Testing non-standard sources of parity violation in jets at the LHC, trialled with CMS Open Data

Christopher G. Lester[a] **and Matthias Schott**[b,c]

[a] *Cavendish Laboratory, University of Cambridge,*
JJ Thomson Avenue, Cambridge, United Kingdom

[b] *Massachusetts Institute of Technology,*
Cambridge, U.S.A.

[c] *Institute of Physics, Johannes Gutenberg University,*
Staudingerweg, Mainz, Germany

E-mail: lester@hep.phy.cam.ac.uk, matthias.schott@cern.ch

ABSTRACT: The Standard Model violates parity, but only by mechanisms which are invisible to Large Hadron Collider (LHC) experiments (on account of the lack of initial state polarisation or spin-sensitivity in the detectors). Nonetheless, new physical processes could potentially violate parity in ways which *are* detectable by those same experiments. If those sources of new physics occur only at LHC energies, they are untested by direct searches. We probe the feasibility of such measurements using approximately $0.2\,\mathrm{fb}^{-1}$ of data which was recorded in 2012 by the CMS collaboration and made public within the CMS Open Data initiative. In particular, we test an inclusive three-jet event selection which is primarily sensitive to non-standard parity violating effects in quark-gluon interactions. Within our measurements, no significant deviation from the Standard Model is seen and no obvious experimental limitations have been found. We discuss other ways that searches for non-standard parity violation could be performed, noting that these would be sensitive to very different sorts of models to those which our method would constrain. We hope that our initial studies provide a valuable starting point for rigorous future analyses using the full LHC datasets at $13\,\mathrm{TeV}$ with a careful and less conservative estimate of experimental uncertainties.

Contents

1 Introduction

It is rightly beyond doubt that the laws of physics violate parity. The elegant experiments of the 1950s [1–3] unambiguously showed that the weak interaction of the Standard Model can tell the difference between our universe and its mirror image. However, no attempts have yet been made to identify whether there are non-standard parity violation mechanisms which operate only at high energies probed by the Large Hadron Collider (LHC).

The absence of tests of parity is largely a pragmatic response to the obstacles presented by the LHC: its beams are not polarised, its detectors are not sensitive to polarizations, and it is mathematically impossible to construct a parity violating spin-averaged matrix element within any CP-conserving Locally Lorentz-Invariant quantum field theory (LLIQFT), effective or otherwise.[1] Given that the only route to probing parity-violation in the Standard

---

[1]To reveal the existence of parity-violation a model must possess at least one matrix element having both a parity-even and a parity-odd part. After trace identities have removed spinor sums, the only parity-odd expressions which can remain in a Lorentz-invariant $|M|^2$ are contractions of the totally antisymmetric alternating tensor with groups of four linearly independent four-momenta: $\epsilon_{\mu\nu\sigma\tau}a^\mu b^\nu c^\sigma d^\tau$. While such terms are parity-odd, they are also time-odd. Assuming CPT-symmetry, such a matrix element therefore also violates CP. CP-conserving local Lorentz-invariant quantum field theories therefore cannot generate parity violating differential cross sections.

Model at the LHC would be from within the C-even part of its (very small) CP-violating sector,[2] it is not surprising that no such analyses have yet been performed.[3]

It would be wrong, however, to conclude that genuine tests of parity are therefore of limited value. On the contrary, we will see that genuine tests of parity-invariance are very straightforward to make and can (in principle) provide arbitrarily large signatures for models not fitting into the class of LLIQFTs.

While it is true that there is an overwhelming theoretical preference for LLIQFTs (the Standard Model itself is one, as are most popular extensions including those featuring supersymmetry, leptoquarks, technicolor, axions, additional gauge interactions, etc.) there is no law of nature which *demands* that new physics be describable only by such theories.

Moreover, given the lack of evidence for new physics found at LHC thus far, the need for the community to search in all possible hiding places is surely greater than ever. In particular, it is hard to imagine any reason why every possible attempt should not be made to test and re-test the fundamental symmetries of nature every time a door opens onto a new energy range.

We use therefore CMS Open Data to make the first LHC search for unequivocal evidence of parity violation. Clearly, the possible physics reach of our study is limited — once by the data statistics used, secondly by the assumed systematic uncertainties; its primary goal is to evaluate potential experimental challenges, while the actual constrains on parity violation are of secondary importance. The Standard Model itself should provide no appreciable signal on account of its approximate CP-symmetry. While sensitivity might exist to LLIQFT extensions of the Standard Model having larger sources of CP-violation, these are probably constrained by data from rare decays. The primary sensitivity, thus, is to extensions which are outside the LLIQFT class altogether.

The paper is structured as follows: section 1 (this introduction) has explained why genuine tests of parity violation in LHC data are important, and discusses the limitations which the LHC imposes on which sources of parity violation are observable; section 2 discusses the properties of the specific parity-odd event-variable, α, used in this paper; section 3 summarises the CMS detector, its relevant reconstruction objects, and the datasets used in this analysis; section 5 discusses the calibration of the relevant physics objects used (in particular jets) used in this analysis; section 6 details the actual analysis strategy, signal selection and background estimation; section 7 both defines the test statistic A_α which is used to provide the a measure of non-standard parity violation in the dataset as

---

[2]One could, in principle, demonstrate parity-violation unambiguously by using a 'genuine CP-odd' observable (such as one of those described in [4]) on a C-even final state.

[3]This limitation does not prevented the LHC from making measurements of parity-violating parameters within models *in which a particular mechanism of parity-violation is present by assumption*. For example, the differences between the axial and vector couplings of the Z-boson in the Standard Model violate parity and were measured in [5, 6]. However, neither of these papers incontrovertibly demonstrates that nature violates parity.

The reason is simple: the angles from which forward-backward asymmetries are calculated are even under parity, unlike primary observables from the experiments of the 1950s. The very same forward backward asymmetries could therefore also be explained, at least in principle, by some alternative parity *conserving* theory.

a whole, and then shows the asymmetries seen in CMS Open Data; section 8 unfolds the preceding results for detector effects. The paper concludes in section 9.

2 Jet-based signatures for non-standard parity violation

One of the evident disadvantages of searching for the breaking of a particular symmetry without a particular model in mind, is that there the number of potential signatures is almost unbounded. If testing for genuine parity violation, any parity-odd variable which is invariant under the action of relevant non-parity symmetries[4] has the capacity (assuming a parity-even event selection) to be sensitive to *some* parity-violating model, albeit not necessarily one with a strong theoretical motivation.

Lacking clear guidance from any well motivated non LLIQFT models, we make no claims of generality or optimality concerning our choice to concentrate on a signal region containing three jets. We argue only that the investigation is interesting for the reasons already given in the introduction.

A simple way of constructing a parity-odd variable from jets which is insensitive to rotations is to use a pseudoscalar. This simplest of these is arguably the scalar triple product $[\vec{p}^{j_1}, \vec{p}^{j_2}, \vec{p}^{j_3}] \equiv (\vec{p}^{j_1} \times \vec{p}^{j_2}) \cdot \vec{p}^{j_3}$, computed from the three jet three-momenta $\vec{p}^{j_1}$, $\vec{p}^{j_2}$ and $\vec{p}^{j_3}$ ordered by the requirement

$$p_T(j_1) \geq p_T(j_2) \geq p_T(j_3). \tag{2.1}$$

Though the scalar triple product has the beneficial property of changing sign under parity, $(\vec{p}^{j_1}, \vec{p}^{j_2}, \vec{p}^{j_3}) \rightarrow (-\vec{p}^{j_1}, -\vec{p}^{j_2}, -\vec{p}^{j_3})$, it has the disbenefit that it varies not only with the directions but also linearly with the magnitudes of each of the momenta. We therefore choose instead to use a normalised variant whose dependence is primarily on the directions of the momenta:[5]

$$\frac{\vec{p}^{j_1} \times \vec{p}^{j_2}}{|\vec{p}^{j_1} \times \vec{p}^{j_2}|} \cdot \frac{\vec{p}^{j_3}}{|\vec{p}^{j_3}|} = \sin \alpha. \tag{2.2}$$

---

[4]For example: an ideal parity-testing variable should be invariant under rotation of an LHC event by any angle about the beam pipe axis, or by 180 degrees about any axis at right angles to the beam pipe, unless sensitivity to Lorentz violating effects is desired.

[5]The quantity $\sin \alpha$ does not depend *exclusively* on the directions of $\vec{p}^{j_1}$, $\vec{p}^{j_2}$ and $\vec{p}^{j_3}$ because (2.1) uses other properties of $\vec{p}^{j_1}$, $\vec{p}^{j_2}$ and $\vec{p}^{j_3}$ (specifically p_T ordering) to define which of them is which. Nonetheless, the normalisation choice makes α completely independent of measurement errors on momentum magnitudes *until* the point at which any measurement errors become large enough to change the p_T ordering. Had the normalisation been achieved by dividing by $|\vec{p}^{j_1}||\vec{p}^{j_2}||\vec{p}^{j_3}|$ instead of by $|\vec{p}^{j_1} \times \vec{p}^{j_2}||\vec{p}^{j_3}|$, the same insensitivity to small changes in momenta magnitudes would still be present, but the resulting variable would no longer have a simple geometric interpretation as the sine of an angle between two planes. We opt for the former to preserve a geometric interpretation of the variable. We re-emphasise that there is nonetheless considerable arbitrariness in our choice of α — which itself is attributable to the lacking concrete models predicting visible parity violation at the LHC. Were such models to exist, we could speak quantitatively of 'better' or 'worse' choices of variable, and make optimisations. At the present time, however, we can only described the consequences and limitations of the choice we have made. The body text provides a number of examples of alternative choices which, if made, would result in variables with very different regions of sensitivity. We hope these ideas and others will be the subject of future investigations.
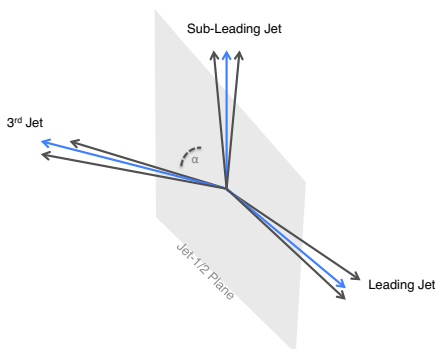
Figure 1. Illustration of the definition of α in a three-jet system, where the leading jet is defined as the jet with the highest p_T.
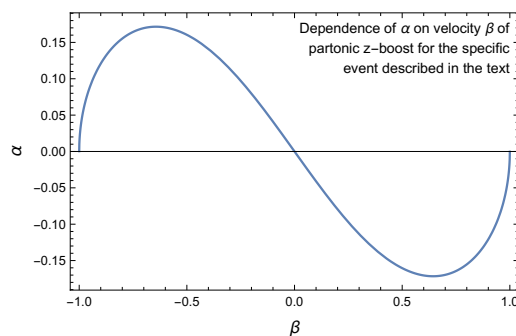
Figure 2. Dependence of α on the longitudinal velocity β of the three-jet system for the example configuration described in the text.

We have denoted this quantity '$\sin\alpha$' since it may be interpreted as the sine of the angle α which momentum $\vec{p}^{j3}$ makes with the plane spanned by $\vec{p}^{j1}$ and $\vec{p}^{j2}$, as shown in figure 1. Positive angles α occur if and only if $\vec{p}^{j3}$ is on the same side of that plane as $\vec{p}^{j1} \times \vec{p}^{j2}$.

Note that magnitude of $\sin\alpha$ is not invariant under longitudinal boosts of the system of three jets. Even the sign of α can be altered by longitudinal boosts for certain jet momenta. Consider, for example, a co-planar set of jet momenta such as $\vec{p}^{j1} = (3,0,0)$, $\vec{p}^{j2} = (0,2,0)$ and $\vec{p}^{j3} = (1,1,0)$. As these are co-planar they have $\alpha = 0$. If those momenta are then given a longitudinal boost, the value of $\sin\alpha$ ceases to be zero, but boosts in one direction generate equal and opposite values of α to boosts in the other direction.[6] Consequently, any search for a new physics process which generates an α-asymmetry in the frame of the hard process must contend with the possibility that such an asymmetry might be partly 'washed away' by the longitudinal boosts which relate the hard-process frame (HPF) to the lab frame (LF). Evidently, the sort of processes most easily washed away would be those which have three co-planar (or almost co-planar) final-state momenta in the HPF and are generated from initial states (such as gluon-gluon or quark-quark) whose hard processes have symmetrically distributed longitudinal boosts.

Given that a large fraction of the total LHC cross section comes from gluon-gluon and given that conservation of momentum would require co-planarity in the HPF of the outgoing partonic momenta in any $2 \to 2$ or $2 \to 3$ parton-level process, one might be concerned that α appears to lack sensitivity to parity-violating effects in some important places!

Such concern is, however, premature. The *only* way in which a set of co-planar momenta can have a notion of handedness akin to parity is if the plane in which they live is orientable.[7] It is for the same reason that the (planar) letter **R** can be distinguished from its mirror image **Я** only if *one* side of the paper on which both are printed is deemed 'special'. The plane separating two *identical* interacting initial-state partons in the HPF

---

[6]Concretely, longitudinal boosts having velocities $\beta_z = \pm 1/\sqrt{2}$ change the momenta in the supplied example to $\vec{p}^{j1} = (3,0,\pm3)$, $\vec{p}^{j2} = (0,2,\pm2)$ and $\vec{p}^{j3} = (1,1,\pm\sqrt{2})$ which have values of $\sin\alpha$ of $\pm(-1/\sqrt{3} + 1/\sqrt{6})$ respectively. A plot showing how α in the above example varies for any β is shown in figure 2.

[7]By orientable, we mean that it has a 'front' which is distinguishable from its 'back'.

is necessarily *non*-orientable.[8] Therefore, although $2 \rightarrow 2$ and $2 \rightarrow 3$ processes using gluon-gluon and quark-quark initial states are those in which one might have feared the aforementioned 'wash out', these same initial states are necessarily those which can carry through no notion of handedness from the intial to the final state in the HPF anyway, let alone one that could be washed away by symmetric boosts from that frame to the LF.[9]

In contrast, asymmetric initial states (such as quark-gluon) *are* capable of distinguishing forward and backward directions in the HPF, and so *are* (at least in principle) able to generate detectable lab-frame asymmetries in α. It is important to realise that such asymmetries could be seen even for Lorentz- and rotationally-invariant $2 \rightarrow 3$ new-physics processes having final state momenta which are co-planar in the HPF. It is differences between the forward and backward boosts from parton distribution functions which would drive such an asymmetry, *exploiting* the dependence of α on longitudinal boosts. Given the asymmetry is generated by this mechanism, there is some motivation for biasing the signal region to events in which the three-jet system has a high absolute rapidity.

In summary: the three-jet variable which we have called α can (in principle) detect parity-violating effects in a broad class of new-physics models at the LHC. As the particular variable we used to illustrate our search is just one of many that could have been chosen, we make no claims of optimality — indeed we have highlighted many situations in which it has no sensitivity to anomalous parity violation whatsoever. Nonetheless, we have: (i) demonstrated that parity violating models to which our variable *is* sensitive do not need to abandon the equivalence principle, Lorentz invariance or isotropy, and (ii) noted that since α is defined in terms of only three objects the parity violating processes which α is sensitive to would probably include only those which couple to asymmetric initial states, such as quark-gluon.

In future work, it may be interesting to consider other event variables which depend on *two* or *four* (rather than three) final state momenta. A simple variable of the first sort might be the product $\Delta(\phi)\Delta(\eta)$ of the azimuthal and pseudorapidity differences between any two final state objects.[10] Variables acting on four objects could show sensitivity to non-

---

[8]... assuming isotropy of the laws of physics.

[9]Another way of explaining the same point is to consider the following question: "Can a $gg \rightarrow jjj$ new-physics processes result at the LHC in more 'right-handed' events of the form 'α_+' with momenta $\{\vec{p}^{j1} = (3,0,3), \vec{p}^{j2} = (0,2,2), \vec{p}^{j3} = (-1,-1,\sqrt{2})\}$ than 'left-handed' events of the form 'α_-' with momenta $\{\vec{p}^{j1} = (0,3,3), \vec{p}^{j2} = (2,0,2), \vec{p}^{j3} = (-1,-1,\sqrt{2})\}$?" [The sets of momenta α_+ and α_- map into each other under the exchange $x \leftrightarrow y$.] In answering this question, we would first note that both of these jet configurations are co-planar in the HPF as they have z-components which are zero there. Secondly, we could observe that in the HPF an event of type α_+ can always be mapped into an event of type α_- by a 180 degree rotation which leaves the gluon-gluon initial state invariant. Therefore, if the equivalence principle holds and the laws of physics are isotropic, then events of type α_+ should occur as frequently as events of type α_- in gluon-gluon interactions, and so the answer to the posed question is 'No!'. If instead the laws of physics are not Lorentz invariant, or not isotropic, then the answer could instead be 'Possibly!'. A similar argument would fail if the initial state were asymmetric (such as quark-gluon) since then the 180-degree rotation would not leave the initial state invariant and it would not be possible to conclude that the number of α_+ and α_- events need be identical.

[10]This amazing variable is invariant under longitudinal boosts and does not need the final state particles to be distinguishable or carry an ordering!

standard parity violating effects which do not need the initial state to provide an reference direction. One example would be the four-jet (longitudinal boost invariant) event-variable δ defined by:

$$\delta \equiv \frac{p^\mu(j_1) p^\nu(j_2) p^\sigma(j_3) p^\tau(j_4) \epsilon_{\mu\nu\sigma\tau}}{p_T(j_1) p_T(j_2) p_T(j_3) p_T(j_4)}. \tag{2.3}$$

The numerator of δ is the archetypical Lorentz-invariant pseudoscalar, so is certainly invariant with respect to longitudinal boosts and capable of demonstrating non-standard parity violating effects. The numerator makes δ dimensionless, and is also invariant under longitudinal boosts because it is built exclusively from transverse quantities. A disadvantage of δ, however, is that it requires the inputs to have been ordered in some way, as in (2.1). It is by no means a requirement that non-standard parity violating effects should come together with an ordering requirement on momenta. It may also be useful, therefore, to consider completely symmetric pseudoscalars. These are pseudoscalars which (unlike δ) do not change their value if their input momenta are re-ordered. Examples include the four completely symmetric pseudoscalars S_1, S_2, S_3 and S_4 defined by

$$S_1(p(1), p(2), p(3), p(4)) = \sum_{i,j,k,l=1}^{3} [i,j,k,l](i \cdot j)(i \cdot j)(i \cdot k), \tag{2.4}$$

$$S_2(p(1), p(2), p(3), p(4)) = \sum_{i,j,k,l=1}^{3} [i,j,k,l](i \cdot j)(i \cdot k)(j \cdot l), \tag{2.5}$$

$$S_3(p(1), p(2), p(3), p(4)) = \sum_{i,j,k,l=1}^{3} [i,j,k,l](i \cdot i)(j \cdot j)(i \cdot k) \qquad \text{and} \tag{2.6}$$

$$S_4(p(1), p(2), p(3), p(4)) = \sum_{i,j,k,l=1}^{3} [i,j,k,l](i \cdot i)(j \cdot k)(i \cdot j), \tag{2.7}$$

in which $(a \cdot b)$ stands for $p^\mu(a) p_\mu(b)$ and $[a,b,c,d]$ stands for $\epsilon_{\mu\nu\sigma\tau} p^\mu(a) p^\nu(b) p^\sigma(c) p^\tau(d)$. S_1 to S_4 are not dimensionless, but they could easily be made so in the same way as δ in (2.3).[11]

3   The CMS detector and reconstructed objects,

The data used in this analysis have been recorded with the CMS detector at the LHC in the year 2012. CMS is a high-energy physics experiment which uses a superconducting solenoid, of 6 m internal diameter, with a magnetic field of 3.8 Tesla. The inner detector (ID) of CMS can reconstruct trajectories of charged particles using a silicon pixel and strip tracker. Electrons and photons are identified and measured in a crystal electromagnetic calorimeter (ECAL), while energies of hadrons or hadronic particle jets are determined in a brass/scintillator hadron calorimeter (HCAL). Muons are identified and measured in the

---

[11]Note that S_3 and S_4 are identically zero if evaluated on four-momenta belonging to massless particles. S_1 and S_2 are therefore more relevance to jet-based variables than S_3 and S_4.

muon system (MS), based on gaseous detectors, which surround the hadronic calorimeter and are embedded in the steel flux-return yoke of the magnet system. CMS uses a right-handed coordinate system. Its origin is defined at the interaction point of the proton collisions, the x-axis is pointing towards the center of the LHC, the y-axis pointing upwards and the z-axis along the counterclockwise-beam direction. The polar angle θ is measured from the positive z-axis, however, mostly expressed in terms of the pseudorapidity η, defined by $\eta = -\ln(\tan(\theta/2))$. The azimuthal angle ϕ is measured in the x-y plane. We refer to [7] for a detailed description of the CMS experiment.

The main objects used in this analysis are muons and particle jets, which are briefly discussed in the following. CMS employs a particle-flow algorithm that provides a complete description of the event and identifies electrons, muons, photons, charged hadrons, and neutral hadrons. Muons are reconstructed from a global fit of hits in the MS and the ID, seeded by tracks in the muon system [8]. In this analysis, we require each muon to have a minimal transverse momentum of $p_T > 25\,\mathrm{GeV}$ within a pseudo-rapidity range of $|\eta| < 2.1$, corresponding to the single muon trigger coverage. In addition, standard quality cuts on the number of hits in the ID and the MS, the χ^2 of the fit as well as on the impact parameters are applied, also following previous CMS measurements [9].

Hadronic jets are reconstructed using an anti-k_t algorithm [10, 11] with a radius parameter of 0.5 based on particle-flow objects [12–14], where the clustering algorithms rejects objects that are coming from a pile-up vertex. A jet area method is used to correct for the remaining pile-up contributions [15]. Since the four-momenta of particle-flow objects is summed, the jets can be massive, In this analyses, we only study jets with a minimal transverse momentum of $p_T > 30\,\mathrm{GeV}$ and a jet rapidity of $|y| < 2.4$, since this region allows for the best jet resolution and pile-up rejection. In addition, certain quality criteria on the reconstructed jet properties, such as energy fraction in the ECAL and HCAL or the number of particle-flow objects is applied, following the standard CMS recommendations. Moreover, jets are required to have a minimal distance of $\Delta R > 0.5$ to all reconstructed electrons, muons and photons candidates. The reconstructed transverse momenta of jets is used to order them as the 1st, 2nd and 3rd jet according to their p_T and denoted as j_1, j_2 and j_3, respectively.

This analysis is built around the CMS Open Data Software Framework available at [16]. Version CMSSW_5_3_32 [17] was used. A dedicated open-source framework, BACON [18], which was used for several published studies of the CMS Collaboration, e.g. [9], is used to read the Analysis Object Data (AOD), extracting information on reconstructed objects as well as generator level data, if available. The BACON software framework is also used to apply a GoodRun-List selection provided by the CMS Open Data project [19], as well as calibration constants for particle jets as well as $\not{E}_T$ observables, leading to a separate output-format based on ROOT-tree objects. The typical event size of one simulated top-quark pair event in the BACON-output format amounts to $5\,\mathrm{kB}$. For this work, we used an additional software package, which reduces the output files of BACON further and transforms them into a plain ROOT-NTuple, denoted as ODNTuple in the following with an average event size of $0.8\,\mathrm{kB}$. Our analysis is based on these ODNTuple data. The full software, which has been used for this work, was previously used and validated for [20].

Data stream / Trigger name	Dataset Name	$\int L dt$ [pb^{-1}]
Incl. Jet-Triggers / `Trigger_HLT_PFJet320`	JetHT_Run2012C [21]	289 / 166

Table 1. Overview of data samples used in this analysis together with the corresponding integrated luminosity and the triggers, which have been used during the data taking.

4 Selected Open Data and simulated Monte Carlo samples

The data acquisition system of CMS records only the event information of collisions with dedicated signatures due to the high-collision rate and the limited bandwidth for data-processing. The data used in this analysis has been collected when significant hadronic activity was present in the event, in particular if the `HLT_PFJet320` trigger had fired, targeted to select events with at least one reconstructed jet with a transverse energy of 320 GeV. In total, files corresponding to 289 pb^{-1} from the CMS Open Data [21] have been processed, of which 166 pb^{-1} fired the `HLT_PFJet320` trigger (table 1). The study of a larger data-set was beyond the available computing resources. The integrated luminosity has been calculated using the public available GoodRun-List as well as the corresponding tools. In total, 831,662 triggered events, which pass the GoodRun-List requirement, are used in this analysis.

An overview of the various signal and background Monte Carlo (MC) samples used in this analysis is given in table 2, indicating the underlying physics process, the dataset name and the corresponding inclusive cross-section. The inclusive QCD MC was generated with PYTHIA6 [22] and covers a jet transverse momentum range between 50 and 3000 GeV. In addition, a 4-jet sample is used as alternative QCD sample, which is based on the ALPGEN generator [23]. The Drell-Yan processes (W/Z) in the muon decay channel were generated using with the POWHEGBOX Monte Carlo program [24, 25] interfaced to the PYTHIA v.6.4.26 [22] parton shower model. All other processes are modeled with the tree-level matrix element event generator MADGRAPH v5.1.3.30 [26] interfaced with PYTHIA v6.4.26. In all samples the CT10 PDF set [27] and the Z2* PYTHIA6 tune [28, 29] are used. The decay of tau-leptons is modeled using the TAUOLA program [30], while the emission of photons from finale state leptons uses PYTHIA6. The strong coupling constant α_s has been set to 0.130 at the Z-boson mass scale for all matrix element calculations. The effect of multiple interactions per bunch crossing (pile-up) has been simulated by overlaying MC-generated minimum bias events. The GEANT4 program was used to simulate the passage of particles through the CMS detector [31].

The simulated event samples are reweighted to describe the distribution of the number of pile-up conditions in the data by reweighting the ρ parameter distribution, where ρ denotes the diffuse offset energy density [32]. Moreover, a reweighting of the longitudinal position of the primary pp collision vertex of the MC samples to data has been performed.

5 Calibration

Even though the full detector simulation of CMS provides a very good description of the expected event signatures, some remaining differences in reconstruction, trigger and

Process	Dataset Name	σ [pb]
QCD (Inclusive, PYTHIA6)	QCD_Pt-15to3000_TuneZ2star [33]	$2 \cdot 10^{10}$
QCD (4-Jets, ALPGEN)	QCD4Jets_Pt-400to5600_TuneZ2Star [34]	–
$pp \to t\bar{t} + X \to 2l2\nu2b + X$	TTJets_FullLeptMGDecays_TuneP11TeV_8TeV [35]	112.3
$pp \to t\bar{t} + X \to 1l1\nu2q2b + X$	TTJets_SemiLeptMGDecays_8TeV [36]	107.2
$pp \to t\bar{t} + X \to 4q2b + X$	TTJets_HadronicMGDecays_TuneP11mpiHi_8TeV [37]	25.8
$pp \to WW + X \to 2l2\nu + X$	WWJetsTo2L2Nu_TuneZ2star_8TeV [38]	5.8
$pp \to WZ + X \to 3l1\nu + X$	WZJetsTo3LNu_8TeV_TuneZ2Star [39]	1.1
$pp \to Z/\gamma^* + X \to \mu^+\mu^- + X$	DYToMuMu_M-20_CT10_TuneZ2star_v2_8TeV [40]	1931

Table 2. Overview of data samples and simulated event samples used in this analysis together with the corresponding inclusive cross-sections. Leptonic decay (e, μ, τ) are denoted with l.

isolation efficiencies as well as in the momentum and energy scales and resolutions between MC simulation and data are present. Dedicated corrections are applied to minimize these differences for the relevant objects in this analysis and are discussed in the following.

The official CMS calibration and corrections for particles jets, in particular the jet energy scale (JES) and the jet energy resolution (JER), has been applied within the BACON framework. These jet corrections and uncertainties were derived from the simulation, and are confirmed with in situ measurements using the energy balance of dijet and photon+jet events [32]. A reduced set of systematic variations is used to estimate JES and JER uncertainties on the final measurement. In particular, the JES is variated by 2% for $y^{\mathrm{jet}} < 1.3$ and by 3% for $y^{\mathrm{jet}} > 1.3$, following [32]. The JER is varied by 20% for $30 < E_{\mathrm{T}}^{\mathrm{jet}} < 100\,\mathrm{GeV}$, by 10% for $100 < E_{\mathrm{T}}^{\mathrm{jet}} < 1\,\mathrm{TeV}$ and by 5% above. The absolute angular ϕ- and η-resolution for reconstructed jets with a transverse momentum of $50\,\mathrm{GeV}$ is between 0.01 and 0.02 and improves for larger p_{T} values. We assume a relative uncertainty of 50% on the both angular resolutions.

Events containing exactly two jets with $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ within $y^{\mathrm{jet}} < 2.4$ are used to test the calibration as well as the assigned uncertainties. A comparison of the $\Delta_{p_T}^{\mathrm{rel}} = (p_{\mathrm{T}}^{j_1} - p_{\mathrm{T}}^{j_2})/(p_{\mathrm{T}}^{j_1} + p_{\mathrm{T}}^{j_2})$ distributions for data and MC, including the assigned systematic uncertainties is shown in figure 3, where agreement between data and MC can be seen.

To further validate the applied corrections, Z-boson candidate events with an associated production of at least one jet have been selected in data and compared to full simulated Monte Carlo samples. $Z \to \mu^+\mu^-$ candidate events are selected by requiring events with exactly two oppositely charged, isolated muons with a minimal p_T of $25\,\mathrm{GeV}$ within $|\eta| < 2.1$. Jets are required to have a minimal transverse momentum of $p_T > 30\,\mathrm{GeV}$ within a rapidity range of $|y| < 2.4$. The invariant mass of these two lepton candidates has to be between 66 and $116\,\mathrm{GeV}$. This selection ensures a nearly background free selection of Z-boson candidates. In order to test the jet calibration, Z-boson events with exactly one reconstructed jet are further divided into evens with $20 < p_{\mathrm{T}}(Z) < 50\,\mathrm{GeV}$ and $50 < p_{\mathrm{T}}(Z) < 100\,\mathrm{GeV}$. The transverse momentum of the Z-boson, precisely measured by its decay leptons, should be balanced in a first approximation by the transverse energy of
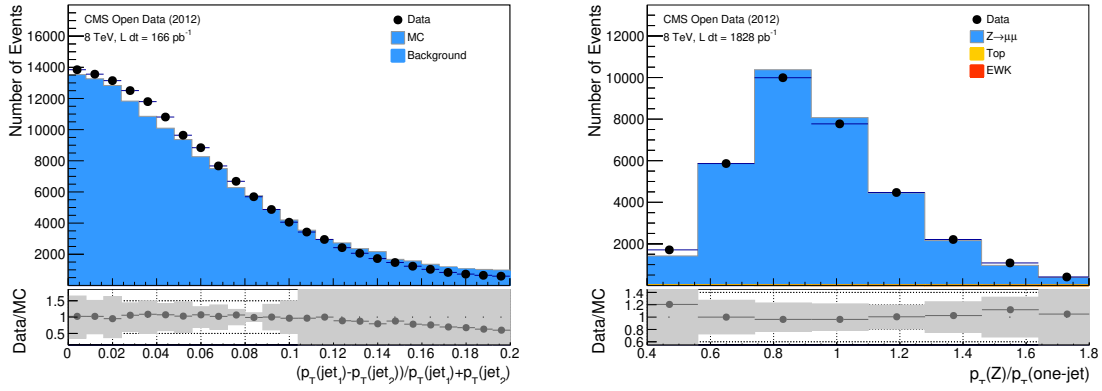
Figure 3. Left: distribution of the $\Delta_{p_T}^{\rm rel} = (p_T^{j_1} - p_T^{j_2})/(p_T^{j_1} + p_T^{j_2})$ in Data and MC. Right: comparison of the ratio of measured $p_T(Z)$ and the measured jet energy E_T for Z-boson events in the muon decay channel with exactly one jet with $E_T > 30\,{\rm GeV}$ and $20 < p_T(Z) < 50\,{\rm GeV}$ for Data and MC. All MC-corrections have been applied. The grey band indicates the corresponding systematic uncertainties.

the select jet, hence the ratio of $p_T(Z)/p_T^{\rm jet}$ should peak around 1. The comparison of data and MC of this ratio is also shown in figure 3, where a good agreement within the assigned systematic uncertainties can be seen. This study has been repeated in [20] with higher jet multiplicities and higher values of $p_T(Z)$ in the final state, all indicating good closure.

6 Analysis strategy, signal selection and background estimation

It is the goal of this analysis to probe a possible asymmetry between positive and negative values of α (defined in section 2) in events containing at least three jets in proton-proton collisions at a center of mass energy of $\sqrt{s} = 8\,{\rm TeV}$. The dependence of this asymmtry will be investigated in bins of the invariant di-jet mass of the leading and sub-leading jet, m_{12}. In addition, the normalized α distribution in data is measured and corrected for detector effects.

Events with at least three jets with a minimal transverse momentum of $p_T > 50\,{\rm GeV}$ within a rapidity range of $|y| < 2.4$ are selected for this analysis. Events with a fourth jet with $p_T > 30\,{\rm GeV}$ within $|y| < 3.0$ are vetoed. Standard noise cleaning cuts as well as cuts to reject jets from pile-up collisions are applied. In addition, it is required that the minimal difference between the transverse momentum of the leading and the sub-leading jet be at least $30\,{\rm GeV}$, i.e. $\Delta p_T^{j_1,j_2} = p_T^{j_1} - p_T^{j_2} > 30\,{\rm GeV}$, and that the minimal transverse momentum difference between the sub-leading and sub-sub-leading be at least $50\,{\rm GeV}$, i.e. $\Delta p_T^{j_2,j_3} = p_T^{j_2} - p_T^{j_3} > 50\,{\rm GeV}$. These requirements ensure that small changes to jet energies (e.g. from reconstruction effects) do not lead to changes in the p_T ordering of the jets, and so do not result in discrete changes in the orientation of the reference plane. Moreover, we veto fully back-to-back events by requiring that the $\Delta\Phi$ value between j_1 and j_2 is smaller than 3.0. This selection reduces uncertainty in the normal of the reference plane, since that normal becomes undefined when j_1 and j_2 are parallel. In total 29,063 events pass this selection. The selected events are further divided along the invariant

mass between the leading and sub-leading jet within the ranges $500 < m_{12} < 700\,\text{GeV}$, $700 < m_{12} < 900\,\text{GeV}$ and $900 < m_{12} < 2000\,\text{GeV}$ as well as for two rapidity regions of the three jet system, y_{jjj}, with $|y_{jjj}| < 0.7$ and $0.7 < |y_{jjj}|$.

Possible contributions to the selected final states involving electroweak decay processes, such as the production of W+jets and Z+jets in the hadronic decay mode or the production of $t\bar{t}$, are estimated by MC simulations and found to be below $0.5\,\%$, limited by the available MC statistics. As an additional cross-check, events with two jets with $p_T^{j_1} > 350\,\text{GeV}$ and $p_T^{j_2} > 100\,\text{GeV}$ and one additional reconstructed isolated muon with a transverse momentum of at least $50\,\text{GeV}$ have been selected. Those events are expected to come from W/Z+jets, $t\bar{t}$ or WW decays, with at least one vector boson decaying leptonically. In total, 31 events pass this selection, of which no events with two muons were found. Assuming that these muons all stem from leptonic W-boson decays, we expect ≈ 310 electroweak background events, i.e. a background contribution of 1%, which we take as an estimate in the following.

Both QCD MC samples (table 2) have been used to study migration effects, i.e. a change in the jet ordering or a wrong charge assignment of α using MC truth information. The minimal p_T difference between the jets as well as the $\Delta\Phi(j_1, j_2)$ requirement results in only 1.5% of events, in which the jet-p_T ordering is different between reconstruction and generator level and 1% of the cases, which lead to wrong sign-assignment of α.

In order to improve the description of the MC samples with data, the transverse momenta distribution of the leading and sub-leading jets have been reweighted to data by a two-dimensional approach on an event-by-event basis using all events containing three jets with $p_T > 50\,\text{GeV}$ within a rapidity range of $|y| < 2.4$. The reweighting is based on two dimensional histograms with the p_T of the leading and sub-leading jet on the axis, which is filled for data and MC. The actual event weights are derived by the ratio of the normalized data histogram over the corresponding normalized histogram of the MC set. It should be noted, that the fiducial volume which is used for the reweighting is significantly larger than the actual fiducial volume of the analysis. Figure 4 shows the reconstructed p_T distributions after the reweighting procedure of all three jets in all selected events as well as the inclusive distribution of α in data as well as MC simulations, where the distributions are normalized. Data and MC agree well within their uncertainties.

7 Raw asymmetry measurements

The α asymmetry, A_α, can be defined as the difference between the sum over the fiducial cross-section in each measurement bin i of α for positive and negative regions of α, normalized by the overall cross section, i.e.

$$A_\alpha = \frac{\sum_{i,\alpha>0} d\sigma_i/\sigma^{\text{fid}} - \sum_{i,\alpha>0} d\sigma_i/\sigma^{\text{fid}}}{\sum_{i,\alpha>0} d\sigma_i/\sigma^{\text{fid}} + \sum_{i,\alpha>0} d\sigma_i/\sigma^{\text{fid}}} = \frac{\sum_{i,\alpha>0} d\sigma_i/\sigma^{\text{fid}} - \sum_{i,\alpha>0} d\sigma_i/\sigma^{\text{fid}}}{\sigma} \quad (7.1)$$

The asymmetry A_α can be approximated by a pure data driven quantity A_α^{data}, that reduces to a pure counting exercise:

$$A_\alpha^{\text{data}} = \frac{N_{\alpha<0} - N_{\alpha>0}}{N_{\alpha<0} + N_{\alpha>0}}, \quad (7.2)$$
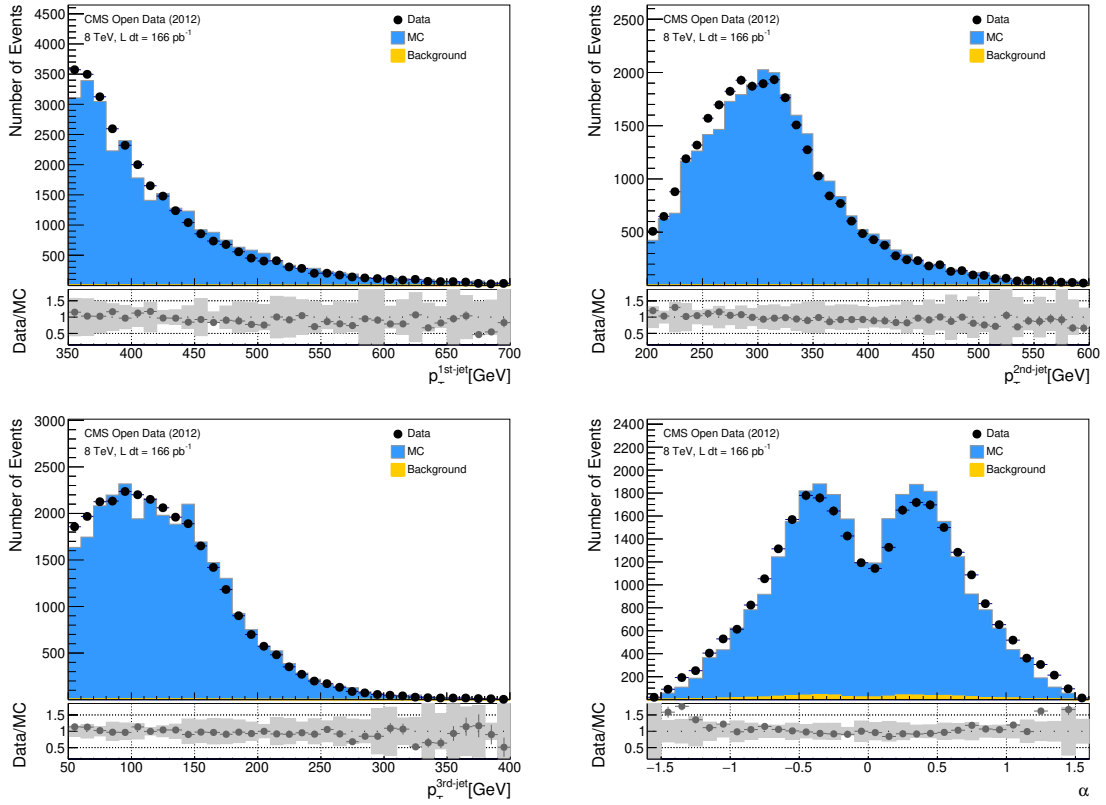
Figure 4. Comparison of the p_T distributions for the leading (upper left), sub-leading (upper right) and sub-sub-leading jets (lower left) for data and MC, as well as the inclusive α distribution (lower right) for data and MC. The grey band indicates the corresponding systematic uncertainties.

where N is the number of events in a particular region of α. In the following we assume that A_α^{data} is, to a good approximation, equal to A_α, since detector related resolution and efficiency effects are expected to be symmetric in α, i.e. affect events with positive and negative values of α in the same way and hence cancel in the ratio. This was tested and confirmed using MC simulations. The resulting asymmetries A_α for the full inclusive selection, as well as for the three invariant mass regions of the leading and sub-leading jet, are shown in table 3. The separation in different invariant di-jet masses allows to probe energy dependencies of parity violating effects. All asymmetries are compatible with the Standard Model expectation of zero, within the statistical uncertainties.

A partial asymmetry can be calculated for different regions of α, for example by comparing the number of events in $[-0.96, -0.64]$ to the number of events in $[0.64, 0.96]$. This is graphically illustrated for the inclusive selection as well as the three m_{jj} mass-ranges and two y_{jjj} rapidity ranges in figure 5 and figure 6, respectively. While the upper part of each plot shows the actual number of reconstructed events with positive and negative values of α, the lower part shows the resulting partial asymmetry including the statistical uncertainty. In general a good agreement with the SM expectation is seen and no deviation above 2.5σ is observed.

Selection	Number of Events ($\alpha < 0$)	Number of Events ($\alpha > 0$)	Asymmetry A_α
m_{jj} inclusive	14661	14402	0.009 ± 0.006
$500 < m_{jj} < 700\,\mathrm{GeV}$	3211	3189	0.003 ± 0.012
$700 < m_{jj} < 900\,\mathrm{GeV}$	5615	5483	0.011 ± 0.009
$900 < m_{jj} < 2000\,\mathrm{GeV}$	5646	5551	0.008 ± 0.009
$|y_{jjj}| < 0.7$	10192	9993	0.010 ± 0.007
$0.7 < |y_{jjj}|$	4469	4409	0.007 ± 0.011

Table 3. Measured asymmetry of α for the inclusive selection as well as three invariant mass regions of the leading and sub-leading jet as well as two rapidity regions of the three jet system y_{jjj}. The uncertainties are statistical only.
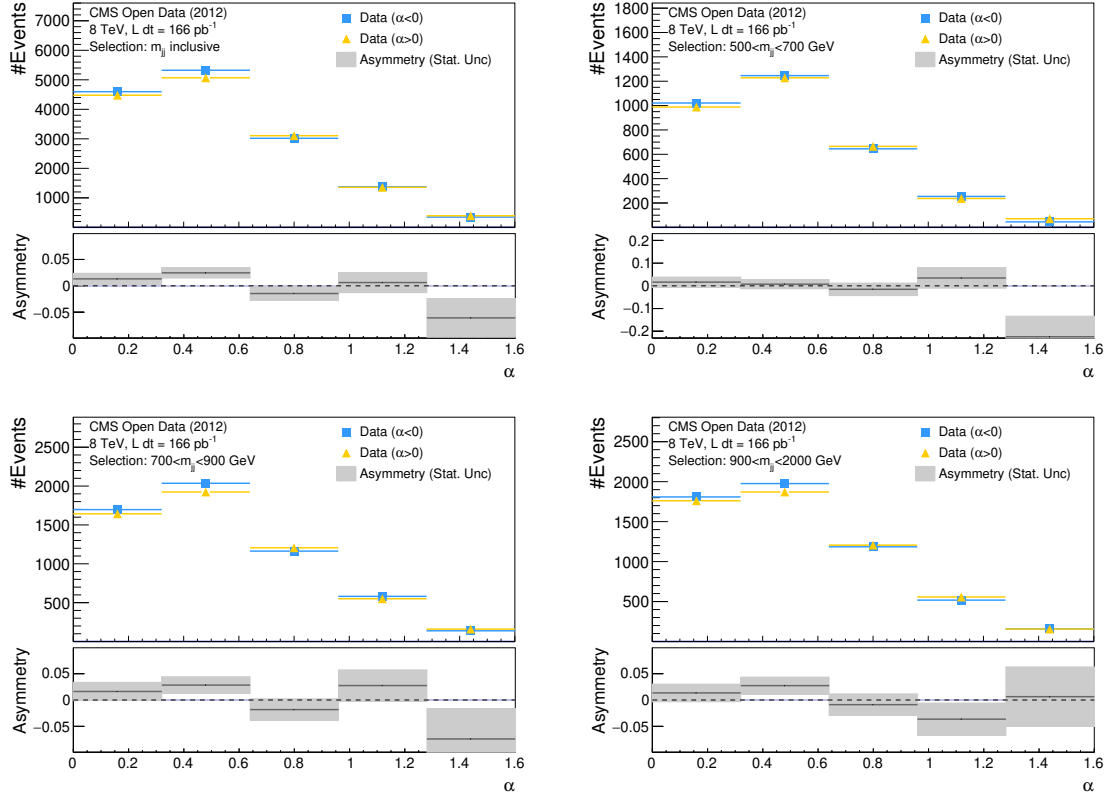
Figure 5. Distribution of reconstructed events with $\alpha < 0$ and $\alpha > 0$ for all events (upper right), as well as three regions in the invariant mass m_{12}; upper right: $0 < m_{12} < 700\,\mathrm{GeV}$, lower left: $700 < m_{12} < 900\,\mathrm{GeV}$, lower right: $900 < m_{12} < 2000\,\mathrm{GeV}$. The ratio in each plots shows the asymmetry value including statistical uncertainties.
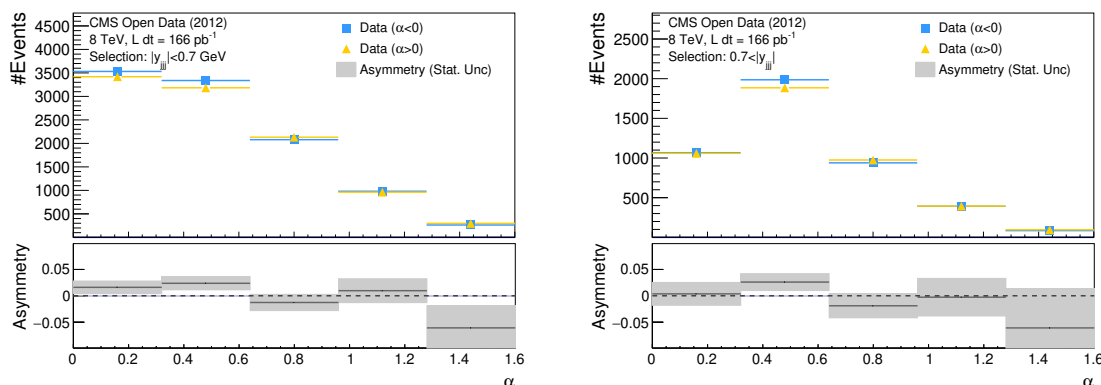
Figure 6. Distribution of reconstructed events with $\alpha < 0$ and $\alpha > 0$ for events with $|y_{jjj}| < 0.7$ (left) and $0.7 < |y_{jjj}|$ (right). The ratio in each plots shows the asymmetry value including statistical uncertainties.

8 Results corrected for detector effects

Jets have also been defined at the generator level, by clustering final-state particles with decay length $c\tau$ ¿ 10 mm, using the anti-k_t algorithm [10] with radius parameter R = 0.5. The generator-level jets are used to define the fiducial volume of this measurement, by selecting events fulfilling the same kinematic requirements at the generator level as at the reconstruction level.

The normalized three-jet production cross-section, is measured as a function of α in the binning $[-1.60, -1.28, -0.96, -0.64, -0.32, 0.00, 0.32, 0.64, 0.96, 1.28, 1.60]$.[12]

The normalized differential distribution within the fiducial volume is corrected for detector effects and bin-to-bin migrations using an iterative Bayesian unfolding method, with two iterations [41, 42]. First, the data are corrected for background contributions as well as events that pass the detector-level selection but not the particle-level selection. Then, the iterative Bayesian unfolding technique is used as a regularized matrix inversion to correct for the detector resolution in events that pass both the detector-level and particle-level selections. The response matrix (figure 7), which connects the distribution at reconstruction and particle level is estimated using the inclusive QCD MC sample. After the application of the response matrix, a final correction is applied to account for events that pass the particle-level but not detector-level selection. The chosen binning of α ensures a purity of over 90% in each bin.

Statistical uncertainties are estimated by toy variations of the input data statistics, systematic uncertainties are estimated by varying each source of uncertainty and repeating the full unfolding procedure. For the studies of the systematic effects of JES and JER related uncertainties, several scenarios are considered: in the full-correlation scenario, we apply the systematic uncertainties equally to all reconstructed jet candidates of each event. In the no-correlation scenario, we apply the systematic variations to each jet individually,

---

[12]The bin boundaries are at exactly at the decimal values listed. Since α cannot exceed $\pi/2 \approx 1.57$ there are no overflow bins.
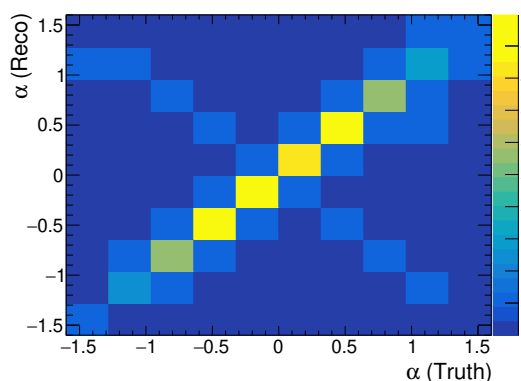
Figure 7. Unfolding Matrix used in the iterative unfolding procedure, where the truth α values are shown on the x-axis and the corresponding reconstructed values on the y-axis.
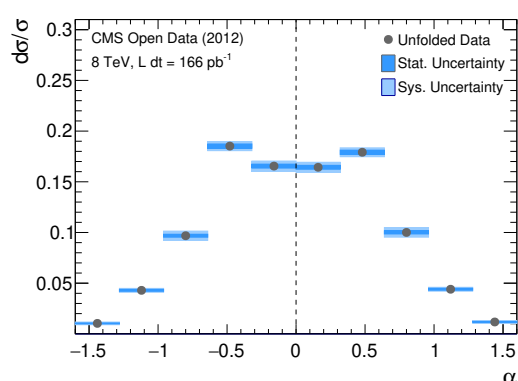
Figure 8. Unfolded, detector-corrected α distribution with statistical and systematic uncertainties. For a conservative interpretation, the uncertainties should treated as uncorrelated.

but leave the other jet candidates of the event unaffected.[13] The raw-calibration scenario is based on jet kinematics, without any applied calibrations. The difference to the nominal results is calculated for each scenario and taken as systematic. The systematic uncertainties due to the angular resolutions are estimated by adding an additional smearing on the reconstructed η and ϕ values, following section 5, and then repeating the unfolding procedure. The modeling bias of the unfolding procedure has been tested in two different ways. First, the MC truth level distribution of α was reweighted to the data distributions and then the unfolding procedure repeated. Second, the 4-jet QCD sample was used as a MC prior. The differences between the nominal and the alternative result was taken as systematic uncertainty and added in quadrature.

The normalized unfolded distributions of α are shown in figure 8 together their overall uncertainties. The numerical values, including a detailed split-up of the associated uncertainties is given in table 4. The total uncertainty ranges between 3% for the central bins to 13% for the bins corresponding to large absolute values of α. While the central bins are dominated by uncertainties of the jet energy scale systematics, the bins for $|\alpha| > 1$ are dominated by the limited statistics of data and MC as well as modeling uncertainties. It is interesting to note, that these modeling uncertainties are not due from migration effects of α, rather than from efficiency corrections that have to be applied between the reconstruction and truth fiducial volumes.

9 Conclusion

We have proposed a way of constraining non-standard sources of parity violation through measurements of the angular distribution, α, of the radiation angle of the lowest energetic jet in three-jet events, and have trialled proposal with using CMS open data proton-proton collisions at a center of mass energy of 8 TeV. Potential parity violating effects beyond the

---

[13]Hence three variations of the results are derived, corresponding to the individual changes to each of the three jets.

Bin	Value	Stat.	MC-Stat.	Model	Jet (Uncor.)	Jet (Cor.)	Jet η Res.	Jet ϕ Res.	Total
(-1.60,-1.28)	0.010 ± 0.001	4.3%	4.9%	7.1%	7.2%	5.3%	2.5%	0.8%	13.5%
(-1.28,-0.96)	0.043 ± 0.002	2.2%	2.2%	0.5%	2.2%	2.8%	0.4%	1.1%	5.0%
(-0.96,-0.64)	0.097 ± 0.005	1.5%	1.4%	3.3%	2.8%	0.0%	0.5%	0.6%	4.9%
(-0.64,-0.32)	0.185 ± 0.005	1.1%	1.0%	0.7%	1.0%	1.7%	0.3%	0.1%	2.6%
(-0.32,0.00)	0.165 ± 0.005	1.1%	1.2%	1.7%	1.9%	0.8%	0.3%	0.1%	3.2%
(0.00,0.32)	0.164 ± 0.005	1.1%	1.1%	1.7%	1.9%	0.8%	0.3%	0.1%	3.2%
(0.32,0.64)	0.179 ± 0.005	1.1%	1.0%	0.7%	1.0%	1.7%	0.3%	0.1%	2.7%
(0.64,0.96)	0.100 ± 0.005	1.4%	1.4%	3.3%	2.8%	0.0%	0.5%	0.6%	4.9%
(0.96,1.28)	0.044 ± 0.002	2.2%	2.1%	0.4%	2.1%	2.8%	0.5%	1.1%	4.9%
(1.28,1.60)	0.012 ± 0.001	4.1%	4.6%	6.8%	7.1%	5.2%	2.3%	0.8%	13.0%

Table 4. Unfolded normalized distribution of α in the fiducial volume including and a summary of statistical and systematic uncertainty.

Standard Model could lead to an asymmetry in the α distribution, which can be quantified by an asymmetry parameter, A_α, which is expected to the zero in the SM. A_α was measured on data in three different kinematic regions, yielding to values which are consistent with zero. No obvious experimental challenges or limitations have been encountered, hence the sensitivity of this measurement could be dramatically improved with higher statistics on data and a better estimate of the corresponding experimental uncertainties. The models to which it is sensitive to could be changed by moving to other variables which probe the final state only, such as the variables δ and S_1 to S_4 defined in equations (2.3) to (2.7). The limitations of the unfolded distribution can be overcome by larger MC samples as well as a more careful evaluation of the associated model uncertainties. We therefore encourage LHC collaborations to perform similar studies in upcoming full Run-2 analysis a center of mass energy of 13 TeV.

References

[1] C.S. Wu, E. Ambler, R.W. Hayward, D.D. Hoppes and R.P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, *Phys. Rev.* **105** (1957) 1413 [ɪɴSPIRE].

[2] R.L. Garwin, L.M. Lederman and M. Weinrich, *Observations of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: The Magnetic Moment of the Free Muon*, *Phys. Rev.* **105** (1957) 1415 [ɪɴSPIRE].

[3] M. Goldhaber, L. Grodzins and A.W. Sunyar, *Helicity of Neutrinos*, *Phys. Rev.* **109** (1958) 1015 [ɪɴSPIRE].

[4] T. Han and Y. Li, *Genuine CP-odd Observables at the LHC*, *Phys. Lett.* **B 683** (2010) 278 [arXiv:0911.2933] [ɪɴSPIRE].

[5] ATLAS collaboration, *Measurement of the forward-backward asymmetry of electron and muon pair-production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09** (2015) 049 [arXiv:1503.03709] [ɪɴSPIRE].

[6] CMS collaboration, *Forward-backward asymmetry of Drell-Yan lepton pairs in pp collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C 76** (2016) 325 [arXiv:1601.04768] [ɪɴSPIRE].

[7] CMS collaboration, *The CMS Experiment at the CERN LHC*, 2008 *JINST* **3** S08004 [ɪɴSPIRE].

[8] CMS collaboration, *Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV*, CMS-PAS-MUO-10-002 (2010).

[9] CMS collaboration, *Measurement of inclusive W and Z boson production cross sections in pp collisions at $\sqrt{s} = 8$ TeV*, *Phys. Rev. Lett.* **112** (2014) 191802 [arXiv:1402.0923] [ɪɴSPIRE].

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

[11] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [arXiv:1111.6097] [ɪɴSPIRE].

[12] CMS collaboration, *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*, CMS-PAS-PFT-10-001 (2010).

[13] CMS collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus and MET*, CMS-PAS-PFT-09-001 (Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus and MET).

[14] CMS collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, 2017 *JINST* **12** P10003 [arXiv:1706.04965] [ɪɴSPIRE].

[15] M. Cacciari and G.P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett.* **B 659** (2008) 119 [arXiv:0707.1378] [ɪɴSPIRE].

[16] CMS collaboration, *Software Framework for CMS Open Data Analysis*, http://opendata.cern.ch/docs/about-cms (2017).

[17] CMS collaboration, *CMS software version 5_3_32* (CMSSW_5_3_32), CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.WYJG.FYK9 (2016).

[18] C. Vernieri et al., *Bacon analysis framework*, https://github.com/ksung25/BaconProd/tree/Run1 (2015).

[19] CMS collaboration, *CMS list of validated runs for primary datasets of 2012 data taking*, CERN Open Data Portal, http://opendata.cern.ch/record/1002 (2017).

[20] A. Apyan, W. Cuozzo, M. Klute, Y. Saito, M. Schott and B. Sintayehu, *Opportunities and Challenges of Standard Model Production Cross Section Measurements at $\sqrt{s} = 8\,TeV$ using CMS Open Data*, arXiv:1907.08197 [INSPIRE].

[21] CMS collaboration, *JetHT primary dataset in AOD format from run of 2012* (/JetHT/Run2012C-22Jan2013-v1/AOD), CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.VZSR.LYZX (2017).

[22] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[23] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A.D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003) 001 [hep-ph/0206293] [INSPIRE].

[24] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO vector-boson production matched with shower in POWHEG*, *JHEP* **07** (2008) 060 [arXiv:0805.4802] [INSPIRE].

[25] 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] [INSPIRE].

[26] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: Going Beyond*, *JHEP* **06** (2011) 128 [arXiv:1106.0522] [INSPIRE].

[27] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, *Phys. Rev.* **D 89** (2014) 033009 [arXiv:1302.6246] [INSPIRE].

[28] CMS collaboration, *Study of the Underlying Event at Forward Rapidity in pp Collisions at $\sqrt{s} = 0.9$, 2.76 and 7 TeV*, *JHEP* **04** (2013) 072 [arXiv:1302.2394] [INSPIRE].

[29] CMS collaboration, *Event generator tunes obtained from underlying event and multiparton scattering measurements*, *Eur. Phys. J.* **C 76** (2016) 155 [arXiv:1512.00815] [INSPIRE].

[30] N. Davidson, G. Nanava, T. Przedzinski, E. Richter-Was and Z. Was, *Universal Interface of TAUOLA Technical and Physics Documentation*, *Comput. Phys. Commun.* **183** (2012) 821 [arXiv:1002.0543] [INSPIRE].

[31] GEANT4 collaboration, *GEANT4: A Simulation toolkit*, *Nucl. Instrum. Meth.* **A 506** (2003) 250 [INSPIRE].

[32] CMS collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, 2017 *JINST* **12** P02014 [arXiv:1607.03663] [INSPIRE].

[33] CMS collaboration, *Simulated dataset* QCD-Pt-15to3000-TuneZ2star-Flat-8TeV-pythia6 *in AODSIM format for 2012 collision data*, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.7Y4S.93A0 (2017).

[34] CMS collaboration, *Simulated dataset* QCD4Jets-Pt-400to5600-TuneZ2Star-8TeV-alpgen *in AODSIM format for 2012 collision data*, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.P2XT.ZX19 (2017).

[35] CMS collaboration, *Simulated dataset* TTJets-FullLeptMGDecays-TuneP11TeV-8TeV -madgraph-tauola *in* AODSIM *format for 2012 collision data*, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.7RZ3.0BXP (2017).

[36] CMS collaboration, *Simulated dataset* `TTJets-SemiLeptMGDecays-8TeV-madgraph` *in AODSIM format for 2012 collision data*, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.FZCE.MBDW (2017).

[37] CMS collaboration, *Simulated dataset* `TTJets-HadronicMGDecays-TuneP11mpiHi-8TeV -madgraph-tauola` *in AODSIM format for 2012 collision data*, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.B91N.86OR (2017).

[38] CMS collaboration, *Simulated dataset* `WWJetsTo2L2Nu-TuneZ2star-8TeV-madgraph-tauola` *in AODSIM format for 2012 collision data*, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.V2C6.O1P4 (2017).

[39] CMS collaboration, *Simulated dataset* `WZJetsTo3LNu-8TeV-TuneZ2Star-madgraph-tauola` *in AODSIM format for 2012 collision data*, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.71R9.VLZA (2017).

[40] CMS collaboration, *Simulated dataset* `DYToMuMu-M-20-CT10-TuneZ2star-v2-8TeV-powheg-pythia6` *in AODSIM format for 2012 collision data*, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.QGC3.PTZ9 (2017).

[41] G. D'Agostini, *A Multidimensional unfolding method based on Bayes' theorem*, *Nucl. Instrum. Meth.* **A 362** (1995) 487 [INSPIRE].

[42] G. D'Agostini, *Improved iterative Bayesian unfolding*, in *Alliance Workshop on Unfolding and Data Correction*, Hamburg, Germany, 27–28 May 2010 (2010) [arXiv:1010.0632] [INSPIRE].