ATLAS CONF Note

ATLAS-CONF-2017-069

20th September 2017

Measurement of colour flow using jet-pull observables in $t\bar{t}$ events with the ATLAS experiment at $\sqrt{s} = 13$ TeV

The ATLAS Collaboration

Previous phenomenological studies and measurements have shown that weighted angular moments derived from jet constituents encode the colour connections between initiating partons which seed the jets. This note presents measurements of two such distributions, the jet-pull angle and jet-pull magnitude, both of which are derived from the jet-pull-angular moment. The measurement is performed in $t\bar{t}$ events with one leptonically decaying W boson and one hadronically decaying W boson using 36.1 fb^{-1} of pp collision data recorded by the ATLAS detector at $\sqrt{s} = 13$ TeV delivered by the Large Hadron Collider. The observables are measured for two di-jet systems, corresponding to the colour-connected daughters of the W boson and the two b-jets from the top quark decays. To allow the comparison of the measured distributions to colour model predictions, they are unfolded to particle level, thereby removing experimental effects introduced by the detector. In general, the observables sensitive to colour flow remain poorly modelled by MC predictions. While good agreement can be found for individual prediction and observables combinations, none of the predictions describes the data well across all observables.

ATLAS-CONF-2017-069
25 September 2017

1 Introduction

In high energy hadron collisions, such as those produced at the Large Hadron Collider (LHC) [1] at CERN, quarks and gluons are produced abundantly. However, due to the confining nature of Quantum Chromodynamics (QCD), the direct measurement of the interactions that occur between these particles is impossible and only colour neutral hadrons can be measured. In practice, the high energy quarks and gluons are measured as jets, which are bunches of collimated hadrons that form in the evolution of the coloured initial particles. The colour connections between the high energy particles affect the structure of emitted radiation and therefore also the structure of the resulting jets. For example, soft gluon radiation is suppressed in some regions of phase space compared to others. A smaller effect arises in the process of hadronisation, where phenomenological models are used to provide an approximate description, as for example the colour string model [2].

Providing evidence for the existence of the connections — the *colour flow* — between particles is important for the validation of the phenomenological description. Using the energy-weighted distributions of particles within and between jets has been a long-standing tool for investigating colour flow, with early measurements at PETRA [3] and LEP [4, 5] using a leptonic initial state. Later, using the abundance of a hadronic initial state at the Tevatron, a precursor of the jet pull was studied [6]. Recently, the colour flow was measured in $t\bar{t}$ events at the LHC with the ATLAS experiment at a centre-of-mass energy of $\sqrt{s} = 8$ TeV [7] using the jet-pull angle.

Despite the long standing history of measurements of the colour connections, they remain a poorly constrained effect of QCD and require further dedicated experimental input to be improved. Furthermore, it may be possible to use the extracted colour information to distinguish between event topologies with different colour structure. In this case, it may provide knowledge complementary to the kinematic properties of jets, allowing to handle otherwise irreducible backgrounds or facilitate the correct assignment of jets to a desired physical process. For example, a colour flow observable could be used to resolve the disambiguity in assigning b-jets to the Higgs boson decay in $t\bar{t}H(\rightarrow b\bar{b})$ events. In this note the observed data are compared to simulated colour flow consistent with the Standard Model (SM) expectation as well as events which contain hypothetically colour-charged W bosons to investigate and illustrate the capabilities of the method.

An observable predicted to encode colour information about a jet is the jet-pull vector $\vec{\mathcal{P}}$ [8], a p_{T}-weighted radial moment of the jet. For a given jet J with transverse momentum p_{T}^{J}, the observable is defined as

$$\vec{\mathcal{P}}\,(J) = \sum_{i \in J} \frac{\left|\vec{\Delta r}_i\right| \cdot p_{\mathrm{T}}^{i}}{p_{\mathrm{T}}^{J}} \vec{\Delta r}_i\,, \tag{1}$$

where the summation runs over the constituents of J which have transverse momentum p_{T}^{i} and are located at $\vec{\Delta r}_i = (\Delta y_i, \Delta\phi_i)$, which is the offset of the constituent from the jet axis (y_J, ϕ_J) in rapidity (y) - azimuth (ϕ) - space.[1] Examples of constituents that could be used in Eq. 1 include calorimeter clusters, inner

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Using these coordinates, the radial distance ΔR between two objects is thus defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ where $\Delta\eta$ and $\Delta\phi$ are the differences in pseudorapidity and azimuthal angle between the two objects, respectively.

detector tracks, and simulated stable particles.

Given two jets J_1 and J_2, the jet-pull vector, $\vec{\mathcal{P}}\,(J_1)$, can be used to construct another observable, the jet-pull angle $\theta_{\mathcal{P}}\,(J_1, J_2)$, which relates the local colour structure information of J_1 to the global superstructure of the two jets. This is defined as the angle between the pull vector $\vec{\mathcal{P}}\,(J_1)$ and the vector connecting J_1 to another jet J_2 in rapidity - azimuth - space, $(y_{J_2} - y_{J_1}, \phi_{J_2} - \phi_{J_1})$. Figure 1 illustrates the jet-pull vector and angle for an idealised di-jet system. As the jet-pull angle is symmetric around zero and takes values ranging from $-\pi$ to π, it is convenient to consider the normalised absolute pull angle $|\theta_{\mathcal{P}}|\,/\pi$ instead as observable. The measurement presented here is performed using this normalisation.

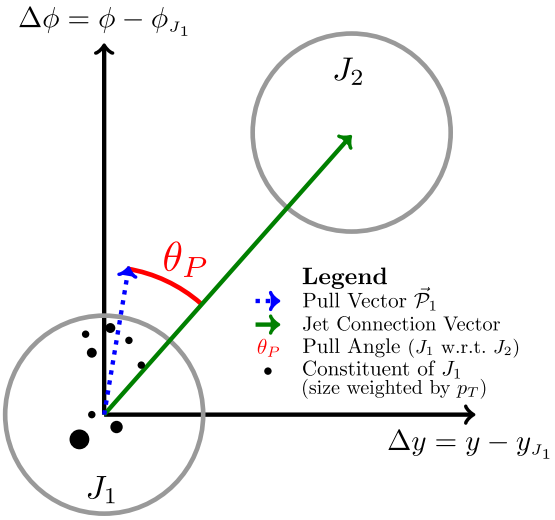


Figure 1: Illustration of jet-pull observables for a di-jet system. For a jet J_1 the jet-pull vector (blue dashed) is calculated using an appropriate set of constituents (tracks, calorimeter clusters, truth particles, ...). The variable of particular sensitivity to the colour structure of J_1 with respect to J_2 is the jet-pull angle (red) which is the angle between the pull vector for J_1 and the vector connecting J_1 to another jet J_2 in localised y-ϕ-space (green).

The jet-pull angle is particularly suited for studying the colour structure of an object decaying into a di-jet system as the inputs into the calculation are well-defined and the observable is expected to be sensitive to the presence or absence of a colour connection.

For such a system of two colour-connected jets, it is expected that $\vec{\mathcal{P}}$ is aligned with the jet connection axis, i.e $\theta_{\mathcal{P}} \sim 0$. Experimentally, the observable is smeared out but should exhibit a peak at small values and a sloped reduction from there on. If $\theta_{\mathcal{P}}$ is calculated for two jets without any particular colour connection, the jet-pull vector and the connection axis are not expected to be aligned and thus $\theta_{\mathcal{P}}$ is expected to be distributed uniformly.

In this note, the normalised jet-pull angle is measured for two different systems of di-jets in $t\bar{t}$ events using 36.1 fb^{-1} of pp collision data recorded by the ATLAS detector at $\sqrt{s}$ = 13 TeV. The first targets the jets originating from the hadronic decay of a W boson and thus from a colour singlet, while the second targets the two b-jets from the top decays, which are not expected to be colour connected. Furthermore, the magnitude of the jet-pull vector is measured. The results are presented as normalised distributions corrected for detector effects.

In Section 2, the ATLAS detector is introduced. Section 3 discusses the data used by this analysis as well as the simulation samples. The reconstruction procedures and event selection are presented in Section 4.

In Section 5 the analysis observables are introduced and discussed in detail. Section 6 introduces the phase-space of the particle-level measurement and the unfolding procedure used to correct the observed data for detector effects. In Section 7 the relevant systematic uncertainties are discussed. Finally, Section 8 presents the results followed by a conclusion in Section 9.

2 The ATLAS Detector

The ATLAS detector [9] uses a system of tracking detectors, which directly enclose the interaction point, to provide highly resolved spacial measurements of charged particles covering a range of $|\eta| < 2.5$. These tracking detectors, collectively called the inner detector, are immersed in a 2 T magnetic field enabling curvature-based track momentum reconstruction. During the Long Shutdown 1, a new innermost layer of the pixel detector has been inserted into the detector, the insertable B-layer (IBL) [10, 11]. Two calorimeter subsystems enclose the inner detector allowing complementary calorimetric measurements of both charged and neutral particles traversing the detector. The calorimeters are enclosed by the muon chambers which provide muon identification, triggering, and (additional) tracking. A more complete description of the ATLAS detector can be found elsewhere [9].

Data are selected for readout and further processing using a two-stage triggering procedure [12] which sequences a hardware-based high-performance trigger using coarser detector information with a software-based second trigger stage which has access to the full detector granularity. This reduces the incoming data from a rate of 40 MHz enforced by the LHC pp collision rate to about 75 kHz after the first stage and 1 kHz after the second stage.

3 Data Sample and Simulation

The data used by this analysis were collected in 2015 and 2016 during pp runs provided by the LHC at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. Stable beams and fully operational subdetectors are required. After data quality requirements, the data correspond to an integrated luminosity of $\mathcal{L}_{\text{Int}} = 36.1\,\text{fb}^{-1}$.

Monte-Carlo (MC) samples are used to evaluate the contribution of background processes to the event selection and evaluate how the detector response affects the analysis observables. A variety of configurations are investigated for different purposes. Table 1 summarises the samples used by the analysis. The $t\bar{t}$ sample in the first row of the Table (the *"nominal"* sample) is used to evaluate the data to MC agreement, predict the number of signal events, and obtain the nominal detector response description. This sample is generated using the Powheg-Box `v2` [13–15] event generator with the top quark mass, m_t, set to 172.5 GeV and the value of the h_{damp} parameter, which controls the p_{T} of the first emission beyond the Born configuration in Powheg, is set to 1.5 m_t. The main effect of this is to regulate the high-p_{T} emission against which the $t\bar{t}$ system recoils. Pythia 8 is used to simulate the parton shower, hadronisation and underlying event. To evaluate the impact of systematic uncertainties coming from signal modelling on the measurements, a variety of alternative signal MC samples are used. These samples are marked with a † in Table 1.

Process	Generator	Type	Version	PDF	Tune[2]
$t\bar{t}$	Powheg-Box v2 [13–15]	NLO ME	r3026	NNPDF 3.0 [16]	–
	+Pythia 8 [17]	+LO PS	v8.186	NNPDF 2.3 [18]	A14 / A14.v1† / A14.v3c† [19]
Single top	Powheg-Box v1	NLO ME	r2819	CT10 [20]	–
	+Pythia 6 [21]	+LO PS	v6.425	CTEQ6L1 [22]	Perugia 2012C [23]
WW, WZ, ZZ	Sherpa [24–26]	LO/NLO multileg ME+PS	v2.1.1	CT10	Default
W/Z + jets	Sherpa	LO/NLO multileg ME+PS	v2.2.1	NNPDF 3.0	Default
$t\bar{t}W/Z$	MG5_aMC@NLO [27]	NLO ME	v2.3.3	NNPDF 3.0	–
	+Pythia 8	+LO PS	v8.210	NNPDF 2.3	A14
$t\bar{t}H$	MG5_aMC@NLO	NLO ME	v2.2.3.p4	NNPDF 3.0	–
	+Pythia 8	+LO PS	v8.210	NNPDF 2.3	A14
$t\bar{t}^{\dagger}$	Powheg-Box v2	NLO ME	r3026	NNPDF 3.0	–
	+Herwig 7 [28]	+LO PS	v7.0.1.a	MMHT 2014 [29]	H7UE
$t\bar{t}^{\dagger}$	MG5_aMC@NLO	NLO ME	v2.3.3.p1	NNPDF 3.0	–
	+Pythia 8	+LO PS	v8.112	NNPDF 2.3	A14
$t\bar{t}^{\star}$	Powheg-Box v2	NLO ME	r2819	CT10	–
	+Pythia 6	+LO PS	v6.428	CTEQ6L1	Perugia 2012
$t\bar{t}^{\star}$	Sherpa	LO/NLO multileg ME+PS	v2.2.1	NNPDF 3.0 NNLO	–

Table 1: Monte Carlo samples used for this analysis. Samples / tunes marked with † refer to alternative signal MC samples used to evaluate signal modelling uncertainties, those marked with ⋆ are used for comparison to the measurement result. The following abbreviations are used: *ME* – matrix element, *PS* – parton shower, *LO* – leading-order calculation in QCD, *NLO* – next-to-leading order calculation in QCD, *PDF* – parton distribution function

To assess the impact of increased or reduced radiation, samples are generated using the A14.v3c up and down tune variations.[2] Additionally, in the A14.v3c up (down) variation sample the renormalisation and factorisation scales are scaled by a factor of 0.5 (2) with respect to the nominal sample and the value of h_{damp} is set to $3m_t$ ($1.5m_t$) [30]. Multiple overlaid proton-proton collisions are simulated with the soft QCD processes of Pythia 8.186 [17] using tune A2 [31] and the MSTW2008LO Parton Distribution Function (PDF) set [32].

Events generated by the MC programs are further processed using the ATLAS detector and trigger simulation [33] which uses Geant4 [34] to simulate the interactions between particles and the detector material.

The samples used to evaluate the detector response and estimate the background contributions are processed using the full ATLAS simulation [33]. Alternative signal MC samples which are used to evaluate signal modelling uncertainties are processed using Atlfast II (AFII) [35]. This detector simulation uses a faster energy deposition modelling method for the simulation of the calorimeter than the full ATLAS detector simulation, while the simulation of the inner detector remains unchanged.

[2] The term *tune* refers to a specific setting of configurable parameters of the MC generator. A tune variation can be used to assess the effects of MC modelling on an analysis as configured by the tuning parameters.

All background MC samples are normalised to their theoretical cross-sections evaluated to at least next-to-leading order (NLO) precision in QCD [36–44]. Signal MC is normalised to a cross-section of 832 ± 46 pb as calculated with the `Top++ 2.0` program [45] to next-to-next-to-leading order in perturbative QCD, including resummation of next-to-next-to-leading-log soft-gluon terms, assuming a top quark mass of 172.5 GeV [46–51]. Normalised signal MC is only used in order to allow comparison of the observed data to the prediction.

A reweighting procedure is applied on an event-by-event basis to the simulation samples to reflect the distribution of the average number of pp interactions per event (pile-up) observed in data.

In order to evaluate the sensitivity of the analysis observables to colour flow and to be able to assess the colour-model dependence of the analysis methods, a dedicated MC sample with a simulated exotic colour flow model is used; this is labelled as *(colour) flipped*. In this sample, the colour-singlet W boson in ordinary signal events is replaced ad-hoc with a colour octet. The procedure is based on the prescription used to create the nominal signal MC. First, hard-scatter signal events are generated using Powheg-Box `v2` and stored in the LHE format [52]. The colour strings are then flipped in such a way that, among the the decay products obtained from the W boson decay, one of them is connected to the incoming top quark while the other one is connected to the outgoing b quark. Pythia 8 is then used to perform the showering and hadronisation on the modified hard-scatter event using the same procedure as the nominal $t\bar{t}$ sample.

4 Event Reconstruction and Selection

To have access to an event sample that predictably contains a hadronically decaying W boson, this analysis targets the $t\bar{t} \to b\bar{b}W(\to \ell\nu)W(\to q\bar{q}')$ final state where ℓ refers to electrons and muons.[3] Such a sample provides access to both a pair of colour-connected ($q\bar{q}'$) and non-connected ($b\bar{b}$) jets.

In the following, the definitions used for the object reconstruction as well as the event selection used to obtain a signal-enriched sample in data is discussed.

4.1 Detector-Level Objects

Primary vertices are constructed from all reconstructed tracks compatible with the interaction region given by LHC beam-spot characteristics [53]. The hard-scatter primary vertex is then selected as the vertex with the largest $\sum p_{\mathrm{T}}^2$ where tracks entering the summation must satisfy $p_{\mathrm{T}} > 0.4$ GeV.

Candidate electrons are reconstructed by matching tracks from the inner detector to energy deposits in the electromagnetic (EM) calorimeter. Electron identification (ID) relies on a likelihood classifier constructed from various detector inputs such as EM shower shape or track quality [54–56]. The electron candidates must satisfy a "tight" ID criterion as defined in [56]. They must further satisfy $E_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.47$ with the region $1.37 \leq |\eta| \leq 1.52$ being excluded. This so called "crack" region is where two calorimeter components connect and as a result the energy resolution is significantly degraded within the exclusion region. Isolation requirements using calorimeter and tracking requirements are applied to reduce background from non-prompt and fake electrons [57]. The resulting efficiency increases linearly with the electron p_{T} starting at approximately 95 % and reaches a plateau of 100 % at approximately $p_{\mathrm{T}} = 70$ GeV.

[3] Electrons and muons produced via an intermediate tau lepton decay are also accepted.

Electrons are also required to have $|d_0^{\text{Sig}}| < 5$ and $|z_0 \sin\theta| < 0.5\,\text{mm}$, where $|d_0^{\text{Sig}}| = |d_0|/\sigma_{d_0}$ is the significance of the transverse impact parameter with respect to the beamline, and z_0 is the difference between the longitudinal impact parameter of the track and the primary vertex relative to the beamline.

Muon candidates are reconstructed by matching tracks in the muon spectrometer to inner detector tracks. Muons must satisfy a "medium" ID criterion as defined in [58]. The muon p_{T} is determined from a fit of all hits associated with the muon track, also taking into account the energy loss due to traversing the calorimeters. Furthermore, muons must satisfy $p_{\text{T}} > 25\,\text{GeV}$ and $|\eta| < 2.5$. Isolation requirements similar to those used for electrons are applied. The respective efficiencies are the same as for electrons. Finally, muon tracks must have $|d_0^{\text{Sig}}| < 3$ and $|z_0 \sin\theta| < 0.5\,\text{mm}$, where the parameters are as defined above.

Jets are reconstructed using the anti-k_t algorithm [59] with radius parameter $R = 0.4$ as implemented by the `FastJet` [60] package. The inputs to the jet algorithm consist of three-dimensional, massless, positive-energy topological clusters [61, 62] constructed from energy deposited in the calorimeters. The jet four-momentum is calibrated using an η- and energy-dependent scheme with *in situ* corrections based on data [63, 64]. Afterwards, they are required to satisfy $p_{\text{T}} > 25\,\text{GeV}$ and $|\eta| < 2.5$. To reduce the number of jets originating from pile-up, an additional selection criterion is applied to jets with $p_{\text{T}} < 60\,\text{GeV}$ and $|\eta| < 2.4$ based on a jet-vertex tagging (JVT) technique [65]. A multivariate discriminant is used to identify jets containing b-hadrons, using track impact parameters, track invariant mass, track multiplicity and secondary vertex information. The b-tagging algorithm [66, 67] is used at a working point which is constructed to operate at an overall b-tagging efficiency of 70 % in simulated $t\bar{t}$ events for jets with $p_{\text{T}} > 20\,\text{GeV}$. The corresponding c-jet and light-jet rejection factors are 12 and 381 respectively, resulting in a purity of 97 %.

Detector information may produce objects which satisfy both the jet and lepton object definition criteria. In order to associate the detector information to a unique physics object hypothesis, an overlap removal procedure is applied: double-counting of electron energy deposits as jets is prevented by discarding the closest jet lying within $\Delta R < 0.2$ to a reconstructed electron. Subsequently, if an electron is within $\Delta R < 0.4$ to a jet, the electron is discarded in order to reduce the impact of non-prompt leptons. Furthermore, if a jet has fewer than three associated tracks and is within $\Delta R < 0.4$ to a muon, the jet is discarded. Conversely, any muon that is within $\Delta R < 0.4$ to a jet with at least three associated tracks is discarded.

The magnitude of the missing transverse momentum $E_{\text{T}}^{\text{miss}}$ is calculated as the transverse component of the negative vector sum of the calibrated momentum of all objects in the event [68, 69]. This sum includes contributions from tracks which have not been associated with any of the physics objects discussed above.

4.2 Event Selection

At first, basic event level quality criteria are applied such as the presence of a primary vertex and the requirement of stable detector conditions. Then, events are selected by requiring that a single electron or muon trigger has fired. The triggers are designed to select well-identified charged leptons with high p_{T}. They require a p_{T} of at least 20 (26) GeV for muons and 24 (26) GeV for electrons for the 2015 (2016) data set and also include requirements on the lepton quality and isolation. These triggers are complemented by triggers with higher p_{T} requirements but loosened isolation and identification requirements to ensure maximum efficiencies at higher lepton p_{T}.

The reconstructed lepton must satisfy $p_{\mathrm{T}} > 27$ GeV and must match the trigger-level object that fired using a geometric matching. No additional lepton may be present. Furthermore, selected events must contain at least four jets. At least two of the jets in the event must be b-tagged. Finally, $E_{\mathrm{T}}^{\mathrm{miss}}$ must exceed 20 GeV.

4.3 Background Determination

Given the event selection, a variety of potential background sources remain: production of $t\bar{t} + X$ with X being either a W, Z, or Higgs boson is an irreducible background which is, however, expected to be small. Similarly, events which contain only a single top quark may contribute.

Events which contain either two electroweak bosons, or one electroweak boson in association with jets can be misidentified as signal. However, only the W + jets component is expected to contribute significantly. Finally, multijet processes where either a jet is misidentified as a lepton or a semileptonic decay is wrongly reconstructed as an isolated lepton enter the signal selection. This last category is collectively called the non-prompt (NP) and fake lepton background.

All backgrounds are modelled using MC simulation with the exception of the non-prompt and fake lepton background which is estimated using the matrix-method [70, 71]. A sample enriched in fake leptons is obtained by loosening the requirements on the standard lepton selections defined in Section 4.1. The efficiency of these "loose" leptons to pass the standard criteria is then measured separately for prompt and fake leptons. For both electrons and muons the efficiency for a prompt loose lepton to pass the standard criteria is measured using a sample of Z boson decays, while the efficiency for fake loose leptons to pass the standard criteria is measured in events with low missing transverse momentum and high lepton impact parameter significance for electrons and muons respectively. These efficiencies allow the number of fake leptons selected in the signal region to be estimated.

The number of selected events is listed in Table 2. The estimated signal purity is approximately 88 % with the single-top and non-prompt and fake lepton backgrounds being the largest remnant contributors.

Sample	Yield		
$t\bar{t}$	1 026 000	±	95 000
$t\bar{t}V$	3 270	±	250
$t\bar{t}H$	1 700	±	100
Single-top	48 400	±	5 500
Diboson	1 440	±	220
W + jets	27 700	±	4 700
Z + jets	8 300	±	1 400
NP/Fake leptons	53 000	±	30 000
Total Expected	1 170 000	±	100 000
Observed	1 153 003		

Table 2: Event yields after selection. The uncertainties are the sum in quadrature of the detector modelling uncertainties and the uncertainties on the cross-section estimate respectively the modelling uncertainties on the data-driven extraction method (non-prompt and fake lepton background). Details of the uncertainties considered can be found in Section 7.

5 Observable Definition and Reconstruction

The jet-pull vector is calculated from inner-detector tracks created using an updated reconstruction algorithm [72] which makes use of the newly introduced IBL [10] and a neural network based clustering algorithm [73, 74] to improve pixel cluster position resolution and the efficiency of reconstructing tracks in jets. A measurement based on the calorimeter clusters of the jet is not considered in this analysis as it suffers from a significantly degraded spatial resolution and is thus not as accurate for a precision measurement as was shown in Ref. [7].

To ensure good quality, reconstructed tracks must satisfy $|\eta| < 2.5$ and $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and further quality cuts are applied to ensure that they originate from and are assigned to the primary vertex [74]. This suppresses contributions from pile-up and tracks with a poor fit that are due to contributions from multiple charged particles. Matching of tracks to jets is performed using a technique called ghost association [75], in which inner detector tracks are included in the jet clustering procedure after having scaled their four-momentum to have infinitesimal magnitude. As a result, the tracks have no effect on the jet clustering result whilst being clustered into the jet that most naturally encloses them according to the jet algorithm used. The jets used in each observable calculation are required to satisfy $|\eta| < 2.1$ so that all associated tracks are within the coverage of the inner detector. Furthermore, at least two tracks must contribute to the pull-vector calculation.

The jet centre axis used in Eq. 1 is recalculated using the ghost-associated tracks rather than using the jet axis which is calculated from the calorimeter clusters that form the jet. This ensures proper correspondence between the pull vector and the components entering the calculation. For consistency, the jet p_{T} in Eq. 1 is also taken from the recalculated jet axis.

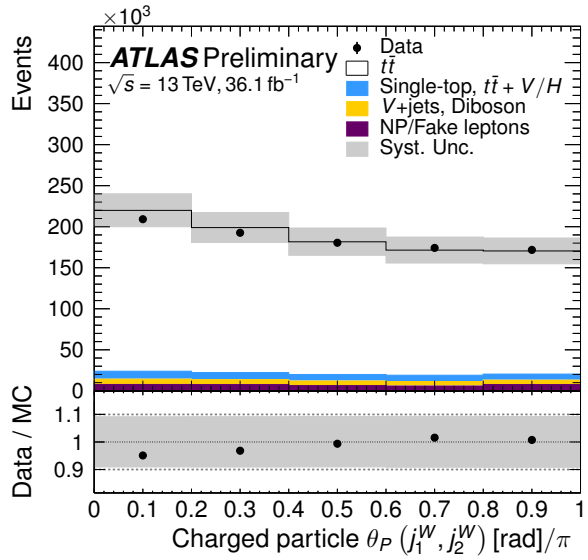
The analysis presented in this note measures the colour flow for two scenarios. The signal colour flow is extracted from an explicitly colour-connected di-jet system while the spurious colour flow is obtained from a jet pair for which no specific colour connection is expected. Table 3 summarises the analysis observables and their definition.

Target colour flow	**Signal colour flow** (j_1 and j_2 are colour connected)	**Spurious colour flow** (j_1 and j_2 are not colour connected)
Jet assignment	j_1^W : leading p_{T} non-b-tagged jet j_2^W : 2$^{\mathrm{nd}}$ leading p_{T} non-b-tagged jet	j_1^b : leading p_{T} b-tagged jet j_2^b : 2$^{\mathrm{nd}}$ leading p_{T} b-tagged jet
Observables	$\theta_{\mathcal{P}}\left(j_1^W, j_2^W\right)$: *"forward pull-angle"* $\theta_{\mathcal{P}}\left(j_2^W, j_1^W\right)$: *"backward pull-angle"* $\lvert\vec{\mathcal{P}}\left(j_1^W\right)\rvert$: *"pull-vector magnitude"*	$\theta_{\mathcal{P}}\left(j_1^b, j_2^b\right)$: *"forward di-b-jet-pull angle"*

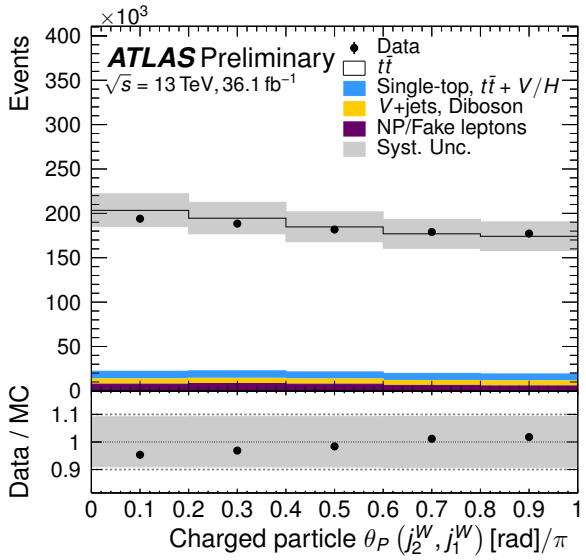
Table 3: Summary of the observable definitions.

Study of the signal colour flow is possible using the daughters of the hadronically decaying W boson from the top-quark decay. In practice, the two highest-p_{T} jets which have not been b-tagged are selected as W boson daughter candidates. A dedicated study using simulated $t\bar{t}$ events has shown that this definition achieves correct matching of both jets in about 30 % of all events with roughly 50 % of all cases having a correct match to one of the two jets.

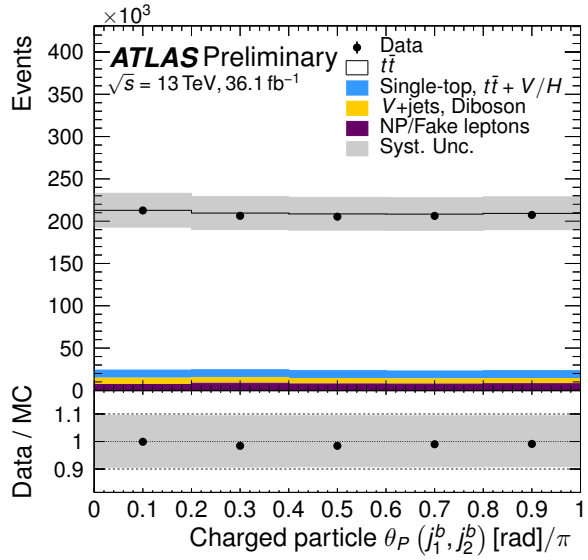
Given the above procedure for selecting the two jets from the hadronically decaying W boson, the jets are labelled as j_1^W and j_2^W with the indices referring to their p_{T} ordering. This allows calculation of two jet pull angles: $\theta_{\mathcal{P}}\left(j_1^W, j_2^W\right)$ and $\theta_{\mathcal{P}}\left(j_2^W, j_1^W\right)$ which are labelled as "forward pull angle" and "backward pull angle" respectively. Although the two observables probe the same colour structure, in practice the two values obtained for a single event have a linear correlation of less than 1 % in data and can be used for two practically orthogonal measurements. Figures 2(a) and 2(b) compare the distributions observed for these two pull angles to those predicted by simulation at detector level.



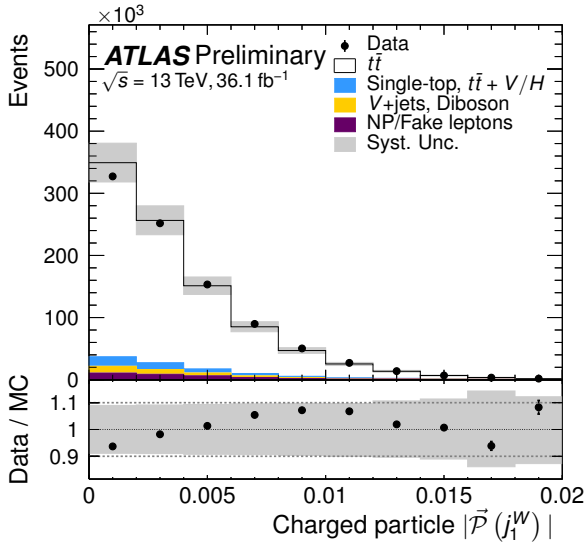
(a) $\theta_{\mathcal{P}}\left(j_1^W, j_2^W\right)$



(b) $\theta_{\mathcal{P}}\left(j_2^W, j_1^W\right)$



(c) $\theta_{\mathcal{P}}\left(j_1^b, j_2^b\right)$



(d) $|\vec{\mathcal{P}}\left(j_1^W\right)|$

Figure 2: Detector-level distributions for the four considered observables: the forward (a) and backward (b) pull angle for the hadronically decaying W boson daughters, the forward di-b-jet-pull angle (c), and the leading W daughter jet-pull vector magnitude (d). Uncertainty bands shown include the experimental uncertainties to the event selection and observable calculation. Details of the uncertainties considered can be found in Section 7.

In addition, the jet-pull-vector magnitude is calculated for the jet with larger transverse momentum: $|\vec{\mathcal{P}}\left(j_1^W\right)|$. A comparison of the observed and predicted distributions for this observable can be found in Figure 2(d) which shows a steeply falling distribution with the bulk of its content being contained in the range below 0.005.

In $t\bar{t}$ events an obvious candidate for measuring spurious colour flow is the structure observed between the two leading b-tagged jets as the hard objects that initiate the b-jets are not expected to have any specific colour connection. For a typical signal event, their colour charge can be traced to the gluon which splits into the $t\bar{t}$ pair. This coloured initial state ensures that the two b quarks are not expected to be colour connected. Therefore, the forward di-b-jet-pull angle is calculated from the two leading p_{T} b-tagged jets: $\theta_{\mathcal{P}}\left(j_1^b, j_2^b\right)$. This definition achieves correct matching on both jets in about 80 % of all events. Figure 2(c) shows a comparison of the distribution observed in data to that predicted by simulation for this observable. Consistent with the expectation, the distribution is flat, unlike in the case of the jet pairs from W boson decays.

6 Unfolding

Particle-level objects are defined for simulated events using definitions which are analogues to those used at detector level as discussed in the previous section. Particle-level objects are defined using particles with mean lifetime greater than 30 ps.

Electrons and muons must not originate from a hadron in the MC truth record, be it either directly or through an intermediate τ decay. Fundamentally, this is equivalent to requiring that the lepton originate from a real W or Z boson. The lepton four-momentum is modified by adding all photons not originating from a hadron within $\Delta R < 0.1$ around the lepton to the original four-momentum to take into account final-state photon radiation. Leptons are then required to satisfy $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$.

Jets are constructed by clustering all stable particles, excluding leptons not from hadron decays and their radiated photons, using the same clustering algorithm and configuration as is used for the detector-level jets. Particle-level jets are furthermore required to satisfy $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$. Classification of jets as having originated from a b-hadron is performed using ghost association [75]. The technique is equivalent to the method used for associating tracks to jets described in Section 5 except that it is applied during particle-level jet clustering and adds ghosts for unstable b-hadrons rather than inner detector tracks. A particle-level jet is considered to be b-tagged if it contains at least one such b-hadron.

An overlap removal procedure is applied which rejects leptons that overlap geometrically with a jet at $\Delta R < 0.4$.

The magnitude of the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ at particle-level is calculated as the transverse component of the four-momentum sum of all neutrinos in the event excluding those from hadron decays, be it either directly or through an intermediate τ decay.

To select events at particle level, the event selection requires exactly one lepton with $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ with no additional lepton, at least four jets of which at least two are b-tagged, as well as $E_{\mathrm{T}}^{\mathrm{miss}} > 20\,\mathrm{GeV}$.

At particle-level, the input to the calculation of the jet-pull vector is simply the collection of jet constituents, as defined by the clustering procedure described in Section 4.1, filtered such that only particles with non-zero electric charge are included. Apart from the inputs to the jet-pull-vector calculation, the procedure applied at detector level is mirrored exactly at particle level.

The signal distributions are unfolded using the Iterative Bayesian (IB) method [76] as implemented by the `RooUnfold` framework [77]. This algorithm iteratively corrects the observed data to an unfolded particle-level distribution given a certain particle-level prior. Initially, this prior is taken to be the particle-level distribution obtained from simulation. However, it is updated after each iteration with the observed posterior distribution. Thus, the algorithm converges onto an unfolding result driven by the observed distribution.

The unfolding procedure essentially consists of two stages: first the background contributions are subtracted bin-by-bin from the observed data. Afterwards, detector effects are unfolded from the signal distribution using a detector response model, the migration matrix, obtained from simulated $t\bar{t}$ events. As part of the second step two correction factors are applied which correct for non-overlap in the fiducial phase-space at reconstructed- and particle-level. The corrections account for events which fall within the fiducial phase-space at one level but not the other. The full unfolding procedure for an observable X can be symbolically summarised by the equation

$$\frac{\mathrm{d}\sigma^t_{\mathrm{Fid}}}{\mathrm{d}X^t} = \frac{1}{\mathcal{L}\cdot\Delta X^t}\cdot\frac{1}{\epsilon^t}\sum_r \mathcal{M}^{-1}_{r,t}\cdot\epsilon^r_{\mathrm{Fid}}\cdot\left(N^r_{\mathrm{Obs}} - N^r_{\mathrm{Bkg}}\right), \tag{2}$$

where t indicates the particle-level bin index, r the detector-level bin index, $\mathcal{L}$ is the integrated luminosity of the data, $\mathcal{M}_{r,t}$ is the migration matrix and the inversion symbolises unfolding using the IB method, N^t_{Obs} is the number of observed, N_{Bkg} the number of expected background events, and ϵ^t and $\epsilon^r_{\mathrm{Fid}}$ are the phase-space correction factors. These last two parameters are defined as

$$\epsilon^t = \frac{N_{\mathrm{Tru}\wedge\mathrm{Rec}}}{N_{\mathrm{Tru}}} \qquad\qquad \epsilon^r_{\mathrm{Fid}} = \frac{N_{\mathrm{Tru}\wedge\mathrm{Rec}}}{N_{\mathrm{Rec}}}\,. \tag{3}$$

The number N_{Tru} (N_{Rec}) indicates the number of events fulfilling the fiducial requirements on particle-level (selection requirements at detector-level), $N_{\mathrm{Tru}\wedge\mathrm{Rec}}$ is the number of events that pass both sets of requirements at their respective level.

The response model and phase-space correction factors are obtained from $t\bar{t}$ simulation.

Some of the background samples considered by this analysis potentially contain true signal colour flow, e.g. the single-top or $t\bar{t}+X$ contributions. However, as their overall contributions are very small, even extreme changes in their respective colour flow are negligible. Therefore, all such contributions are ignored and the estimated backgrounds are simply subtracted from the data.

The binning chosen for the observables is determined by optimisation studies performed with simulated samples. A good binning choice should result in a mostly diagonal migration matrix with bin widths appropriate given the observed resolution. The optimisation therefore imposes a requirement of having at least 50 % of events on-diagonal for each truth-bin of the migration matrix.

The number of iterations used by the unfolding method is chosen such that the total uncertainty composed of the statistical uncertainty and the bias is minimised.

7 Systematic Uncertainties

A variety of systematic uncertainties affect the measurements discussed above. The different sources are grouped into four categories, experimental uncertainties, uncertainties related to the modelling of the background prediction, uncertainties related to the modelling of the signal process, and an uncertainty relating to the unfolding procedure.

The deviations observed from the systematic variations are used to calculate a covariance matrix for each source individually. This covariance matrix combines the deviations from all measured observables simultaneously and therefore also includes the cross-correlations between observables. The total covariance matrix is then calculated by summation over the covariances obtained from all systematic sources. The deviations observed for a systematic source are symmetrised prior to calculating the covariance. For one-sided variations, the deviation is taken as a symmetric uncertainty. For two-sided variations, which variation is used to infer the sign is completely arbitrary as long as it is done consistently. In this analysis, the sign is taken from the up variation while the value is taken as the maximum deviation. Furthermore, it is assumed that all uncertainties, including modelling uncertainties, are gaussian distributed.

7.1 Experimental Uncertainties

Systematic uncertainties due to the modelling of the detector response and other experimental sources affect the signal reconstruction efficiency, the unfolding procedure, and the background estimate. Each source of experimental uncertainty is treated individually by repeating the full unfolding procedure using inputs which have been varied within the detector response associated with the systematic source. The unfolding result is then compared to the nominal result and the difference is taken as systematic uncertainty.

Uncertainties due to lepton identification, isolation, reconstruction, and trigger requirements are evaluated by scaling the efficiencies and kinematics calibrations in simulation by uncertainties derived from data in control regions enriched in $Z \to \ell\ell$ events [58, 78–80].

The uncertainties due to the jet energy scale (JES) and resolution (JER) are derived using a combination of simulation, test-beam data, and *in situ* measurements [63, 81–84]. In addition, contributions from η-intercalibration, single-particle response, pile-up, jet flavour composition, punch-through, varying calorimeter response to different jet flavours, and differences between the fast and full detector simulation are taken into account. This results in a scheme of 21 systematic variations.

Efficiencies relating to the performance of the b-tagging procedure are corrected in simulation to account for differences between data and simulation. The corresponding scale factors are extracted from simulated $t\bar{t}$ events. This is done separately for b-, c-, and light jets, hereby accounting for mis-tags. Uncertainties relating to this are propagated by varying the scale factors within their uncertainty [66, 85, 86].

The uncertainties on the $E_{\mathrm{T}}^{\mathrm{miss}}$ due to systematic shifts in the corrections for leptons and jets are accounted for in a fully correlated way in their evaluation for those physics objects. Uncertainties due to track-based terms in the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation, i.e. those that are not associated with any other reconstructed object, are treated separately [87].

All uncertainties on the reconstructed tracks directly enter the observable calculation as defined in Eq. 1. Uncertainties are either expressed as change in the tracking efficiency or smearing of the track momentum [72, 74]. This also includes effects due to fake tracks and lost tracks in the core of jets.

Corrections and scale factors have been extracted using simulated data as well as experimental data obtained from minimum bias and $Z \to \mu\mu$ selections. The effect of the individual systematic uncertainty is in most cases parameterised in the track p_{T} and η.

The uncertainty on the combined 2015 and 2016 integrated luminosity is 2.1 % which is derived following a methodology similar to that detailed in Ref. [88], from a calibration of the luminosity scale using x-y beam-separation scans performed in August 2015 and May 2016. It affects the scaling of the background prediction which is subtracted from the observed data. The uncertainty relating to the pile-up reweighting is evaluated by varying the scale factors by their uncertainty based on the reweighting of the average number of interactions per pp collision.

The data statistical uncertainty and bin-to-bin correlations are evaluated using the bootstrap method [89]. Bootstrap replicas of the measured data are propagated through the unfolding procedure and their variance is used to assess the statistical uncertainty. These replicas can also be used to calculate the statistical component of the covariance of the measurement as well as the statistical bin-by-bin correlations of the pre- or post-unfolding distributions.

7.2 Background Modelling Uncertainties

Systematic uncertainties related to the background modelling affect the number of background events subtracted from data prior to the unfolding.

The normalisation of the background contributions obtained from MC simulation is varied within the uncertainties obtained from the corresponding cross-section calculation. For the single-top background, the normalisation uncertainty ranges from 3.6 to 5.3 % [37–39], and for the $t\bar{t}Z$ and $t\bar{t}W$ backgrounds it is 12 and 13 % respectively [42, 90]. In the case of the W/Z + Jets backgrounds, the uncertainties include a contribution from the overall cross-section normalisation (4 %) as well as an additional 24 % uncertainty added in quadrature for each jet [91, 92]. For the diboson background, the normalisation uncertainty is 6 % [93]. The uncertainty of the normalisation for the $t\bar{t}H$ background is chosen to be 100 %.

The uncertainty arising from the modelling of the non-prompt and fake lepton background is assessed by varying the normalisation by 50 % as well as by changing the efficiency parameterisation used by the matrix method to obtain a shape uncertainty. These uncertainties were found to adequately cover any disagreement between data and prediction in various background dominated control regions.

7.3 Signal Modelling Uncertainties

The following systematic uncertainties related to the modelling of the $t\bar{t}$ system are considered: the choice of matrix-element generator, the choice of PDF, the hadronisation model, the amount of initial- and final-state radiation (ISR/FSR), and the amount and strength of colour reconnection (CR).

Signal modelling uncertainties are evaluated individually using different signal MC samples. Detector-level distributions from the alternative signal MC are unfolded using the nominal response model. The unfolding result is then compared to the particle-level prediction of the alternative MC and the deviation is used as uncertainty. Table 1 lists the alternative signal MC samples used for assessing the generator, hadronisation, ISR/FSR systematics (`A14.v3c` tune variations), and CR (`A14.v1` tune variations) systematic uncertainties.

The uncertainty arising from the choice of PDF is evaluated by creating reweighted pseudo-samples, in which the weight variations for the PDF sets are according to the PDF4LHC [94] prescription. The unfolding results obtained for the pseudo-samples are then combined in accordance with the PDF4LHC procedure to obtain a single systematic shift.

A summary of the uncertainties affecting $\theta_{\mathcal{P}}\left(j_1^W, j_2^W\right)$ are shown in Table 4. The total uncertainty is dominated by systematic uncertainties, with uncertainties due to $t\bar{t}$ modelling being dominant in most bins. Uncertainties that directly affect the inputs to the pull-vector calculation, such as the JES, JER and track uncertainties are generally sub-dominant.

$\Delta\theta_P\left(j_1^W, j_2^W\right)$ [%]	$\theta_P\left(j_1^W, j_2^W\right)$			
	0.0 – 0.21	0.21 – 0.48	0.48 – 0.78	0.78 – 1.0
Hadronisation	0.63	0.22	0.27	0.09
Generator	0.37	0.24	0.50	0.06
Colour Reconnection	0.11	0.26	0.03	0.53
b-Tagging	0.35	0.12	0.20	0.31
Non-Closure	0.25	0.07	0.08	0.30
ISR / FSR	0.32	0.12	0.15	0.01
Other	0.25	0.20	0.11	0.18
JER	0.12	0.13	0.21	0.03
JES	0.13	0.06	0.13	0.07
Tracks	0.09	0.04	0.05	0.07
Syst.	0.97	0.52	0.68	0.72
Stat.	0.22	0.18	0.17	0.26
Total	0.99	0.55	0.71	0.76

Table 4: Statistical and systematic uncertainties affecting the measurement of $\theta_{\mathcal{P}}\left(j_1^W, j_2^W\right)$. The category "Other" summarises various smaller uncertainty components. Uncertainties are ordered by the mean value of the uncertainty across all bins and are expressed in percent of the measured value.

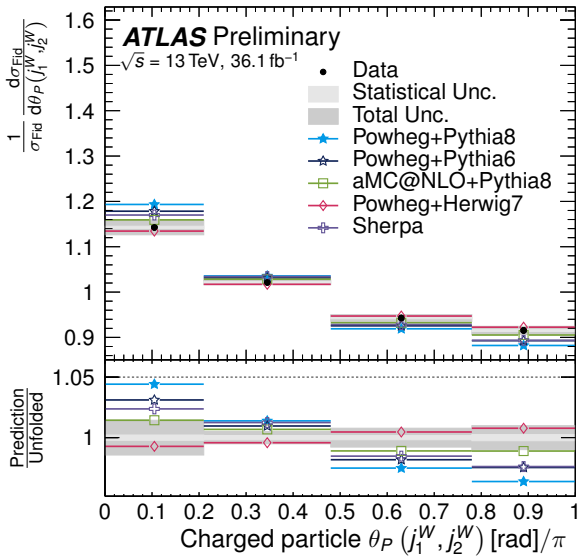
7.4 Unfolding Procedure Systematic

The uncertainty arising from the unfolding procedure, also called non-closure uncertainty, is assessed using a data-driven approach. For each measured distribution, particle-level events are reweighted using a linear weight function such that the corresponding detector-level distributions are in better agreement with the data. The weights are propagated to the correspond detector-level events and the resulting distributions are unfolded using the nominal detector response model. Deviations of these unfolded distributions from the reweighted particle-level distributions are then assigned as non-closure uncertainty.

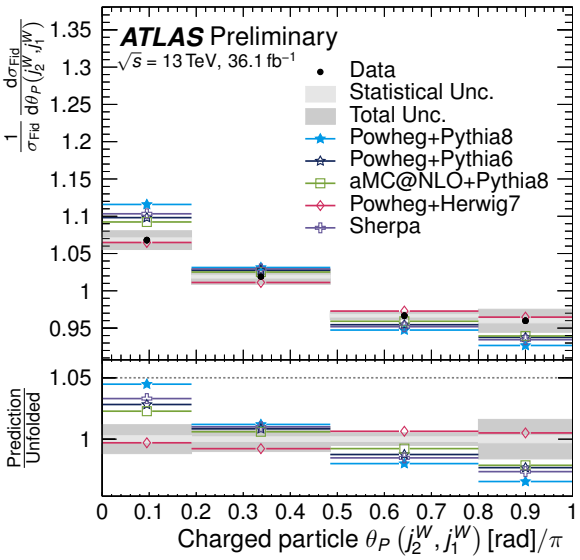
8 Results

Figure 3 compares the unfolded data to several SM predictions for all four observables. Three SM predictions use POWHEG to generate the hard-scatter events and then differ for the subsequent hadronisation,

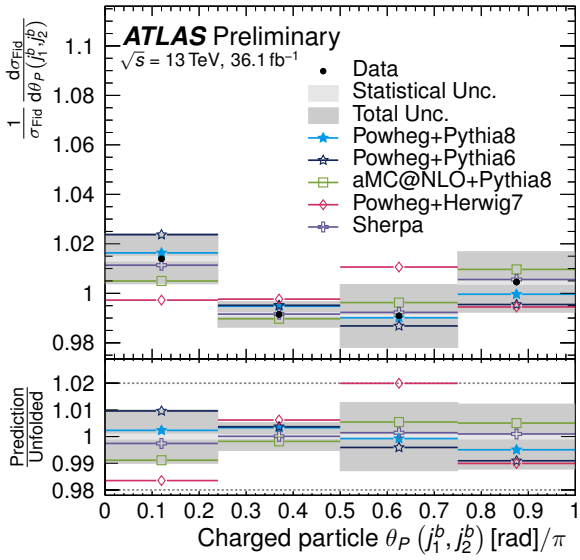
namely Pythia 6, Pythia 8, and Herwig 7. One main difference between these predictions is that the Pythia family uses a fundamentally different hadronisation model than the Herwig family. One SM prediction uses MadGraph5_aMC@NLO to produce the hard-scatter event, the hadronisation is then performed using Pythia 8. Finally, one SM prediction is obtained from events generated with Sherpa.



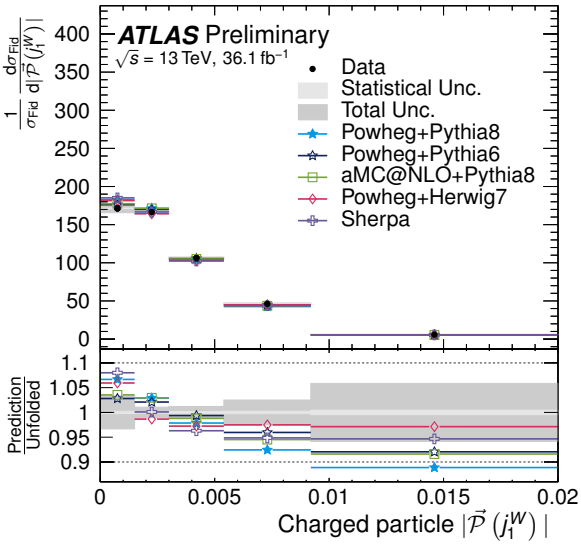
(a) $\theta_{\mathcal{P}}\left(j_1^W, j_2^W\right)$



(b) $\theta_{\mathcal{P}}\left(j_2^W, j_1^W\right)$



(c) $\theta_{\mathcal{P}}\left(j_1^b, j_2^b\right)$

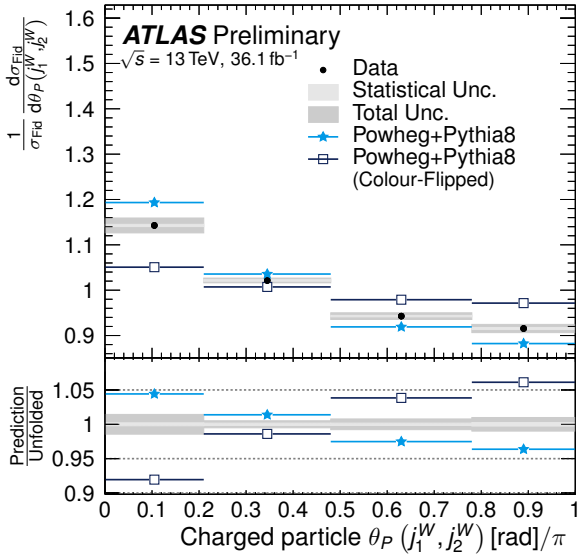


(d) $|\vec{\mathcal{P}}\left(j_1^W\right)|$

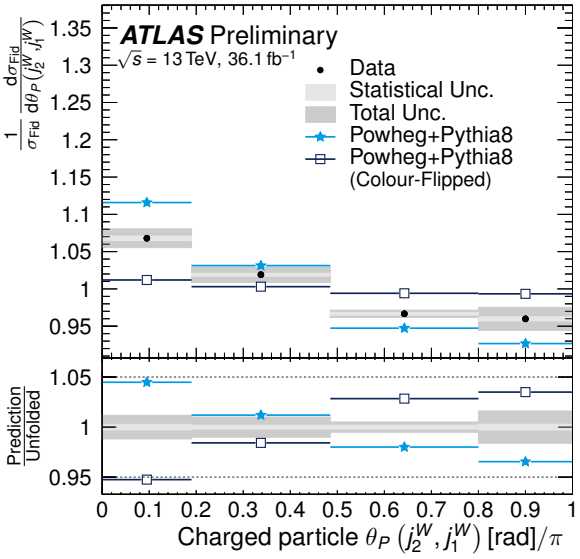
Figure 3: Normalised fiducial differential cross-sections as a function of the forward (a) and backward (b) pull angle for the hadronically decaying W boson daughters, the forward di-b-jet-pull angle (c), and the leading W daughter jet-pull-vector magnitude (d). The data are compared to various SM predictions. The statistical uncertainties on the predictions are smaller than the marker size.

Figure 4 compares the unfolded data to the SM prediction as well as a prediction obtained from the exotic model with flipped colour flow described in Section 3. Both predictions are obtained from MC generated

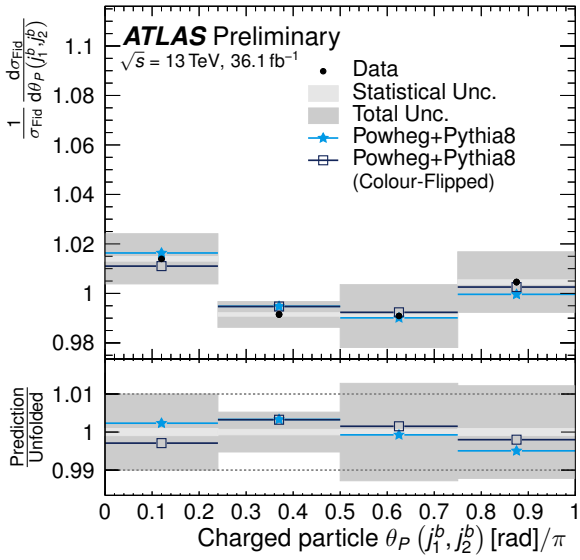
by Powheg + Pythia 8. The data agrees better with the SM prediction than the colour flipped sample.



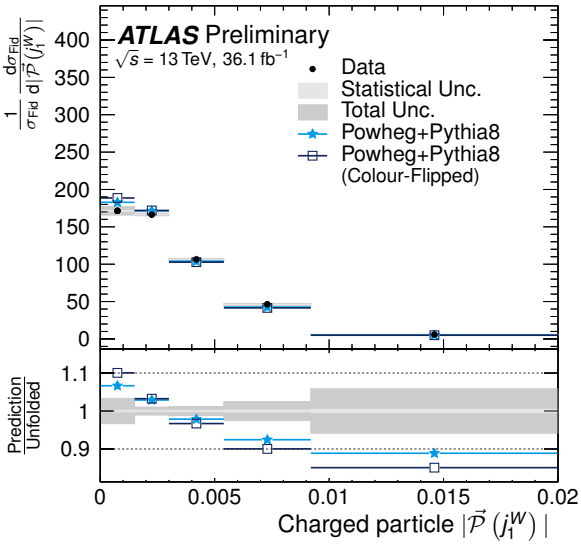
(a) $\theta_{\mathcal{P}}\left(j_1^W, j_2^W\right)$



(b) $\theta_{\mathcal{P}}\left(j_2^W, j_1^W\right)$



(c) $\theta_{\mathcal{P}}\left(j_1^b, j_2^b\right)$



(d) $|\vec{\mathcal{P}}\left(j_1^W\right)|$

Figure 4: Normalised fiducial differential cross-sections as a function of the forward (a) and backward (b) pull angle for the hadronically decaying W boson daughters, the forward di-b-jet-pull angle (c), and the leading W daughter jet-pull-vector magnitude (d). The data are compared to a Standard Model prediction produced by Powheg + Pythia 8 as well as the model with exotic colour flow also created with Powheg + Pythia 8. The uncertainty bands presented on these plots combine the baseline systematics set with effects due to considering the alternative colour-flipped model as a source of signal modelling uncertainty. The statistical uncertainties on the predictions are smaller than the marker size.

The uncertainty bands on the unfolding results shown in Figure 4 have been inflated by a "colour model uncertainty". This uncertainty is obtained using the same procedure as used for the signal modelling uncertainties with the sample with exotic colour flow as alternative $t\bar{t}$ MC and has a similar size as the

dominant signal modelling uncertainties.

A goodness-of-fit procedure is employed in order to quantify the level of agreement between the measured distributions and those predicted by the MC generators. A χ^2 test statistic is calculated for each pairing of observable and theoretical prediction individually using the full covariance matrix of the experimental uncertainties but excluding any uncertainties on the theoretical predictions. Given the unfolded data D, the model prediction M, and the covariance Σ, the χ^2 is given by

$$\chi^2 = (D^T - M^T) \cdot \Sigma^{-1} \cdot (D - M) \,. \tag{4}$$

Subsequently, p-values can be calculated from the χ^2 and number of degrees of freedom (NDF) which describe a probability to obtain a χ^2 value that is larger than or equal to the observed value.

The fact that the analysis measures normalised distributions removes one degree of freedom from the χ^2 calculation. Consequently, one of the N elements of D and M is dropped and the covariance is reduced from dimensionality $N \times N$ to $(N-1) \times (N-1)$ by discarding one column and row. The χ^2 value does not depend on the choice of discarded elements. Table 5 lists the resulting χ^2 values and derived p-values.

Sample	$\theta_{\mathcal{P}}\left(j_1^W, j_2^W\right)$		$\theta_{\mathcal{P}}\left(j_2^W, j_1^W\right)$		$\theta_{\mathcal{P}}\left(j_1^b, j_2^b\right)$		$\lvert\vec{\mathcal{P}}\left(j_1^W\right)\rvert$	
	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value
Powheg+Pythia8	50.7 / 3	< 0.001	20.5 / 3	< 0.001	1.5 / 3	0.690	26.2 / 4	< 0.001
Powheg+Pythia6	24.0 / 3	< 0.001	8.2 / 3	0.041	3.0 / 3	0.385	9.7 / 4	0.045
aMC@NLO+Pythia8	6.2 / 3	0.104	4.6 / 3	0.200	1.9 / 3	0.597	18.4 / 4	0.001
Powheg+Herwig7	2.5 / 3	0.478	2.1 / 3	0.543	5.7 / 3	0.128	10.0 / 4	0.041
Sherpa	23.4 / 3	< 0.001	11.7 / 3	0.008	0.2 / 3	0.974	14.6 / 4	0.006
Powheg+Pythia8$\star$	21.0 / 3	< 0.001	20.5 / 3	< 0.001	0.9 / 3	0.821	13.0 / 4	0.011
Flipped Powheg+Pythia8$\star$	48.1 / 3	< 0.001	35.6 / 3	< 0.001	2.8 / 3	0.427	18.5 / 4	< 0.001

Table 5: The χ^2 and resulting p-values for the measured normalised cross-sections obtained by comparing the different predictions to the unfolded data. When comparing the data with the prediction for the exotic flipped colour-flow model, the model itself is considered as an additional source of signal modelling uncertainty and thus added to the covariance matrix. Calculations which include this additional systematic are marked with $\star$.

For the signal pull-angles $\theta_{\mathcal{P}}\left(j_1^W, j_2^W\right)$ and $\theta_{\mathcal{P}}\left(j_2^W, j_1^W\right)$, the predictions obtained from Powheg + Herwig 7 agree best with the observed data. A general trend is that simulation predicts a more sloped distribution, i.e. a stronger colour-flow effect. The jet-pull-vector magnitude is poorly modelled in general, with the prediction obtained from Powheg + Pythia 6 agreeing best with data. As with the signal pull-angles, the observed mismodelling shows a similar trend for the different MC predictions: data favours larger values of the pull-vector magnitude. Predictions from Powheg + Pythia 6 are in measurably better agreement with the data than those obtained from Powheg + Pythia 8 for the signal pull-angles and jet-pull-vector magnitude. However, it has to be noticed that the choice of hadronisation model is not the only difference between those two samples.

The signal-pull angles and the jet-pull-vector magnitude can be used to distinguish the case of a SM-like colour flow from that of the exotic flipped colour-flow scenario constructed in Section 3. The data favour the SM prediction over the colour-flipped prediction.

The forward di-b-jet-pull angle is modelled relatively well by most predictions. It can be noted that the prediction obtained from Powheg + Herwig 7, which in general shows relatively good modelling for the other three observables, agrees poorly with the distribution obtained from data. Indeed, it is the only prediction that is consistently outside of the estimated uncertainty bands.

9 Conclusion

A measurement of four observables sensitive to the colour flow in $t\bar{t}$ events using $\sqrt{s} = 13$ TeV ATLAS data measured at the LHC has been presented: the forward and backward jet-pull angle for the W boson daughters, the jet-pull angle between the b-tagged jets and the jet-pull magnitude.

The observables sensitive to colour flow discussed in this note are poorly modelled by most MC predictions. The default SM prediction (Powheg + Pythia 8) agrees poorly with the data. However, alternative SM predictions exhibit much better agreement. In particular, the prediction obtained by Powheg + Herwig 7 performs rather well. Predictions from Powheg + Pythia 6 are in measurably better agreement with the data than those obtained from Powheg + Pythia 8.

References

[1] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[2] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31.

[3] JADE Collaboration, W. Bartel et al., *Particle Distribution in Three Jet Events Produced by e+ e- Annihilation*, Z. Phys. **C21** (1983) 37.

[4] L3 Collaboration, P. Achard et al., *Search for color reconnection effects in $e^+e^- \to W^+W^- \to$ hadrons through particle flow studies at LEP*, Phys. Lett. **B561** (2003) 202, arXiv: hep-ex/0303042 [hep-ex].

[5] DELPHI Collaboration, J. Abdallah et al., *Investigation of colour reconnection in WW events with the DELPHI detector at LEP-2*, Eur. Phys. J. **C51** (2007) 249, arXiv: 0704.0597 [hep-ex].

[6] D0 Collaboration, B. Abbott et al., *Evidence of color coherence effects in W + jets events from $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Lett. **B464** (1999) 145, arXiv: hep-ex/9908017 [hep-ex].

[7] ATLAS Collaboration, *Measurement of colour flow with the jet pull angle in $t\bar{t}$ events using the ATLAS detector at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **750** (2015) 475, arXiv: 1506.05629 [hep-ex].

[8] J. Gallicchio and M. D. Schwartz, *Seeing in Color: Jet Superstructure*, Phys. Rev. Lett. **105** (2010) 022001, arXiv: 1001.5027 [hep-ph].

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

[10] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, (2010), URL: https://cds.cern.ch/record/1291633.

[11] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report Addendum*, (2012), Addendum to CERN-LHCC-2010-013, ATLAS-TDR-019, URL: https://cds.cern.ch/record/1451888.

[12] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. **C77** (2017) 317, arXiv: 1611.09661 [hep-ex].

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

[14] 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].

[15] 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].

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

[17] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput.Phys.Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

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

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

[20] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky et al., *New parton distributions for collider physics*, Phys.Rev. **D82** (2010) 074024, arXiv: 1007.2241 [hep-ph].

[21] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv: hep-ph/0603175 [hep-ph].

[22] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, arXiv: hep-ph/0201195 [hep-ph].

[23] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82** (2010) 074018, arXiv: 1005.3457 [hep-ph].

[24] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph].

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

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

[27] 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) 158, arXiv: 1405.0301 [hep-ph].

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

[29] 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. **C75** (2015) 204, arXiv: 1412.3989 [hep-ph].

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

[31] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012, URL: https://cds.cern.ch/record/1474107.

[32] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189, arXiv: 0901.0002 [hep-ph].

[33] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [hep-ex].

[34] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250 , ISSN: 0168-9002.

[35] E. Richter-Was, D Froidevaux and L. Poggioli, *ATLFAST 2.0 a fast simulation package for ATLAS*, (1998), URL: https://cds.cern.ch/record/683751.

[36] 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].

[37] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. **D81** (2010) 054028, arXiv: 1001.5034 [hep-ph].

[38] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. **D82** (2010) 054018, arXiv: 1005.4451 [hep-ph].

[39] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. **D83** (2011) 091503, arXiv: 1103.2792 [hep-ph].

[40] J. M. Campbell and R. K. Ellis, *An Update on vector boson pair production at hadron colliders*, Phys. Rev. **D60** (1999) 113006, arXiv: hep-ph/9905386 [hep-ph].

[41] J. M. Campbell, R. K. Ellis and C. Williams, *Vector boson pair production at the LHC*, JHEP **07** (2011) 018, arXiv: 1105.0020 [hep-ph].

[42] 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].

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

[44] ATLAS Collaboration, *Monte Carlo Generators for the Production of a W or Z/γ^* Boson in Association with Jets at ATLAS in Run 2*, ATL-PHYS-PUB-2016-003, 2016, URL: https://cds.cern.ch/record/2120133.

[45] 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].

[46] 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. **B710** (2012) 612, arXiv: 1111.5869 [hep-ph].

[47] M. Beneke, P. Falgari, S. Klein and C. Schwinn,
Hadronic top-quark pair production with NNLL threshold resummation,
Nucl. Phys. **B855** (2012) 695, arXiv: 1109.1536 [hep-ph].

[48] P. Bärnreuther, M. Czakon and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001,
arXiv: 1204.5201 [hep-ph].

[49] 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].

[50] 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].

[51] 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].

[52] J. Alwall et al., *A Standard format for Les Houches event files*,
Comput. Phys. Commun. **176** (2007) 300, arXiv: hep-ph/0609017 [hep-ph].

[53] 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.

[54] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data*,
Eur. Phys. J. C **74** (2014) 2941, arXiv: 1404.2240 [hep-ex].

[55] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton–proton collision data*, ATLAS-CONF-2014-032, 2014,
URL: https://cds.cern.ch/record/1706245.

[56] ATLAS Collaboration,
Electron identification measurements in ATLAS using $\sqrt{s} = 13$ TeV data with 50 ns bunch spacing,
ATL-PHYS-PUB-2015-041, 2015, URL: https://cds.cern.ch/record/2048202.

[57] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
Phys. Lett. B **761** (2016) 136, arXiv: 1606.02699 [hep-ex].

[58] 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].

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

[60] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, Eur. Phys. J. **C72** (2012) 1896, arXiv: 1111.6097 [hep-ph].

[61] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. **C77** (2017) 490, arXiv: 1603.02934 [hep-ex].

[62] ATLAS Collaboration, *Properties of jets and inputs to jet reconstruction and calibration with the ATLAS detector using proton–proton collisions at* $\sqrt{s}$ = 13 *TeV*, ATL-PHYS-PUB-2015-036, 2015, URL: https://cds.cern.ch/record/2044564.

[63] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at* $\sqrt{s}$ = 7 *TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 17, arXiv: 1406.0076 [hep-ex].

[64] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at* $\sqrt{s}$ = 13 *TeV with the ATLAS detector*, (2017), arXiv: 1703.09665 [hep-ex].

[65] 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].

[66] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, 2015, URL: https://cds.cern.ch/record/2037697.

[67] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: https://cds.cern.ch/record/2160731.

[68] ATLAS Collaboration, *Performance of algorithms that reconstruct missing transverse momentum in* $\sqrt{s}$ = 8 *TeV proton–proton collisions in the ATLAS detector*, Eur. Phys. J. C **77** (2017) 241, arXiv: 1609.09324 [hep-ex].

[69] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton–proton collisions at* $\sqrt{s}$ = 13 *TeV*, ATL-PHYS-PUB-2015-027, 2015, URL: https://cds.cern.ch/record/2037904.

[70] ATLAS Collaboration, *Measurement of the top quark pair production cross-section with ATLAS in the single lepton channel*, Phys. Lett. B **711** (2012) 244, arXiv: 1201.1889 [hep-ex].

[71] ATLAS Collaboration, *Measurements of normalized differential cross-sections for* $t\bar{t}$ *production in pp collisions at* $\sqrt{s}$ = 7 *TeV using the ATLAS detector*, Phys. Rev. D **90** (2014) 072004, arXiv: 1407.0371 [hep-ex].

[72] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 Data at* $\sqrt{s}$ = 13 *TeV*, ATL-PHYS-PUB-2015-051, 2015, URL: https://cds.cern.ch/record/2110140.

[73] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, JINST **9** (2014) P09009, arXiv: 1406.7690 [hep-ex].

[74] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC run 2*, (2017), arXiv: 1704.07983 [hep-ex].

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

[76] G. D'Agostini, *Improved iterative Bayesian unfolding*, ArXiv e-prints (2010), arXiv: 1010.0632 [physics.data-an].

[77] T. Adye, 'Unfolding algorithms and tests using RooUnfold', *Proceedings of the PHYSTAT 2011 Workshop, CERN, Geneva, Switzerland, January 2011, CERN-2011-006, pp 313-318*, 2011 313, arXiv: 1105.1160 [physics.data-an].

[78] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton–proton collision data*, Eur. Phys. J. C **72** (2012) 1909, arXiv: 1110.3174 [hep-ex].

[79] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2016-015, 2016, URL: https://cds.cern.ch/record/2203514.

[80] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, URL: https://cds.cern.ch/record/2157687.

[81] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at* $\sqrt{s} = 7$ *TeV*, Eur. Phys. J. C **73** (2013) 2304, arXiv: 1112.6426 [hep-ex].

[82] ATLAS Collaboration, *Jet energy resolution in proton–proton collisions at* $\sqrt{s} = 7$ *TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. C **73** (2013) 2306, arXiv: 1210.6210 [hep-ex].

[83] ATLAS Collaboration, *Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) 2305, arXiv: 1203.1302 [hep-ex].

[84] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-015, 2015, URL: https://cds.cern.ch/record/2037613.

[85] ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004, 2014, URL: https://cds.cern.ch/record/1664335.

[86] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046, 2014, URL: https://cds.cern.ch/record/1741020.

[87] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction in proton–proton collisions at* $\sqrt{s} = 7$ *TeV with ATLAS*, Eur. Phys. J. C **72** (2012) 1844, arXiv: 1108.5602 [hep-ex].

[88] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653, arXiv: 1608.03953 [hep-ex].

[89] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7** (1979) 1.

[90] ATLAS Collaboration, *Modelling of the* $t\bar{t}H$ *and* $t\bar{t}V(V = W, Z)$ *processes for* $\sqrt{s} = 13$ *TeV ATLAS analyses*, ATL-PHYS-PUB-2016-005, 2016, URL: https://cds.cern.ch/record/2120826.

[91] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z Boson Production Cross Sections in pp Collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Detector*, ATLAS-CONF-2015-039, 2015, URL: https://cds.cern.ch/record/2045487.

[92] F. Berends, H. Kuijf, B. Tausk and W. Giele, *On the production of a W and jets at hadron colliders*, Nuclear Physics B **357** (1991) 32 , ISSN: 0550-3213.

[93] ATLAS Collaboration, *Multi-boson simulation for 13 TeV ATLAS analyses*, ATL-PHYS-PUB-2016-002, 2016, URL: https://cds.cern.ch/record/2119986.

[94] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. **G43** (2016) 023001, arXiv: 1510.03865 [hep-ph].