Measurement of colour flow with the jet pull angle in $t\bar{t}$ events using the ATLAS detector at $\sqrt{s} = 8$ TeV

ATLAS Collaboration

1. Introduction

Due to the confining nature of the strong force, directly measuring the quantum chromodynamic (QCD) interactions between quarks and gluons is not possible. The strength and direction of the strong force depends on the colour charge of the particles involved. To a good approximation, the radiation pattern in QCD can be described through a colour-connection picture, which consists of colour strings connecting quarks and gluons of one colour to quarks and gluons of the corresponding anti-colour. An important question is whether there is evidence of these colour connections \textit{(colour flow)} in the observable objects: colour-neutral hadrons and the jets they form. The study of energy distributions inside and between jets in various topologies has a long history, dating back to the discovery of gluons in three-jet events at PETRA [1–4]. Colour connections are still a poorly constrained QCD effect, which motivates the dedicated study presented in this Letter. If well understood, experiments can exploit colour flow to aid Standard Model (SM) measurements and searches for physics beyond the SM (BSM). As a test that the colour flow can be extracted from the observable final state, the data are compared to models with simulated $W$ bosons that are colour-charged or colour-neutral.

One observable predicted to contain information about the colour representation of a dijet resonance like the $W$, $Z$, or Higgs boson, is the \textit{jet pull vector} [5]. The pull vector for a given jet $J$ with transverse momentum, $p_T^J$, is defined as

$$\vec{v}_p^J = \sum_{i \in J} \frac{p_{T,i}^J r_{i},}{p_T^J}$$

The sum in Eq. (1) runs over jet constituents with transverse momentum $p_T^i$ and location $r_i = (\Delta y_i, \Delta \phi_i)$, defined as the vector difference between the constituent and the jet axis $(y_J, \phi_J)$ in rapidity $(y) - \text{azimuthal angle} (\phi)$ space.\(^1\) Given the pull vector for jet $J_1$, the angle formed between this pull vector and the vector connecting $J_1$ and another jet $J_2$, $(y_{J_2} - y_{J_1}, \phi_{J_2} - \phi_{J_1})$, is expected to be sensitive to the underlying colour connections between the jets. This is shown graphically in Fig. 1, and the angle is called the \textit{pull angle}, denoted $\theta_p(J_1, J_2)$. The pull angle is symmetric around zero when it takes values between $-\pi$ and $\pi$ and so henceforth $\theta_p(J_1, J_2)$ refers to the magnitude of the angle in $(\Delta y, \Delta \phi)$ space with $0 < \theta_p \leq \pi$. If the pull vector is computed using jets originating from colour-connected quarks, $\theta_p \sim 0$ since the radiation is predicted to fall mostly between the two jets.

\(^1\) 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 $(r, \phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle $\theta$ as $\eta = -\log \tan (\theta/2)$. Separation between objects in $(\eta, \phi)$ space is defined as $\Delta \eta = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$. The rapidity of a four-vector $y = 0.5 \log \left( \frac{E + p_z}{E - p_z} \right)$, where $E$ is the energy and $p_z$ the component of the momentum parallel to the beam axis.
the vector is computed using two jets which do not originate from colour-connected quarks, there is no reason to expect \( \theta_p \) to be small. Thus \( \theta_p \) should be useful for determining colour connections.

One of the challenges in studying colour flow is the selection of a final state with a known colour composition. Colour-singlet \( W \) bosons from \( tt \) events provide an excellent testing ground because these bosons have a known initial (colourless) state and such events can be selected with high purity. The first study of colour flow using \( W \) bosons from top quark decays was performed by the DØ Collaboration [6]. In the DØ analysis, calorimeter cells clustered within jets were used as the constituents in Eq. (1) and the resulting distribution was compared to \( W \) singlet (nominal) and \( W \) octet templates. The impact of the colour flow on the observable energy distributions is very subtle; the DØ result was statistically limited and was not able to significantly distinguish singlet and octet radiation patterns.

The analysis presented in this Letter is a measurement of the jet pull angle in \( \sqrt{s} = 8 \) TeV \( pp \) collisions at the Large Hadron Collider (LHC) with the ATLAS detector. Instead of comparing the reconstructed jet pull angle in data directly to simulated templates, as was done by the DØ Collaboration, the jet pull angle distribution is first corrected for detector resolution and acceptance effects. This allows for direct comparison with particle-level predictions for models of physics beyond the SM as well as simulations of non-perturbative physics effects with various tunable parameter values.

2. Object and event selection

The ATLAS detector independently measures the inclusive and charged-particle energy distributions in jets. This allows the jet pull angle to be constructed using only the charged constituents of jets, or both the charged and neutral constituents. In this analysis, both options are used in order to provide independent measurements with different experimental systematic uncertainties. Charged-particle momenta are measured in a series of tracking detectors (collectively called the inner detector), covering a range of \( |\eta| < 2.5 \) and immersed in a 2 T magnetic field. Electromagnetic and hadronic calorimeters surround the inner detector, with forward calorimeters allowing electromagnetic and hadronic energy measurements up to \( |\eta| = 4.5 \). A detailed description of the ATLAS detector can be found elsewhere [7].

The anti-\( k_T \) algorithm [8] with radius parameter \( R = 0.4 \) is used to reconstruct jets from clusters of calorimeter cells [9] with de-

![Fig. 1. Diagram illustrating the construction of the jet pull angle for jet \( J_1 \) with respect to \( J_2 \).](image_url)

Table 1

<table>
<thead>
<tr>
<th>Process</th>
<th>Number of events</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau \bar{\tau} )</td>
<td>95 400 ± 14 000</td>
</tr>
<tr>
<td>( Wt )-channel single top</td>
<td>2730 ± 600</td>
</tr>
<tr>
<td>( s\bar{s} )- and ( t\bar{t} )-channel single top</td>
<td>150 ± 10</td>
</tr>
<tr>
<td>( W + ) jets</td>
<td>3710 ± 120</td>
</tr>
<tr>
<td>( Z + ) jets</td>
<td>560 ± 270</td>
</tr>
<tr>
<td>Dibosons</td>
<td>190 ± 40</td>
</tr>
<tr>
<td>Multijets</td>
<td>2500 ± 910</td>
</tr>
<tr>
<td>Total SM</td>
<td>105 000 ± 14 000</td>
</tr>
<tr>
<td>Data</td>
<td>102 987</td>
</tr>
</tbody>
</table>

2. The primary vertex is defined as the vertex with the highest \( \sum p_T^2 \) of associated tracks.
shows the predicted composition compared to the data yield. More details about the various contributions are given below.

There are two pull angles calculated per event, one calculated from the all-particles pull vector and one from the charged-particles pull vector of the highest $p_T$ jet assigned to the hadronic $W$ boson decay. The all-particles and charged-particles pull angles are the angles that the corresponding pull vectors make with the direction from $J_1$ to $J_2$. Figs. 2(a) and 2(b) show comparisons between data and simulation for the all-particles and charged-particles pull angles, calculated at detector level, i.e. from reconstructed objects. Both distributions are broadly flat with an enhancement at small angles, consistent with the SM prediction.

3. Event simulation

Monte Carlo (MC) samples are produced in order to determine how the detector response affects the pull angle and to estimate some of the non-$\tilde{t}\tilde{t}$ contributions in the data. The details of the samples used are shown in Table 2.

Aside from the $W+$jets background, all MC samples are normalised to their theoretical cross-sections, calculated to at least next-to-leading order (NLO) precision in QCD [33–38]. For the purpose of comparison between data and the SM prediction before unfolding, $\tilde{t}\tilde{t}$ events are normalised to a cross-section of $253 \pm 15$ pb, calculated at next-to-next-to-leading order (NNLO) in QCD including next-to-next-to-leading logarithmic (NNLL) soft gluon terms [39], assuming a top-quark mass of 172.5 GeV.

Generated events are processed with a full ATLAS detector and trigger simulation [40] based on GEANT4 [41] and reconstructed using the same software as the experimental data. The effects of pileup are modelled by adding to the generated hard-scatter events multiple minimum-bias events simulated with Pythia 8.160 [42], the A2 set of tuned MC parameters (tune) [43] and the MSTW2008LO Parton Distribution Function (PDF) set [44]. The distribution of the number of interactions is then weighted to reflect the pileup distribution in the data.

To test the sensitivity of the jet pull angle to the singlet nature of the $W$ boson, a sample was generated with a colour-octet $W$ boson. The octet $W$ boson is simulated using the same setup as
the nominal \( \bar{t}t \) setup described in Table 2. Using the partons produced with POWHEG (in the LHE [45] format), the colour flow is inverted such that one of the \( W \) decay daughters shares a colour line with the \( b \)-quark and the other shares a line with the top quark, as demonstrated schematically in Fig. 3. This sample is referred to as colour flipped in the rest of this Letter.

4. Unfolding

In order to make direct comparisons with various theoretical models, the data are unfolded to particle-level objects. Particle-level jets are clustered from simulated particles with a mean lifetime \( t > 30 \) ps, before taking into account interactions with the detector. The particle-level inputs to the all-particles pull angle are all of the charged and neutral particles clustered within particle-level jets. Only the charged particles clustered within the particle-level jets are used for the charged-particles pull angle. Additional information about the particle-level object definitions can be found in Ref. [46].

The particle-level event selection is analogous to the detector-level selection described in Section 2 with detector-level objects replaced with particle-level objects. Exactly one electron or muon and at least four jets are required, each with \( E_T > 25 \) GeV and \( |\eta| < 2.5 \), with events discarded if the electron or muon is within \( \Delta R = 0.4 \) of a jet. The particle-level missing transverse momentum, \( E_T^{\text{miss}} \), calculated using the sum of neutrino four-momenta, is required to be \( E_T^{\text{miss}} > 20 \) GeV and the sum of \( E_T^{\text{miss}} + m_t > 60 \) GeV. At least two of the selected jets are required to be identified as \( b \)-jets using the same definition as that found in Ref. [46]. As with the detector-level calculation of the pull angle, the two leading-\( p_T \) particle-level non-\( b \)-jets with \( |\eta| < 2.1 \) are labelled as the jets from the hadronically decaying \( W \) boson.

The first step of the unfolding procedure is to subtract from the data an estimate of the non-\( t\bar{t} \) backgrounds bin-by-bin in the pull angle distribution. The \( W+\text{jets} \) and multijet backgrounds are estimated from the data using the charge asymmetry and matrix methods, respectively [47]. The other backgrounds are estimated from simulation. Single-top \( Wt \) events have hadronically decaying \( W \) bosons and are thus sensitive to colour flow; however, even large changes in the colour flow through the single-top contribution are subdominant compared to other uncertainties, and are thus ignored.

After subtraction, the data are unfolded using an iterative Bayesian (IB) technique [48] as implemented in the RooUnfold framework [49]. The IB method iteratively applies Bayes’ theorem using the response matrix to connect the prior to posterior at each step. The response matrix is constructed from the nominal \( t\bar{t} \) simulation (also used for the prior) and describes the bins migrations between the particle-level and detector-level pull angle distributions. The binning of the response matrix and the number of iterations in the IB method are chosen to minimise the overall uncertainty, described in Section 5. In addition to correcting for bin migrations, the unfolding procedure also corrects for events that pass either the detector-level or particle-level selection, but not both. The corrections are estimated using simulated \( tt \) events. Approximately 70% of events that pass the detector-level selection also pass the particle-level selection, while approximately 20% of events that pass the particle-level selection also pass the detector-level selection. These corrections are found to be largely independent of the pull angle.

The all-particles pull angle has a stronger dependence on the colour flow, but has a worse resolution than the charged-particles pull angle and is more sensitive to the particle-level spectrum used as the IB prior. Three bins and 15 iterations are used for the all-particles pull angle and four bins and three iterations are used for the charged-particles pull angle. More iterations are required for the all-particles pull angle to reduce the sensitivity to the IB prior, at the cost of a larger statistical uncertainty. The size of each bin is comparable to the pull angle resolution. The uncertainty on the unfolding procedure is determined in a data-driven way by reweighting the particle-level distributions so that the corresponding detector-level distributions are in better agreement with the data [50]. The method uncertainty is then estimated using the nominal response matrix by comparing the unfolded, reweighted simulation with the reweighted particle-level distribution.

5. Systematic uncertainties

There are various systematic uncertainties which can impact the measured jet pull angle. The sources of uncertainty can be classified into two categories: experimental uncertainties and theoretical modelling uncertainties. In the first category, some uncertainties impact the pull angle directly while the others impact only the acceptance.

The uncertainty in the energy scale and angular distribution of clusters of calorimeter cells are primary sources of uncertainty that directly impact the pull angle. To estimate the uncertainty in the angular resolution of clusters, isolated tracks are matched to clusters in events enriched in \( Z(\rightarrow \mu\mu) + \) jets events. The distribution of \( \Delta R \) (track, cluster) is investigated for various bins of track \( p_T \) and \( \eta \), and in extrapolations to various layers of the calorimeter. The observed resolution uncertainty is \( \lesssim 1 \) mrad, but to account for the possible effects of multi-particle clusters, 5 mrad is used, as in Ref. [51]. Estimations of the cluster energy scale uncertainty are based on measurements of the ratio of the cluster energy to the \( p_T \) of matched tracks \( E/p \) [52]. The ratio \( E/p \) in 8 TeV data is compared with simulation and the differences are used to estimate the uncertainty, which is parameterised as \( \alpha(1 + \beta p_t) \), where \( \alpha \) and \( \beta \) depend on \( \eta \). Typical values are \( \alpha = 0.05 \) and \( \beta = 0.5 \) GeV. To account for an uncertainty in the energy lost due to inactive material in the detector and noise thresholds, low-energy clusters are randomly removed with energy-dependent probabilities, as in Ref. [53].

For the charged-particles pull angle, there are uncertainties associated with track reconstruction efficiency, which are estimated by randomly removing tracks with an \( \eta \)-dependent probability [54]. Furthermore, tracks are randomly removed from jets with a \( p_T \)-dependent factor to estimate the effect of uncertainties in the efficiency of track reconstruction inside jets. For jets

---

1 The RMS of the all-particles (charged-particles) pull angle is found to be 0.35\( \pi \) (0.28\( \pi \)) radians.
with $p_T < 500$ GeV, the uncertainty on the track reconstruction efficiency is less than 1%.

As the pull vector definition uses the calorimeter jet $p_T$, both the all-particles and charged-particles pull angle are affected by the uncertainty in the jet energy scale [55,56] and resolution [57]. However, changes in the jet energy scale and resolution do not impact the pull angle, but do impact the results via the acceptance due to $p_T$ thresholds. Similarly, uncertainties in the lepton energy scale, trigger efficiency, $E_{T\text{miss}}$ resolution and $b$-tagging efficiencies [17,58–60] indirectly affect the results through changes in acceptance. Other sources of uncertainty on the acceptance, which impact the measurement through the background subtraction, include those related to the luminosity [61], the multijet estimation, and the normalisation and heavy flavour content of the $W$ + jets background [47].

The modelling of the $t\bar{t}$ sample is a primary source of theoretical uncertainty. For instance, there are different treatments of the matrix element calculation (POWHEG+HERWIG versus M@NLO+HERWIG) and the parton shower model (POWHEG+PYTHIA versus POWHEG+HERWIG). The change in the result when using a response matrix constructed from the flipped $t\bar{t}$ simulation is considered as a source of uncertainty on the colour model. Other sources of modelling uncertainties are evaluated, including initial/final-state radiation by varying the radiation simulated with A@CtMC 3.8 [62] constrained by Ref. [63], multiple simultaneous parton collisions using the PERUGIA 2012 M@PHI tune [25], colour reconnections with the PERUGIA 2012 LOWCR tune, overlap between single top and $t\bar{t}$ by comparing the diagram removal scheme with the diagram subtraction scheme [64], and PDF uncertainties [65]. Colour flow describes the colour representation of the hard-scatter process while colour reconnection is a phenomenological model for implementing a finite number of colours in the parton shower. The observable consequences of both processes are similar, but colour reconnection has been constrained by charged-particle distributions in experimental data [25]. Varying the top quark mass by $\pm 1$ GeV has a negligible impact on this measurement.

The systematic uncertainties are estimated by unfolding the data with varied response matrices or by subtracting varied background predictions from the data. The sources of uncertainty described above are summarised in Table 3 for the first bin of the pull angle for both the all-particles and charged-particles pull angle. In general uncertainties are found to be larger for the all-particles pull angle than the charged-particles pull angle due to the remaining sensitivity to the IB prior, as discussed in Section 4. All bins carry information about the underlying colour flow, but the first bin has the largest expected difference between the flipped and nominal models.

### 6. Results

The unfolded data are shown in Fig. 4 (all-particles) and Fig. 5 (charged-particles) and compared to SM $t\bar{t}$ predictions simulated with POWHEG+PYTHIA and POWHEG+HERWIG, and a flipped $t\bar{t}$ simulation generated using POWHEG+PYTHIA.

The data favour the SM predictions over the prediction of the flipped model. Of the two SM predictions, POWHEG+PYTHIA offers a slightly better description of the data. The level of agreement between the data and the POWHEG+PYTHIA models is quantified using $\Delta \chi^2$ as a test statistic. A $p$-value is computed taking into account systematic and statistical uncertainties and their correlations by generating pseudo-datasets from the full covariance matrix and then normalising the pull angle distributions. This $p$-value is then converted into a one-sided Gaussian equivalent $Z$-score.
unfolded data are found to agree with the nominal SM colour flow (with POWHEG-+PYTHIA) at the 0.8σ (0.9σ) level and differ from the flipped model at 2.9σ (3.7σ) for the all-particles (charged-particles) pull angle. Although the statistical uncertainty is largely uncorrelated between the all-particles and charged-particles pull angle measurements, the systematic uncertainty is not. Since the charged-particles pull angle measurement is more sensitive than the all-particles pull angle measurement, a full combination of the two measurements does not provide increased sensitivity.

7. Summary

The analysis presented in this Letter describes a measurement of the orientation of radiation from jets identified as originating from a W boson in tt̄ events. The measurement uses 20.3 fb−1 of √s = 8 TeV pp collision data recorded by the ATLAS detector at the LHC. To quantify the distribution of energy inside one jet relative to another, the distribution of the jet pull angle is extracted from the data using information from both the ATLAS calorimeter and tracking detectors. The jet pull angle is found to correctly characterise the W boson as a colour singlet, with data disfavouring an alternative colour-octet model at greater than 3σ. This illustrates the potential to use the jet pull angle in future SM measurements and BSM searches. The jet pull angle measurement is presented as a normalised fiducial tt̄ differential cross-section, allowing the results to be used to constrain implementations of colour connection.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CINPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF, DNSRC and Lundbeck Foundation, Denmark; EPLANET, ERC and NSRF, European Union; IN2P3-CNRS, CEA-DSM/IRFU, France; GNSF, Georgia; BMBF, DBF, HGF, MPG and AvH Foundation, Germany; GSKT and NSRF, Greece; RGC, Hong Kong SAR, China; ISF, MINERVA, GIF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; FOM and NWO, Netherlands; BRF and RCN, Norway; MNISW and NCN, Poland; GRICES and FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MSTD, Serbia; MSSR, Slovakia; ARIES and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenius Foundation, Sweden; SER, SNSF and Cantons of Bern and Geneva, Switzerland; NSC, Taiwan; TAEK, Turkey; STFC, the Royal Society and Leverhulme Trust, United Kingdom; DOE and NSF, United States of America.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN and the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA) and in the Tier-2 facilities worldwide.

References


ATLAS Collaboration
