CMS Physics Analysis Summary

Contact: cms-pag-conveners-top@cern.ch 2022/05/14

Measurement of the $t\bar{t}$ charge asymmetry in highly boosted events in the single-lepton channel at 13 TeV

The CMS Collaboration

Abstract

The measurement of the charge asymmetry for highly boosted top quark pairs decaying to a single lepton and jets is presented. The analysis is performed using 138 fb ^{1} of data collected in pp collisions at $\bar{s}$ 13 TeV with the CMS detector during Run 2 of the Large Hadron Collider. The selection is optimized for top quark-antiquark pairs produced with large Lorentz boosts, resulting in non-isolated leptons and overlapping jets. The top quark charge asymmetry is measured for events with $t\bar{t}$ invariant mass larger than 750 GeV and corrected for detector and acceptance effects using a binned maximum likelihood fit. The measured top quark charge asymmetry is in good agreement with the standard model prediction at next-to-next-to-leading order in perturbation theory with next-to-leading order electroweak corrections. Differential distributions for two invariant mass ranges are also presented.

c 2022 CERN for the benefit of the CMS Collaboration. CC-BY-4.0 license

1 Introduction

Top quarks are produced at hadron colliders primarily in the form of $t\bar{t}$ pairs that originate from a $gt\bar{t}$ vertex via the strong interaction [1, 2]. A fundamental difference between the $t\bar{t}$ production in the Tevatron $p\overline{p}$ collisions and the LHC pp collisions is that the former is dominated by $q\overline{q}$ annihilation and the latter by gluon fusion. At leading order, the standard model (SM) predicts that the $t\bar{t}$ production is forward-backward symmetric in $q\overline{q}$ annihilation. However, higher-order SM effects result in a small (6.6%) positive forward-backward asymmetry A_{FB}, such that the top quark (antiquark) is preferentially emitted in the direction of the incoming quark (antiquark) [3]. There is no asymmetry in the gluon fusion $t\bar{t}$ production that dominates at the LHC, but because valence quarks carry, on average, larger momentum than antiquarks (from the sea), the rapidity distribution of top quarks is expected to be broader than that of top antiquarks [4, 5]. The central-forward $t\bar{t}$ charge asymmetry is defined as

$$A_C \quad \frac{N \quad y \quad 0 \quad N \quad y \quad 0}{N \quad y \quad 0 \quad N \quad y \quad 0}, \tag{1}$$

where $y \quad y_t \quad y_{\bar{t}}$ is the difference between the absolute value of the top quark and top anti-quark rapidities. This distribution could be modified by beyond the standard model (BSM) production mechanisms that exchange new bosons. A_C is expected to be on the order of 1% in the SM.

Since the relative contribution of valence quarks increases at high momentum transfer [6], we expect that measuring A_C in a sample of highly boosted $t\bar{t}$ events will lead to a more stringent probe of quantum chromodynamics (QCD) predictions and higher sensitivity to BSM physics processes that might alter the charge asymmetry [7]. Several models predict enhancements with respect to the SM prediction in the presence of new particles, including axigluons [8, 9], Z bosons [10–12], and models that predict the existence of heavy W bosons, scalar isodoublets, color triplet scalars, and color sextet scalars [13, 14]. These models introduce new spin-0 and spin-1 particles in the interaction, modifying A_C by exchanging the new particles through interference terms and dedicated loops. Along with specific BSM models, deviations from the SM prediction can also be explained through an effective field theory (EFT) approach in which new physics contributions are described via a fixed set of dimension-six operators added to the SM Lagrangian [15].

Early Tevatron A_{FB} measurements [16, 17], based on about half of the full data set, sparked wider interest when they showed larger asymmetries than those predicted by the SM at the time [18]. These measurements were especially interesting because the discrepancies grew with larger top quark pair masses and rapidity differences. Measurements using the full Tevatron data set and combining the results from the two collaborations, CDF and D0, later became available [19]. Even though all measurements favored larger positive asymmetries than the predictions [20], none of the observed differences were larger than 2 standard deviations. The ATLAS and CMS collaborations have combined their inclusive and differential measurements of A_C at two center-of-mass energies (7 and 8 TeV), obtaining $A_C \quad 0.005 \quad 0.007\,(\text{stat}) \quad 0.006\,(\text{syst})$ and $A_C \quad 0.0055 \quad 0.0023\,(\text{stat}) \quad 0.0025\,(\text{syst})$ at 7 and 8 TeV, respectively. The two independent uncertainties reported are the statistical uncertainty in the observed data, labeled "stat", and the systematic uncertainty, labeled "syst". These combined measurements show good agreement with the respective SM predictions and uniquely restrict the phase space of possible BSM phenomena that would produce large asymmetries [13].

The measurement presented in this publication is the first one that uses proton-proton colli-

sion data at $\bar{s}$ 13 TeV and optimizes the reconstruction for highly Lorentz-boosted $t\bar{t}$ events with invariant mass above 750 GeV. We target the single-lepton channel, in which both top quarks decay as t bW, one W boson decays leptonically (W), and the other decays hadronically (W $q\bar{q}$). The highly boosted top quarks yield collimated decay products that are partially or fully merged. For the leptonically decaying top quark, this results in the lepton (electron or muon) appearing as non-isolated because of its proximity to the b quark. Dedicated jet and lepton cleaning at the trigger and offline levels allows us to reconstruct the decay products of the boosted leptonically-decaying top quarks without applying an isolation requirement on the leptons, while the multijet background is controlled with topological requirements [21]. The behavior of hadronically decaying top quarks depends on the magnitude of their transverse momentum (p_T). At the high end of the p_T spectrum, the top quark decay products have angular distances between partons that can be smaller than the jet clustering distance parameter and are thus reconstructed as a single, large-radius jet. In contrast, at the low end of the p_T near the kinematic threshold, each parton is matched to a single jet. For intermediate p_T values, the partons from the hadronic W decay are merged into a single jet, but the b quark is reconstructed separately. All three topologies are considered in this analysis, and are referred to as "Boosted", "Semi-resolved", and "Resolved".

2 Apparatus and object reconstruction

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity ( ) coverage provided by the barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. Events of interest are selected using a two-tiered trigger system [22, 23]. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [24].

The offline event reconstruction is based on a particle-flow algorithm [25], which combines information from each subdetector to identify electrons, photons, and charged or neutral hadrons. To recover inefficiencies observed in data for very high p_T particle-flow muons, we use muons that are reconstructed first in the muon system and then fitted to tracks [26]. The primary vertex (PV) is taken to be the vertex corresponding to the hardest scattering in the event, evaluated using tracking information alone, as described in Section 9.4.1 of Ref. [27]. Charged hadrons associated with other vertices are removed from further consideration. The remaining particle-flow candidates are clustered into jets using the anti-k_T algorithm [28, 29] with distance parameters of 0.4 (AK4) and 0.8 (AK8). If a lepton is found with R $\sqrt{^2 \quad ^2}$ 0.4 with an AK4 jet, where is the azimuthal angle, its four-momentum is subtracted from that jet [21]. The missing transverse momentum vector p_T^{miss} is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event, and its magnitude is denoted as p_T^{miss} [30]. Corrections are applied to improve the jet energy scale and resolution, and the p_T^{miss} is modified to account for these corrections [31].

Specialized techniques use AK8 jets and jet substructure information, including the mass of the leading three subjets [32] and the number of subjets [33], to separate the hadronic decay of boosted top quarks into two exclusive categories: hadronically decaying top quarks (t-tag) in which the three partons are merged into a single AK8 jet, and hadronically decaying W bosons (W-tag) in which the two partons from the W boson are merged into a single AK8 jet, but the

bottom quark is reconstructed as a separate AK4 jet. The identification of jets originating from the decay of B-hadrons (b-tag) employs a deep neural network multi-classification algorithm that relies on information from the tracker and the calorimeters [34]. The b-tagging algorithm is applied to AK4 jets with $p_{\mathrm{T}}^{j_{\mathrm{AK4}}}$ 50 GeV and $^{j_{\mathrm{AK4}}}$ 2.5; the t-tagging and W-tagging algorithms are applied to AK8 jets with $p_{\mathrm{T}}^{j_{\mathrm{AK8}}}$ 400 GeV and $^{j_{\mathrm{AK8}}}$ 2.5.

3 Collider data and simulated samples

We analyze data collected by the CMS detector during Run 2 (2018, 2017, and 2016) and corresponding to a total integrated luminosity of 138 fb ^{1} [35–37]. Events in the muon channel (jets) are selected with a single muon trigger that requires p_{T} 50 GeV. Events in the electron channel (e jets) are selected by either a single electron trigger with $p_{\mathrm{T}}^{\mathrm{e}}$ 115 GeV or a trigger requiring one electron with $p_{\mathrm{T}}^{\mathrm{e}}$ 50 GeV and one jet with p_{T}^{j} 165 GeV. As the e jets trigger was not available during the early running period in 2017, the integrated luminosity available for the 2017 e jets channel is reduced by 5 fb ^{1}.

In the offline reconstruction, we select events for the jets (e jets) channel that contain exactly one muon with p_{T} 55 GeV and 2.4 (one electron with $p_{\mathrm{T}}^{\mathrm{e}}$ 80 GeV and $^{\mathrm{e}}$ 2.5) and at least two jets with $p_{\mathrm{T}}^{j_1}$ 150 GeV ($p_{\mathrm{T}}^{j_1}$ 185 GeV), $p_{\mathrm{T}}^{j_2}$ 50 GeV, and ^{j} 2.4. To preserve the identification efficiency of $\mathrm{t\bar{t}}$ decay products in the highly boosted topology, no isolation requirement is imposed on the leptons either at trigger or offline level. To reduce the background from QCD multijet events, we apply a two-dimensional (2D) selection that requires leptons to satisfy the condition R_{min} , j 0.4 or $p_{\mathrm{T,rel}}$, j 25 GeV, where R_{min} , j is the minimum R between the lepton and all AK4 jets, and $p_{\mathrm{T,rel}}$, j is the transverse momentum of the lepton with respect to the axis of the nearest AK4 jet with p_{T} 25 GeV and 2.4 [21]. Finally, events need to satisfy $p_{\mathrm{T}}^{\mathrm{miss}}$ 50 GeV and $p_{\mathrm{T}}^{\mathrm{miss}}$ p_{T} 150 GeV ($p_{\mathrm{T}}^{\mathrm{miss}}$ 120 GeV) in the jets (e jets) channel. To suppress the contribution from the W jets background, at least one of the AK4 jets has to be b-tagged.

Monte Carlo (MC) samples for the $\mathrm{t\bar{t}}$ and single top quark processes (ST) are produced with the next-to-leading-order (NLO) POWHEG [38] generator. W jets, Drell-Yan (DY) Z jets and jets, and QCD multijet processes are generated with MADGRAPH5_aMC@NLO [39]. All samples are interfaced to PYTHIA8 [40] for parton showering with the CP5 tune [41]. Vector boson pair (Diboson) events are simulated with PYTHIA8. All samples include the simulation of additional inelastic pp interactions (pileup) within the same bunch crossing and additional contributions from the previous and next bunch crossings. Small corrections are applied to all MC samples to improve the agreement with the observed data, derived from data control samples that are orthogonal to the candidate selection.

4 Event reconstruction

The $\mathrm{t\bar{t}}$ system is reconstructed by assigning the four-vectors of the final-state objects to either the leptonic t or hadronic $\mathrm{t_h}$ leg of the $\mathrm{t\bar{t}}$ decay. For events with a t-tag, the t-tagged jet is taken as the $\mathrm{t_h}$ and only AK4 jets with R 0.8 from the $\mathrm{t_h}$ are considered candidates for the t . For events with a W-tag, the W-tagged jet is assigned to the $\mathrm{t_h}$. AK4 jets with R 0.8 from the W-tag can be assigned to either the t or the $\mathrm{t_h}$. For events with no t-tag and no W-tag, all possible assignments of AK4 jets are considered for both the t and the $\mathrm{t_h}$. No b-tagging information for individual jets is used in this process. Finally, one $\mathrm{t\bar{t}}$ hypothesis is selected for each event as the one that minimizes a ^{2} variable that minimizes the difference between

the reconstructed t_ℓ and t_h masses and the true top mass (within uncertainties). Background processes typically result in large values of χ^2 and are rejected from the signal selection.

The signal candidate sample is separated into three topologies based on the presence of t-tagged or W-tagged jets: "Boosted" contains events with one t-tag and no W-tag; "Semi-resolved" contains events with one W-tag and no t-tag; and "Resolved" contains the rest of the events that have no t-tag and no W-tag. Only events with an invariant mass of the top quark-antiquark pair ($M_{t\bar{t}}$) greater than 750 GeV are retained for further study. Figure 1 shows comparisons between data and the SM prediction for events in our candidate sample, where the boosted nature of the events becomes evident: the $M_{t\bar{t}}$ range extends to multi-TeV values, events with 2 and 3 jets are reconstructed, and events with leptons closer to the nearest jet axis than the jet size are retained. Good agreement between prediction and data is observed in all cases.

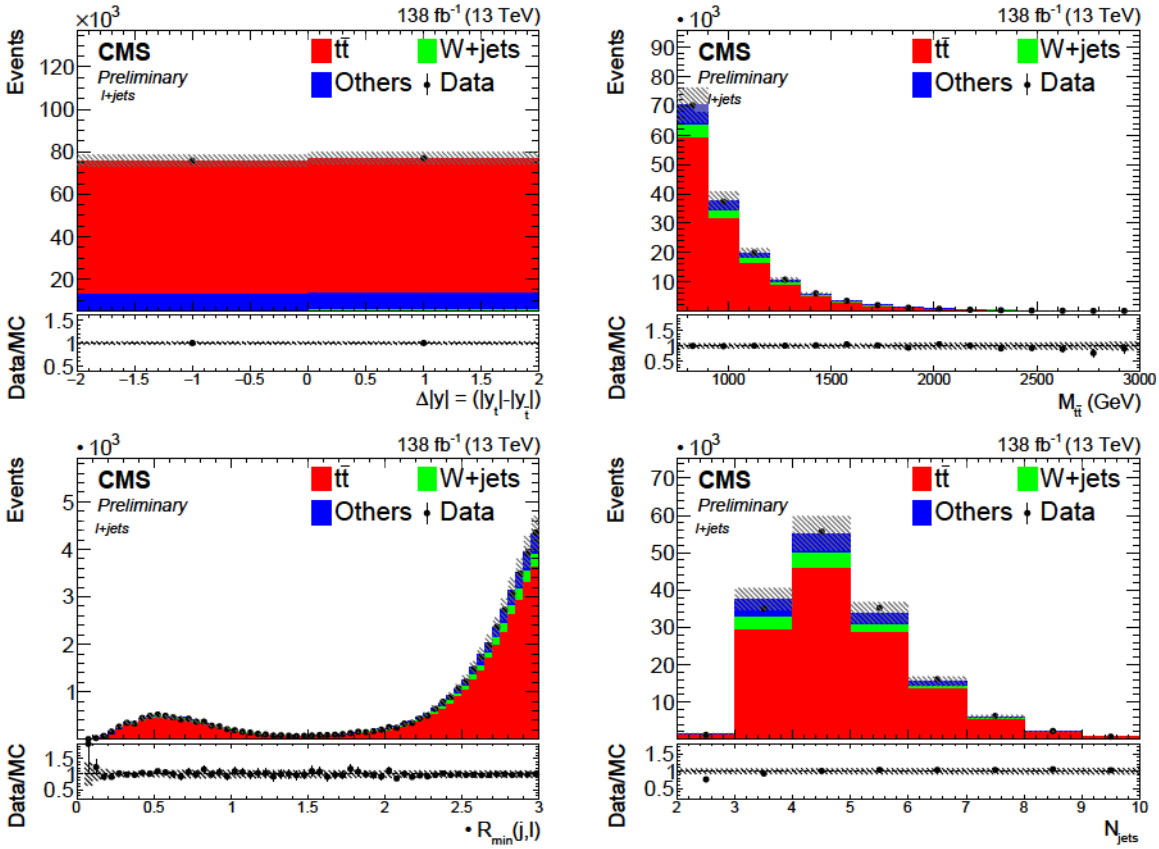

Figure 1: Comparison between data and SM prediction for the events in the signal candidate sample in the combined ℓ + jets channel after the likelihood normalization (see Section 6) for several quantities: $\Delta|y|$ (top left), reconstructed $M_{t\bar{t}}$ (top right), distance between the lepton and the closest AK4 jet $\Delta R_{\text{min}}(\ell, j)$ (bottom left), and the number of AK4 jets (bottom right). Data points are shown with their statistical uncertainty. The shaded band combines the MC statistical uncertainty and the systematic uncertainty (see Section 5). Overall, good agreement between data and simulation is observed in all variables.

5 Systematic uncertainties

Systematic uncertainties from numerous sources can affect the normalization and the shape of the distributions of physical observables in both signal and background simulated samples.

The systematic uncertainties affecting only the normalization come from the SM theoretical cross section values for each process and the luminosity normalization. All MC samples are normalized according to their respective SM cross section values and assigned a rate uncertainty of 30% for background processes and 5% for the $t\bar{t}$ signal. Additionally, uncertainties in the integrated luminosity vary per year—2.5%, 2.3%, and 2.5% for 2016, 2017, and 2018 respectively— and include both correlated and uncorrelated components across the three years.

All other systematic uncertainties affect both the normalization and the shape. Uncertainties from experimental sources are applied to both signal and background samples. All MC samples are reweighted to match the pileup distribution in data, which is generated by using the instantaneous luminosity per bunch crossing for each luminosity section, with a minimum bias cross section of 69.2 mb; an uncertainty of 4.6% is applied to this value. All electrons and muons in the simulated samples have uncertainties associated with the trigger and the reconstruction and identification. These uncertainties are uncorrelated across lepton flavors but correlated across years and are parameterized as a function of the p_{T} and of the leptons. There is a flat uncertainty in the efficiency of the 2D selection that rejects QCD background, which is uncorrelated across lepton flavors and years. Uncertainties in the jet energy scale (JES) and resolution (JER) are parameterized in terms of the jet p_{T} and and considered correlated across years. The uncertainty in the tagging scale factors is parametrized as a function of the jet p_{T}. The uncertainties in t-tagging and W-tagging are 100% correlated across years, but the uncertainty in b-tagging has both correlated and uncorrelated components. There are different scale factors to account for the cases when the tagging algorithms incorrectly identify some jets, so a separate mistagging uncertainty is also assigned.

In addition to the experimental sources, we consider uncertainties affecting the SM simulations. Renormalization ($_R$) and factorization ($_F$) scales at the matrix element level are varied by a factor of 2 or 0.5 to take into account the effect of higher-order corrections in the $t\bar{t}$ and W jets simulations. The matrix element and parton shower (ME-PS) matching scale (h_{damp}) regulates the high-p_{T} radiation by damping real emissions from the POWHEG generator; this effect is only taken into account for $t\bar{t}$ and evaluated using independent simulated samples. Uncertainties related to the initial and final state radiation (ISR and FSR) modeling in the parton shower are taken into account by varying the strong coupling constant $_s$ at the scale Q^2 for the $t\bar{t}$ samples. Finally, an uncertainty in the correction to the top quark p_{T} in simulated $t\bar{t}$ samples, which depends on the generator-level top quark transverse momentum, is evaluated as a one-sided variation computed from the difference between the top quark p_{T} distribution with and without the correction. For all these uncertainties, those originating from the same source are considered as 100% correlated between channels and those arising from different sources are considered to be 100% uncorrelated.

6 Unfolded results

The top quark charge asymmetry is obtained by performing a simultaneous binned maximum likelihood fit to data in all bins and categories of the signal candidate sample. Statistical uncertainties due to the limited MC sample size are treated separately in each bin with the Barlow–Beeston-lite [42] approach. Each source of systematic uncertainty is included in the likelihood as a unique nuisance parameter. For contributions that apply to multiple analysis channels, the nuisance parameters are fully correlated, allowing better constraints to be placed on the systematic uncertainties. This unfolding approach also has the advantage that the contributions from the background processes are constrained by the fit, resulting in smaller systematic uncertainties than a direct background subtraction. For a given channel k in our analysis, the

channel likelihood function $\mathcal{L}_k$ is defined as:

$$\mathcal{L}_k \quad \prod_{j\ 1}^{N_{\text{reco}}} P \quad n_j; \sum_{i\ 1}^{N_{\text{gen}}} A_{ji} \quad _u \quad _i \quad b_j \quad N \quad _u \ , \tag{2}$$

where

$P\ n;$ represents the Poisson probability of observing n events when are expected.

the index i runs over the truth bins at generator level, and the index j runs over the bins at reconstruction level. In this analysis, we use two bins corresponding to the positive and negative difference between the absolute value of the top quark and top antiquark rapidities y . Correspondingly, N_{reco} and N_{gen} are both set equal to 2.

the response matrix A_{ji} gives the probability for an event produced in bin i to be measured in bin j.

n_j corresponds to the number of data events in bin j.

b_j represents the number of background events predicted in bin j.

$N \ _u$ represents the priors from the nuisance parameters taken as normal distributions of the individual uncertainty sources $_u$. Normalization uncertainties are taken as log-normal distributions.

Each analysis channel is defined by a range of $M_{\text{t}\bar{\text{t}}}$ values and a specific year and lepton flavor. The final result is given by the product of the individual likelihoods with the index k running over all 12 channels: two lepton flavors (jets and e jets), 3 years (2018, 2017, and 2016), and two mass regions ($M_{\text{t}\bar{\text{t}}}$ 750, 900 GeV and $M_{\text{t}\bar{\text{t}}}$ 900 GeV). Table 1 shows the signal and background yields in our final candidate sample after the likelihood fit, separated into the two mass regions. The contributions to our candidate sample from background processes (ST, W jets, DY, and QCD multijet) are taken from simulation. The Diboson background yield is negligible and therefore omitted from the table. Figure 2 shows y for each of these 12 channels both before and after the likelihood normalization. As can be observed, the likelihood fit reduces the total uncertainty significantly and improves the agreement between data and the SM prediction.

Combinations of subsets of these channels are also possible and allow us to obtain results for the two mass regions separately. In all cases, the unfolding performs a multi-dimensional maximum likelihood fit of the simulation to observed data and returns two signal strengths, r_{pos} and r_{neg}, corresponding to the y 0 and y 0 regions, respectively. We choose to define r_{pos} in terms of r_{neg} and A_C^{fid}, which allows us to obtain from the likelihood fit directly:

$$r_{\text{pos}} \quad r_{\text{neg}} \quad \frac{N_{\text{truth}} \quad y \quad 0}{N_{\text{truth}} \quad y \quad 0} \quad \frac{1 \quad A_C^{\text{fid}}}{1 \quad A_C^{\text{fid}}} \ , \tag{3}$$

where

$N_{\text{truth}} \quad y$ is the number of events in a given y bin at the generator level,

r_{neg} is the signal strength that scales the contribution of the events with y 0, and

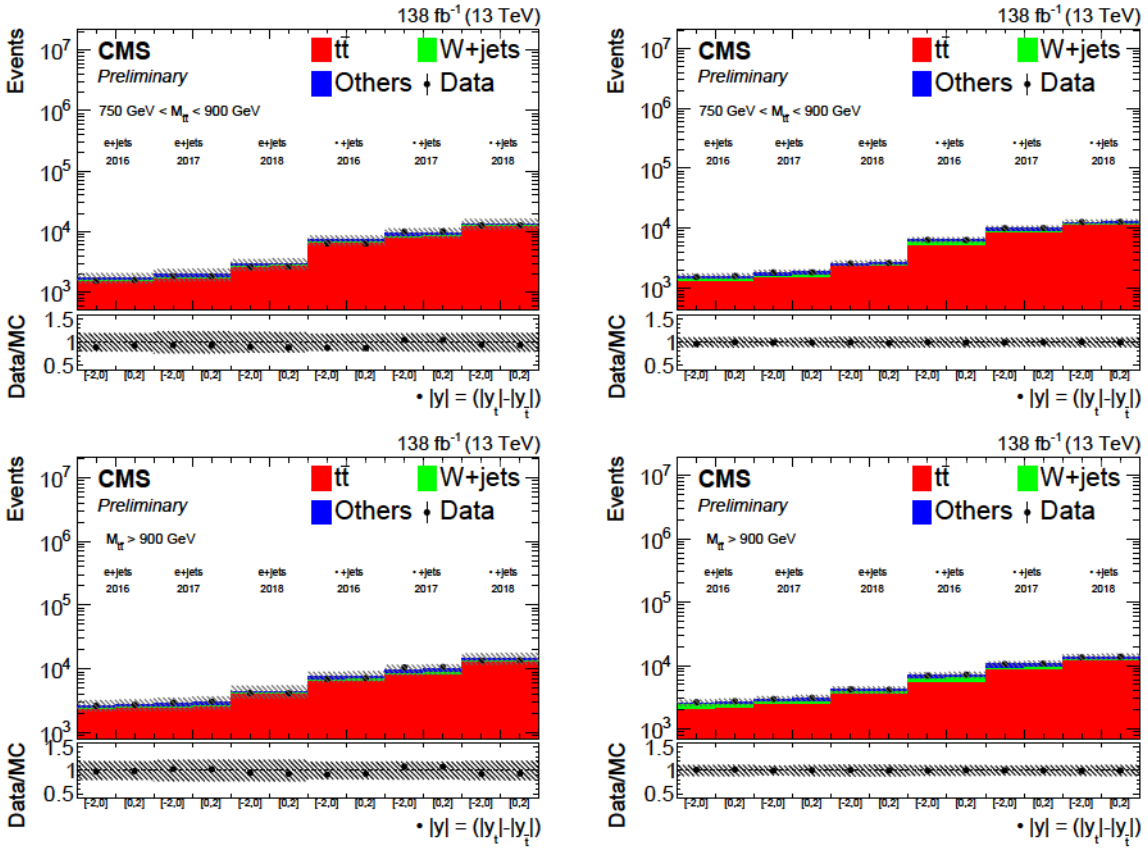

Figure 2: Comparison between data and SM prediction for $\Delta|y|$ for each of the 12 analysis channels both before (left) and after (right) the likelihood normalization. The plots in the top row correspond to 750 GeV $\leq M_{t\bar{t}} \leq$ 900 GeV, and the plots in the bottom row to $M_{t\bar{t}} >$ 900 GeV. Data points are shown with statistical uncertainty, and the shaded band combines the MC statistical uncertainty and the systematic uncertainty. As can be observed, these uncertainties are reduced significantly after the likelihood fit, and the agreement between data and simulation is improved. Overall, excellent agreement in all channels is observed.

Table 1: Event yields after the likelihood fit for each of the 12 channels used in the analysis (jets, e jets and 3 years: 2018, 2017, and 2016), separated into the two mass regions, for events that pass the signal sample selection. The errors shown include both the MC statistical and the total systematic uncertainty.

Process	2018		2017		2016		e$_{2018}$		e$_{2017}$		e$_{2016}$	
					750 GeV $M_{t\bar{t}}$ 900 GeV							
$t\bar{t}$	22226.67	1950.71	16434.01	1403.42	10368.72	965.88	4586.29	445.32	2952.65	258.24	2564.28	237.28
ST	1618.11	146.22	2154.47	190.31	912.80	77.74	413.35	34.46	506.82	43.33	286.85	25.73
W jets	966.31	107.78	1152.92	124.98	1248.75	271.95	244.14	26.93	223.16	25.24	320.43	69.12
DY	85.51	16.55	44.35	8.51	53.16	10.87	15.42	3.00	6.15	1.19	9.31	1.72
QCD multijet	415.36	101.54	270.23	55.02	182.20	43.53	5.64	2.15	1.85	1.42	0.00	0.00
Total	25311.96	1961.85	20055.99	1422.86	12765.63	1007.44	5264.85	447.48	3690.64	263.07	3180.87	248.48
Data	25417		20052		12735		5219		3674		3127	
					$M_{t\bar{t}}$ 900 GeV							
$t\bar{t}$	23343.26	2269.82	17124.51	1641.34	10697.60	1059.26	7144.49	745.33	4875.68	485.69	4111.02	425.51
ST	1646.12	143.62	2022.92	166.04	924.42	84.56	609.30	55.66	689.51	59.91	422.01	36.27
W jets	1448.65	171.68	1326.88	159.09	1968.35	449.51	518.91	63.82	435.07	52.74	741.77	161.82
DY	106.53	21.54	61.12	11.42	66.78	13.14	31.31	5.60	14.36	2.78	18.86	3.66
QCD multijet	856.61	122.48	811.55	133.86	473.15	74.80	13.17	2.58	34.19	8.59	40.16	10.30
Total	27401.17	2284.21	21346.98	1662.80	14130.29	1156.29	8317.18	750.15	6048.81	492.28	5333.82	456.81
Data	27298.00	165.22	21358.00	146.14	14157.00	118.98	8361.00	91.44	6066.00	77.88	5385.00	73.38

A_C^{fid} is the value of the unfolded charge asymmetry in the fiducial phase space and its uncertainty.

The measured charge asymmetry in the fiducial phase space is found to be consistent, within uncertainty, with the expectations. This is shown in Table 2 and Fig. 3 (left), which together summarize the A_C^{fid} values for different mass regions with their statistical and systematic uncertainties, compared to the theoretical values obtained using Asimov data.

Table 2: Measured unfolded charge asymmetry at fiducial phase level in individual channels compared with the SM predictions.

$M_{t\bar{t}}$ (GeV)	A_C (measured)	stat	syst	MC stat	total	A_C (theory)	
			A_C in fiducial phase space				
750	0.0022	0.0044	0.0034 0.0043	0.0032	0.0064 0.0069	0.0072	0.0064 0.0061
750 900	0.0039	0.0066 0.0065	0.0039 0.0056	0.0043 0.0044	0.0088 0.0096	0.0060	0.0097 0.0091
900	0.0118	0.0058	0.0055 0.0075	0.0041	0.0090 0.0103	0.0083	0.0085 0.0082
			A_C in full phase space				
750	0.0069	0.0044	0.0034 0.0042	0.0032	0.0065 0.0069	0.0094	0.0005 0.0007
750 900	0.0243	0.0065	0.0029 0.0064	0.0045 0.0043	0.0084 0.0101	0.0087	0.0006 0.0008
900	0.0037	0.0058	0.0055 0.0072	0.0041 0.0040	0.0090 0.0101	0.0101	0.0006 0.0007

The measured A_C is also presented in the full phase space using a correction based on , the product of the acceptance measured at generator level times the event selection efficiency. The number of unfolded signal events in each channel is divided by the corresponding to correct from the fiducial phase space of that channel to the full phase space common to all 12 channels. In this case, the relation:

$$r_{\text{pos}} \quad r_{\text{neg}} \quad \frac{\text{pos}}{\text{neg}} \quad \frac{N_{\text{truth}} \quad y \quad 0}{N_{\text{truth}} \quad y \quad 0} \quad \frac{1 \quad A_C^{\text{full}}}{1 \quad A_C^{\text{full}}} \tag{4}$$

allows us to obtain the signal strengths r_{neg} and A_C^{full} directly from the likelihood fit. A_C^{full} is the

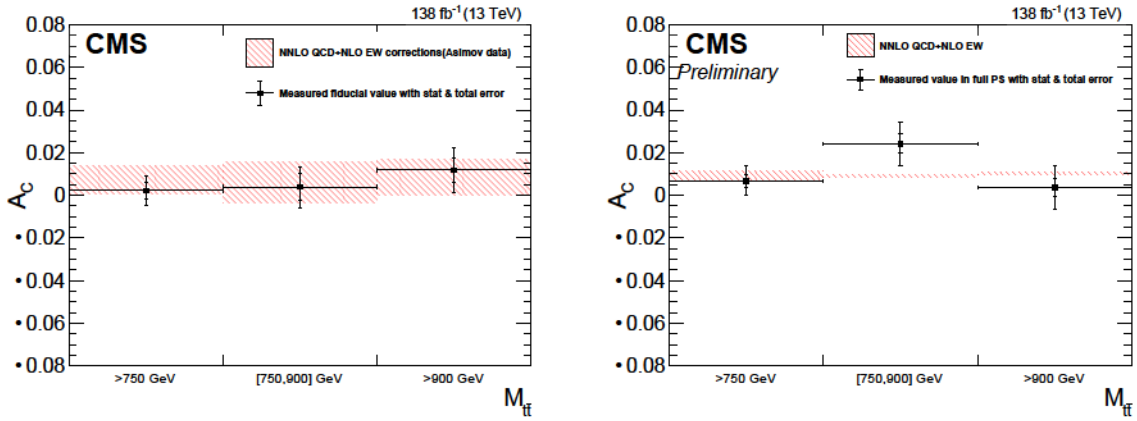

Figure 3: The measured A_C values in different mass regions, combining the μ + jets and e + jets channels, compared with the prediction in the fiducial region obtained by fitting Asimov data (left) and the theoretical prediction including NNLO QCD and NLO EW corrections from Ref. [4] (right).

value of the unfolded charge asymmetry at parton level corrected back to the full phase space and its uncertainty. The uncertainty in the acceptance arising from theoretical uncertainties in $t\bar{t}$ generation is several orders of magnitude smaller than the systematic uncertainties and therefore neglected.

Table 2 and Fig. 3 (right) show the measured top quark charge asymmetry after unfolding to parton level in the full phase space, compared with the theoretical prediction at NNLO QCD and NLO EW corrections from Ref. [4]. Good agreement between the data and the SM prediction is observed. Figure 4 shows the ranking of the main systematic uncertainties for the inclusive A_C measurement.

7 Summary

The first measurement of the charge asymmetry for highly boosted top quark-antiquark pairs in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ has been presented based on $138\,\text{fb}^{-1}$ of data. The selection was optimized for top quark-antiquark pairs produced with large Lorentz boosts and decaying to a single lepton + jets, resulting in non-isolated leptons and overlapping jets. The top quark charge asymmetry is corrected for detector and acceptance effects using a binned maximum likelihood fit. The resulting unfolded charge asymmetry for $t\bar{t}$ events with $M_{t\bar{t}} \geq 750\,\text{GeV}$ corrected to the full phase space is $A_C^{\text{full}} = 0.0069^{+0.0065}_{-0.0069}$. The corresponding theoretical prediction at NNLO in perturbation theory with NLO electroweak corrections from Ref. [4] is $0.0094^{+0.0005}_{-0.0007}$. Good agreement between the data and the SM prediction is observed.

References

[1] 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, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[2] S. Catani et al., "Top-quark pair production at the LHC: Fully differential QCD predictions at NNLO", *JHEP* **07** (2019) 100, doi:10.1007/JHEP07(2019)100, arXiv:1906.06535.

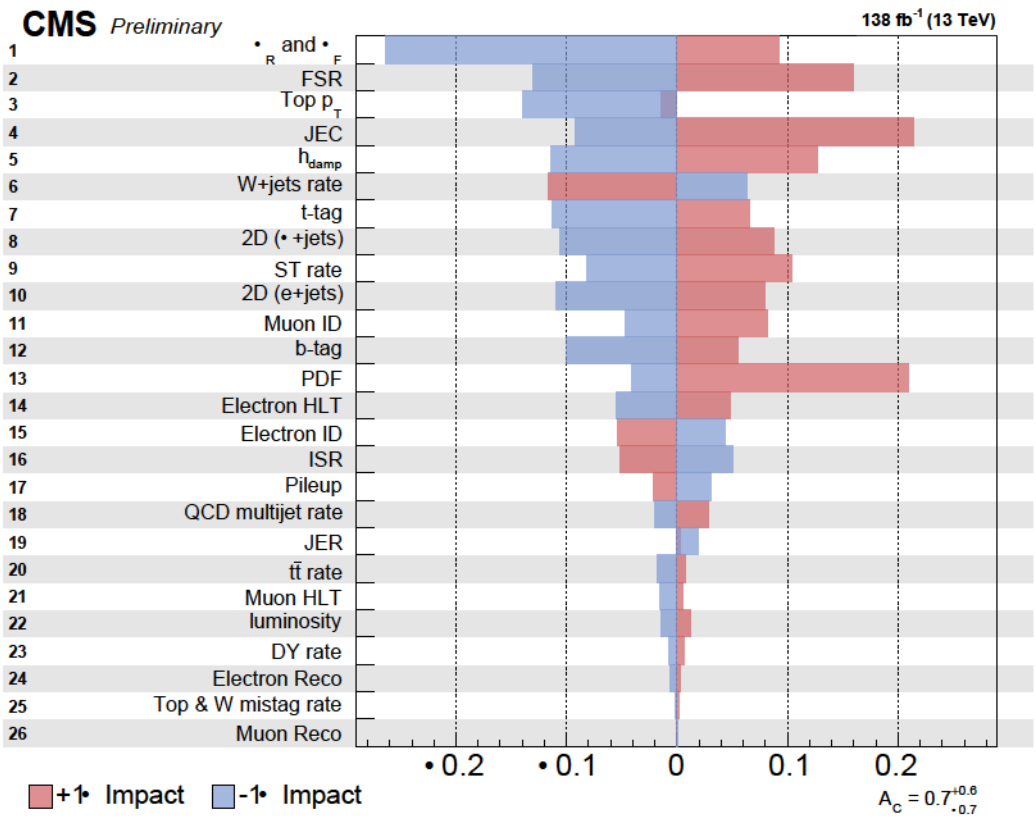

Figure 4: The impacts of the nuisance parameters corresponding to the systematic uncertainties for the inclusive A_C measurement for $M_{t\bar{t}} \geq 750\,\text{GeV}$. The blue and red bars show the effect on the unfolded A_C values for up and down variations of the systematic uncertainty. MC statistical uncertainties are omitted here.

[3] M. Czakon, P. Fiedler, and A. Mitov, "Resolving the Tevatron top quark forward-backward asymmetry puzzle: Fully differential next-to-next-to-leading-order calculation", *Phys. Rev. Lett.* **115** (2015) 052001, doi:10.1103/physrevlett.115.052001, arXiv:1411.3007.

[4] M. Czakon et al., "Top-quark charge asymmetry at the LHC and Tevatron through NNLO QCD and NLO EW", *Phys. Rev. D* **98** (2018) 014003, doi:10.1103/physrevd.98.014003, arXiv:1711.03945.

[5] M. Czakon, D. Heymes, and A. Mitov, "High-precision differential predictions for top-quark pairs at the LHC", *Phys. Rev. Lett.* **116** (2016) 082003, doi:10.1103/physrevlett.116.082003, arXiv:1511.00549.

[6] J. Rojo et al., "The PDF4LHC report on PDFs and LHC data: results from Run I and preparation for Run II", *J. Phys. G* **42** (2015) 103103, doi:10.1088/0954-3899/42/10/103103, arXiv:1507.00556.

[7] J. A. Aguilar-Saavedra, A. Juste, and F. Rubbo, "Boosting the $t\bar{t}$ charge asymmetry", *Phys. Lett. B* **707** (2012) 92, doi:10.1016/j.physletb.2011.12.007, arXiv:1109.3710.

[8] O. Antunano, J. H. Kuhn, and G. Rodrigo, "Top quarks, axigluons and charge asymmetries at hadron colliders", *Phys. Rev. D* **77** (2008) 014003, doi:10.1103/PhysRevD.77.014003, arXiv:0709.1652.

[9] P. H. Frampton, J. Shu, and K. Wang, "Axigluon as possible explanation for $p\bar{p}$ $t\bar{t}$ forward–backward asymmetry", *Phys. Lett. B* **683** (2010) 294, doi:10.1016/j.physletb.2009.12.043, arXiv:0911.2955.

[10] J. L. Rosner, "Prominent decay modes of a leptophobic Z'", *Phys. Lett. B* **387** (1996) 113, doi:10.1016/0370-2693(96)01022-2, arXiv:hep-ph/9607207.

[11] P. Ferrario and G. Rodrigo, "Massive color-octet bosons and the charge asymmetries of top quarks at hadron colliders", *Phys. Rev. D* **78** (2008) 094018, doi:10.1103/physrevd.78.094018, arXiv:0809.3354.

[12] P. Ferrario and G. Rodrigo, "Constraining heavy colored resonances from top-antitop quark events", *Phys. Rev. D* **80** (2009) 051701, doi:10.1103/physrevd.80.051701, arXiv:0906.5541.

[13] J. A. Aguilar-Saavedra and M. Pérez-Victoria, "Asymmetries in $t\bar{t}$ production: LHC versus Tevatron", *Phys. Rev. D* **84** (2011) 115013, doi:10.1103/PhysRevD.84.115013, arXiv:1105.4606.

[14] J. A. Aguilar-Saavedra and M. Perez-Victoria, "Simple models for the top asymmetry: Constraints and predictions", *JHEP* **09** (2011) 097, doi:10.1007/JHEP09(2011)097, arXiv:1107.0841.

[15] C. Zhang and S. Willenbrock, "Effective-field-theory approach to top-quark production and decay", *Phys. Rev. D* **83** (2011) 034006, doi:10.1103/PhysRevD.83.034006, arXiv:1008.3869.

[16] T. Aaltonen et al., "Evidence for a mass dependent forward-backward asymmetry in top quark pair production", *Phys. Rev. D* **83** (2011) 112003, doi:10.1103/PhysRevD.83.112003, arXiv:1101.0034.

[17] V. M. Abazov et al., "Forward-backward asymmetry in top quark-antiquark production", *Phys. Rev. D* **84** (2011) 112005, doi:10.1103/PhysRevD.84.112005, arXiv:1107.4995.

[18] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.

[19] T. Aaltonen et al., "Combined forward-backward asymmetry measurements in top-antitop quark production at the Tevatron", *Phys. Rev. Lett.* **120** (2018) 042001, doi:10.1103/PhysRevLett.120.042001, arXiv:1709.04894.

[20] J. A. Aguilar-Saavedra, D. Amidei, A. Juste, and M. Pérez-Victoria, "Asymmetries in top quark pair production at hadron colliders", *Rev. Mod. Phys.* **87** (2015) 421, doi:10.1103/RevModPhys.87.421, arXiv:1406.1798.

[21] CMS Collaboration, "Search for resonant $\mathrm{t\bar{t}}$ production in proton-proton collisions at $\bar{s}$ 13 TeV", *JHEP* **04** (2019) 031, doi:10.1007/JHEP04(2019)031, arXiv:1810.05905.

[22] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\bar{s}$ 13 TeV", *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.

[23] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[24] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[25] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[26] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[27] CMS Collaboration, "Technical proposal for the Phase-II upgrade of the Compact Muon Solenoid", CMS Technical Proposal CERN-LHCC-2015-010, CMS-TDR-15-02, 2015.

[28] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{T} jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[29] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[30] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\bar{s}$ 13 TeV using the CMS detector", *JINST* **14** (2019) P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

[31] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[32] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop", *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.

[33] J. Thaler and K. Van Tilburg, "Identifying boosted objects with N-subjettiness", *JHEP* **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.

[34] E. Bols et al., "Jet flavour classification using DeepJet", *JINST* **15** (2020), no. 12, P12012, doi:10.1088/1748-0221/15/12/P12012, arXiv:2008.10519.

[35] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s}$ 13 TeV in 2015 and 2016 at CMS", *Eur. Phys. J. C* **81** (2021) 800, doi:10.1140/epjc/s10052-021-09538-2, arXiv:2104.01927.

[36] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}$ = 13 TeV", CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[37] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s}$ = 13 TeV", CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.

[38] S. Alioli, P. Nason, C. Oleari, and Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[39] 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, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[40] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[41] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[42] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219, doi:10.1016/0010-4655(93)90005-W.