit. on p. 2).

[19] C. Severi, C. D. E. Boschi, F. Maltoni and M. Sioli, *Quantum tops at the LHC: from entanglement to Bell inequalities*, Eur. Phys. J. C **82** (2022) 285, arXiv: `2110.10112 [hep-ph]` (cit. on p. 2).

[20] Y. Afik and J. R. M. de Nova, *Quantum information with top quarks in QCD*, Quantum **6** (2022) 820, arXiv: `2203.05582 [quant-ph]` (cit. on pp. 2, 3).

[21] J. A. Aguilar-Saavedra and J. A. Casas, *Improved tests of entanglement and Bell inequalities with LHC tops*, Eur. Phys. J. C **82** (2022) 666, arXiv: `2205.00542 [hep-ph]` (cit. on p. 2).

[22] Y. Afik and J. R. M. de Nova, *Quantum Discord and Steering in Top Quarks at the LHC*, Phys. Rev. Lett. **130** (2023) 221801, arXiv: `2209.03969 [quant-ph]` (cit. on pp. 2, 10).

[23] R. Ashby-Pickering, A. J. Barr and A. Wierzchucka, *Quantum state tomography, entanglement detection and Bell violation prospects in weak decays of massive particles*, JHEP **05** (2023) 020, arXiv: `2209.13990 [quant-ph]` (cit. on p. 2).

[24] J. A. Aguilar-Saavedra, *Post-decay quantum entanglement in top pair production*, (2023), arXiv: `2307.06991 [hep-ph]` (cit. on p. 2).

[25] R. L. Workman et al., *Review of Particle Physics*, PTEP **2022** (2022) 083C01 (cit. on pp. 3, 6).

[26] G. Mahlon and S. J. Parke, *Angular correlations in top quark pair production and decay at hadron colliders*, Phys. Rev. D **53** (1996) 4886, arXiv: `hep-ph/9512264` (cit. on p. 3).

[27] ATLAS Collaboration, *Observation of spin correlation in t$\bar{t}$ events from pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector*, Phys. Rev. Lett. **108** (2012) 212001, arXiv: `1203.4081 [hep-ex]` (cit. on p. 3).

[28] ATLAS Collaboration, *Measurement of Spin Correlation in Top–Antitop Quark Events and Search for Top Squark Pair Production in pp Collisions at $\sqrt{s}$ = 8 TeV Using the ATLAS Detector*, Phys. Rev. Lett. **114** (2015) 142001, arXiv: `1412.4742 [hep-ex]` (cit. on p. 3).

[29] ATLAS Collaboration, *Measurements of top-quark pair spin correlations in the eμ channel at $\sqrt{s}$ = 13 TeV using pp collisions in the ATLAS detector*, Eur. Phys. J. C **80** (2020) 754, arXiv: `1903.07570 [hep-ex]` (cit. on p. 3).

[30] CMS Collaboration, *Measurements of t$\bar{t}$ spin correlations and top-quark polarization using dilepton final states in pp collisions at $\sqrt{s}$ = 7 TeV*, Phys. Rev. Lett. **112** (2014) 182001, arXiv: `1311.3924 [hep-ex]` (cit. on p. 3).

[31] CMS Collaboration, *Measurement of the top quark polarization and t$\bar{t}$ spin correlations using dilepton final states in proton–proton collisions at $\sqrt{s}$ = 13 TeV*, Phys. Rev. D **100** (2019) 072002, arXiv: `1907.03729 [hep-ex]` (cit. on p. 3).

[32] A. Brandenburg, Z. G. Si and P. Uwer, *QCD corrected spin analyzing power of jets in decays of polarized top quarks*, Phys. Lett. B **539** (2002) 235, arXiv: `hep-ph/0205023` (cit. on p. 3).

[33] W. Bernreuther, M. Flesch and P. Haberl,
Signatures of Higgs bosons in the top quark decay channel at hadron colliders,
Phys. Rev. D **58** (1998) 114031, arXiv: hep-ph/9709284 (cit. on p. 3).

[34] Y. Kiyo, J. H. Kuhn, S. Moch, M. Steinhauser and P. Uwer,
Top-quark pair production near threshold at LHC, Eur. Phys. J. C **60** (2009) 375,
arXiv: 0812.0919 [hep-ph] (cit. on pp. 3, 9).

[35] J. H. Kuhn and E. Mirkes, *QCD corrections to toponium production at hadron colliders*,
Phys. Rev. D **48** (1993) 179, arXiv: hep-ph/9301204 (cit. on p. 3).

[36] A. Petrelli, M. Cacciari, M. Greco, F. Maltoni and M. L. Mangano,
NLO production and decay of quarkonium, Nucl. Phys. B **514** (1998) 245,
arXiv: hep-ph/9707223 (cit. on p. 3).

[37] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
JINST **3** (2008) S08003 (cit. on p. 4).

[38] 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. 4),
Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL:
https://cds.cern.ch/record/1451888.

[39] 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. 4).

[40] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*,
ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187
(cit. on p. 4).

[41] 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. 4).

[42] 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 pp. 4, 15).

[43] 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 pp. 4, 15).

[44] 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 pp. 4, 15).

[45] ATLAS Collaboration, *Studies of the muon momentum calibration and performance of the ATLAS
detector with pp collisions at $\sqrt{s} = 13\,TeV$*, Eur. Phys. J. C **83** (2022) 686,
arXiv: 2212.07338 [hep-ex] (cit. on pp. 4, 15).

[46] 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. 4, 16).

[47] 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. 4).

[48] 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. 4).

[49] 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. 4).

[50] 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. 4, 16).

[51] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, Eur. Phys. J. C **83** (2022) 681, arXiv: `2211.16345 [physics.data-an]` (cit. on p. 4).

[52] 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 pp. 4, 16).

[53] 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 p. 4).

[54] 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. 5).

[55] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. B **659** (2008) 119, arXiv: `0707.1378 [hep-ph]` (cit. on p. 5).

[56] 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 pp. 5, 12).

[57] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: `hep-ph/0409146` (cit. on pp. 5, 11, 12).

[58] 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 pp. 5, 11, 12).

[59] 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 pp. 5, 11, 12).

[60] The NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv: `1410.8849 [hep-ph]` (cit. on pp. 5, 11, 12).

[61] 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. 5, 11, 16).

[62] T. Ježo and P. Nason, *On the Treatment of Resonances in Next-to-Leading Order Calculations Matched to a Parton Shower*, JHEP **12** (2015) 065, arXiv: `1509.09071 [hep-ph]` (cit. on p. 5).

[63] T. Ježo, J. M. Lindert, P. Nason, C. Oleari and S. Pozzorini, *An NLO+PS generator for $t\bar{t}$ and Wt production and decay including non-resonant and interference effects*, Eur. Phys. J. C **76** (2016) 691, arXiv: `1607.04538 [hep-ph]` (cit. on p. 5).

[64] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv: `1410.3012 [hep-ph]` (cit. on pp. 5, 11, 12).

[65] 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 pp. 5, 11, 12).

[66] NNPDF Collaboration, R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on pp. 5, 11).

[67] M. Bähr et al., *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639, arXiv: 0803.0883 [hep-ph] (cit. on p. 5).

[68] J. Bellm et al., *Herwig 7.2 release note*, Eur. Phys. J. C **80** (2020) 452, arXiv: 1912.06509 [hep-ph] (cit. on p. 5).

[69] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on pp. 5, 11, 12).

[70] B. A. Betchart, R. Demina and A. Harel, *Analytic solutions for neutrino momenta in decay of top quarks*, Nucl. Instrum. Meth. A **736** (2014) 169, arXiv: 1305.1878 [hep-ph] (cit. on p. 6).

[71] B. Abbott et al., *Measurement of the top quark mass using dilepton events*, Phys. Rev. Lett. **80** (1998) 2063, arXiv: hep-ex/9706014 (cit. on p. 6).

[72] R. Aoude, E. Madge, F. Maltoni and L. Mantani, *Quantum SMEFT tomography: Top quark pair production at the LHC*, Phys. Rev. D **106** (2022) 055007, arXiv: 2203.05619 [hep-ph] (cit. on p. 10).

[73] M. Fabbrichesi, R. Floreanini and E. Gabrielli, *Constraining new physics in entangled two-qubit systems: top-quark, tau-lepton and photon pairs*, Eur. Phys. J. C **83** (2023) 162, arXiv: 2208.11723 [hep-ph] (cit. on p. 10).

[74] C. Severi and E. Vryonidou, *Quantum entanglement and top spin correlations in SMEFT at higher orders*, JHEP **01** (2023) 148, arXiv: 2210.09330 [hep-ph] (cit. on p. 10).

[75] R. Aoude, E. Madge, F. Maltoni and L. Mantani, *Probing new physics through entanglement in diboson production*, (2023), arXiv: 2307.09675 [hep-ph] (cit. on p. 10).

[76] ATLAS Collaboration, *ATLAS Computing Acknowledgements*, ATL-SOFT-PUB-2023-001, 2023, URL: https://cds.cern.ch/record/2869272 (cit. on p. 10).

[77] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on p. 11).

[78] S. Agostinelli et al., *Geant4 – a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 11).

[79] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, 2010, URL: https://cds.cern.ch/record/1300517 (cit. on p. 11).

[80] 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. 11).

[81] 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. 11).

[82] 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. 11).

[83] 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. 11, 16).

[84] 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. 11).

[85] N. Kidonakis, 'Top Quark Production', *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013)* (JINR, Dubna, Russia, 15th–28th July 2013) 139, arXiv: 1311.0283 [hep-ph] (cit. on pp. 11, 16).

[86] 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. 11).

[87] 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. 11).

[88] E. Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Phys. **7** (2019) 034, arXiv: 1905.09127 [hep-ph] (cit. on p. 11).

[89] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 11).

[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. 11).

[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. 11).

[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. 11).

[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. 11).

[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. 11).

[95] F. Buccioni et al., *OpenLoops 2*, Eur. Phys. J. C **79** (2019) 866, arXiv: 1907.13071 [hep-ph] (cit. on p. 11).

[96] 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. 11).

[97] 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. 11).

[98] F. Buccioni, S. Pozzorini and M. Zoller, *On-the-fly reduction of open loops*,
Eur. Phys. J. C **78** (2018) 70, arXiv: 1710.11452 [hep-ph] (cit. on p. 11).

[99] C. Anastasiou, L. Dixon, K. Melnikov and F. Petriello, *High-precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*,
Phys. Rev. D **69** (2004) 094008, arXiv: hep-ph/0312266 (cit. on p. 11).

[100] 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 p. 12).

[101] S. Frixione, E. Laenen, P. Motylinski and B. R. Webber, *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*, JHEP **04** (2007) 081,
arXiv: hep-ph/0702198 (cit. on pp. 12, 14).

[102] 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 pp. 12, 14).

[103] 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. 12).

[104] H. B. Hartanto, B. Jäger, L. Reina and D. Wackeroth,
Higgs boson production in association with top quarks in the POWHEG BOX,
Phys. Rev. D **91** (2015) 094003, arXiv: 1501.04498 [hep-ph] (cit. on p. 12).

[105] 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, arXiv: 2106.10987 [hep-ph] (cit. on p. 14).

[106] M. Czakon, D. Heymes and A. Mitov,
High-precision differential predictions for top-quark pairs at the LHC,
Phys. Rev. Lett. **116** (2016) 082003, arXiv: 1511.00549 [hep-ph] (cit. on p. 14).

[107] ATLAS Collaboration, *Measurement of the top-quark mass using a leptonic invariant mass in pp collisions at $\sqrt{s} = 13\,TeV$ with the ATLAS detector*, JHEP **06** (2022) 019,
arXiv: 2209.00583 [hep-ex] (cit. on p. 14).

[108] ATLAS Collaboration,
Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO,
ATL-PHYS-PUB-2017-007, 2017, url: https://cds.cern.ch/record/2261938
(cit. on p. 14).

[109] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001,
arXiv: 1510.03865 [hep-ph] (cit. on p. 15).

[110] 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 p. 15).

[111] 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. 16).

[112] 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. 16).

[113] G. Gustafson and J. Hakkinen, *Color interference and confinement effects in W pair production*, Z. Phys. C **64** (1994) 659 (cit. on p. 18).

[114] L. Lonnblad, *Reconnecting colored dipoles*, Z. Phys. C **70** (1996) 107 (cit. on p. 18).

[115] C. Friberg, G. Gustafson and J. Hakkinen, *Color connections in e+ e- annihilation*, Nucl. Phys. B **490** (1997) 289, arXiv: `hep-ph/9604347` (cit. on p. 18).

[116] B. R. Webber, *Color reconnection and Bose-Einstein effects*, J. Phys. G **24** (1998) 287, ed. by P. J. Dornan and B. R. Webber, arXiv: `hep-ph/9708463` (cit. on p. 18).