Measurement of the Charge Asymmetry in Top Quark Pair Production

The CMS Collaboration

Abstract

We present a measurement of the charge asymmetry in top quark pair production using an integrated luminosity of 1.09 fb$^{-1}$ collected with the CMS detector. Top quark pairs with a signature of one electron or muon and four or more jets, at least one of them $b$ tagged, are selected. At LHC a small charge asymmetry in the rapidity distributions of top and antitop quark is predicted, with a slightly broader rapidity distribution for top quarks, while antitop quarks are produced more centrally and therefore possess a narrower rapidity distribution. We use two different sensitive variables, one based on pseudorapidities ($\eta$) of the two top quarks and one based on rapidities ($y$). The measured charge asymmetries of $A_\eta^C = -0.016 \pm 0.030$ (stat.)$^{+0.010}_{-0.019}$ (syst.) and $A_y^C = -0.013 \pm 0.026$ (stat.)$^{+0.026}_{-0.021}$ (syst.) are within the uncertainties consistent with the small positive asymmetries predicted by the Standard Model.
1 Introduction

The top quark is the only known fermion with a mass of the order of the electroweak symmetry breaking (EWSB) scale and therefore plays a special role in many beyond Standard Model (BSM) theories. In addition to the production via quark-antiquark annihilation and gluon-gluon fusion, in some BSM theories top quarks can also be produced by the exchange of yet unknown heavy particles. Possible candidates for such heavy particles are axigluons [1, 2], Z’ bosons [3] or colored Kaluza Klein excitations of gluons [4, 5]. Such new exchange particles might show up as a resonance in the invariant t ¯t mass spectrum in case of s channel production of top quark pairs. If these hypothetical new particles are exchanged in the t or u channel, alternative approaches are needed in order to search for new top quark production modes [6]. A property of t ¯t production, which is sensitive to such additional production modes is the t ¯t charge asymmetry.

In the Standard Model (SM), the interference between the leading order (LO) Feynman diagram and box diagrams and between initial-state-radiation and final-state-radiation leads to a small charge asymmetry in the t ¯t production in the quark antiquark annihilation mode [7], linking the flight direction of the (anti)top quark to the direction of motion of the initial (anti)quark. At the Tevatron this leads due to the asymmetric initial state of proton-antiproton collisions to an observable forward-backward asymmetry, where the top quark tends to fly into the direction of motion of the incoming proton and the antitop quark flies in the direction of the initial antiquark. This asymmetry is accessible by the difference of the rapidities (y) of top and antitop quarks, y_t − y_\bar{t}. Recent measurements [8–11] by the CDF and D0 collaborations report an asymmetry which is about 2σ larger than the SM theory predicted value of about 5% [7]. In the region with high invariant masses (M_{t\bar{t}} > 450 GeV/c^2) the CDF collaboration finds an asymmetry which is 3.4σ above the SM prediction [9]. In various theory papers [12–22] it has been speculated that such a large asymmetry might be generated by potential new exchange particles with different vectorial and axial couplings to top and antitop quarks.

Due to the symmetric initial state of proton-proton collisions at the LHC, the charge asymmetry manifests itself no more in terms of a forward backward asymmetry. The rapidity distributions of top and antitop quarks are symmetrically distributed around zero. But since the quarks in the initial state are mainly valence quarks, while the antiquarks are always sea quarks, the different averaged momentum fractions of quarks and antiquarks are transferred to different widths of the rapidity distributions of top and antitop quarks. Thus, in the SM, the rapidity distribution of top quarks is broader compared to that of the more centrally produced antitop quarks. This asymmetry can be observed in the difference of the absolute values of the pseudo-rapidities of top and antitop quarks \Delta(\eta) = |\eta_t| − |\eta_\bar{t}| or using \Delta(y^2) = (y_t − y_\bar{t}) \cdot (y_t + y_\bar{t}). The latter variable can be interpreted as the variable used at the Tevatron multiplied by a factor accounting for the boost of the t ¯t system and is motivated in [23]. In both variables one can define the charge asymmetry

\[ A_C = \frac{N^+ - N^-}{N^+ + N^-}, \] (1)

where N^+ is the number of events with a positive value of the sensitive variable and N^- is the number of events with negative values, respectively. Since the SM charge asymmetry is a NLO effect in quark anti-quark annihilation and since at LHC the top quark pairs are mainly produced by gluon-gluon fusion processes the expected SM asymmetry at the LHC is even smaller than the 5% predicted for the Tevatron. For a center-of-mass energy of 7 TeV the current prediction for an asymmetry in the \(|\eta_t| − |\eta_\bar{t}|\) variable is \(A_C^{7\text{TeV}} = 0.013 \pm 0.001\) [24] and
$A_C^\eta = 0.011 \pm 0.001$ [24] using rapidities instead of pseudo-rapidities (For the sake of simplicity we focus on $\Delta(|\eta|)$ when describing the method but quote uncertainties and results for both variables). The theory predictions are an update of [7] using a top quark mass of $m_t = 173.3 \pm 1.1$ GeV/$c^2$ [25] and the MSTW2008 [26] structure functions. For the estimation of the uncertainties different choices of the parton distribution functions and different choices of the factorization and normalization scale as well as variations of the top quark mass within its experimental uncertainty have been taken into account. As at the Tevatron, the existence of new exchange particles with different vectorial and axial couplings to top and antitop quark could enhance this asymmetry.

Using the difference of the absolute values of the pseudo-rapidities of the top quark pair an analysis on the 2010 dataset corresponding to an integrated luminosity of 36 pb$^{-1}$ has been performed [27]. In this first measurement of this property of top quark production at the LHC an asymmetry of $A_C^\eta = 0.060 \pm 0.134$ (stat.)$^{+0.028}_{-0.025}$ (syst.) has been found, in agreement with the SM theory predictions. The precision on the result is dominated by the statistical uncertainty.

The analysis documented in this note is an update of the measurement reported in [27] using about 30 times more data. While the previous analysis used lepton+jets events with one isolated high-energetic charged lepton (electron or muon) and four or more high-energetic jets, without the requirement on a $b$ tag, we now require at least one jet to be tagged as stemming from a $b$ decay to enhance the $t\bar{t}$ purity of the selected data sample.

After the selection of $t\bar{t}$ candidate events, the four-momenta of the two top quarks are reconstructed in order to derive the distributions of the sensitive variables. In order to allow for a direct comparison between measured and theoretically predicted asymmetry we subtract the expected background contributions and apply a regularized unfolding technique on the reconstructed distributions before the asymmetry is finally computed.

## 2 Data and Simulation

In this analysis, proton-proton collisions at a center of mass energy of 7 TeV taken with the CMS detector up to July 2011 have been analyzed (for a description of the CMS detector see [28]). The amount of data corresponds to an integrated luminosity of $1.09 \pm 0.07$ fb$^{-1}$. In order to compare the measured distributions with predictions we make use of simulated data samples. Top quark pair events are generated with the tree-level matrix element generator MADGRAPH [29] interfaced to PYTHIA [30] for the parton showering using the MLM [31] matching algorithm. Spin correlation in the top quark decays is taken into account and higher order tree-level gluon and quark production is described via the matrix element for up to three extra jets beyond the top quark pair system. For the description of the main SM backgrounds to top quark pair production the same combination of MADGRAPH and PYTHIA is used. $W$ and $Z$ boson production is simulated in association with jets (abbreviated as $W$+jets and $Z$+jets in the following). The radiation of up to four jets is simulated with the matrix element. Also the electroweak production of single top quarks is simulated using the MADGRAPH generator. All generated events were fully simulated and reconstructed via the CMS simulation and reconstruction software.

## 3 Event Selection

We select $t\bar{t}$ candidate events, where one $W$ boson stemming from a top quark decay subsequently decays leptonically into a muon or electron, and the other $W$ boson decays hadronically into hadronic jets. For the reconstruction of electrons, muons and jets we use the particle
flow (PF) algorithm which does not identify and reconstruct all these objects separately but rather aims at reconstructing globally the whole event based on PF objects [32]. We use only events selected by special trigger algorithms which search for electrons or muons together with at least three jets with at least 30 GeV/$c$ of transverse momentum. In addition, the presence of a primary vertex (PV) reconstructed in a cylindrical region with $|z| < 24$ cm and $r < 2$ cm around the nominal interaction point for which the weighted sum of the number of tracks used for its construction is larger than 4 is required.

In the electron+jets selection the electron candidates have to have a transverse energy larger than $E_T > 30$ GeV and have to lie within the region $|\eta| < 2.5$, excluding the transition region between the barrel and the endcap of the electromagnetic calorimeter (ECAL), $1.4442 < |\eta_{sc}| < 1.5660$, where $\eta_{sc}$ is the pseudorapidity of the supercluster of the electron candidate. The latter is the cluster of energy deposits from the electron and possible Bremsstrahlung photons. The transverse impact parameter of the electron’s track with respect to the beamspot is required to be smaller than 0.02 cm. The energy of the hadronic calorimeter (HCAL) cell mapped to the supercluster must be less than 2.5% of the total calorimeter energy associated with the supercluster. Additional requirements are made on the shower shape and the angular separation between the ECAL supercluster and the matching track. The $z$ position of the electron’s vertex is required to lie within 1 cm around the $z$ position of the primary vertex to ensure that the electron comes from the primary interaction. Electron candidates have to be isolated. The relative isolation of a lepton is based on particle flow objects and defined as

$$I_{\ell}^{\text{Rel}} = \frac{I_{\ell}^{CH} + I_{\ell}^{NH} + I_{\ell}^{Ph}}{p_{T,\ell}},$$

where $I_{\ell}^{CH}$ is the energy deposited by stable charged hadrons in a cone of $\Delta R = 0.4$ around the lepton’s track and $I_{\ell}^{NH}$ and $I_{\ell}^{Ph}$ are the respective energies of the neutral hadrons and photons. We require electron candidates to have $I_{\ell}^{\text{Rel}} < 0.125$.

Events with exactly one electron candidate satisfying these quality criteria are selected for further consideration. Two handles are used to reject events in which the found electron candidate comes from the conversion of a photon into an electron-positron pair. The first approach uses the fact that photons travel through the tracker before they convert and therefore there may be layers of tracker material between the beam line and the conversion vertex which do not have hits which are valid to the reconstructed tracker hits of the electron. We therefore require the number of missed inner tracker layers of the electron track to be exactly zero (i.e. there are no missed layers before the first hit of the electron track from the beamline). In the second approach we reject any event in which the selected electron is flagged as a conversion using the partner track conversion veto, a dedicated conversion removal algorithm which identifies conversions by a geometric reconstruction of the conversion event of two tracks.

In the event selection for the muon+jets channel, muons have to have transverse momentum greater than 20 GeV/$c$ and must lie within the muon trigger acceptance ($|\eta| < 2.1$). The transverse impact parameter of the muon’s track with respect to the beamspot is required to be smaller than 0.02 cm and the $z$ position of the muon’s vertex must lie within 1 cm around the $z$ position of the primary vertex. The muon candidate is required to have at least a minimum number of hits in both the silicon tracking system and the muon chambers and must be isolated from any other activity in the event, thus $I_{\mu}^{\text{Rel}} < 0.125$. Exactly one muon passing all these criteria is required.

Dilepton $t\bar{t}$ decays and Z boson events are removed in both channels by discarding events
that contain additional more loosely defined charged leptons. Therefore we reject all events containing any additional electron candidates with $E_T > 15 \text{ GeV}$, $|\eta| < 2.5$, and $I_{\text{Rel}}^e < 0.25$ or muon candidates with $p_T > 10 \text{ GeV/c}$, $|\eta| < 2.5$, and $I_{\text{Rel}}^\mu < 0.25$.

We use jets reconstructed with the anti-$k_T$ [33–35] jet algorithm with the distance parameter $R= 0.5$ and particle flow objects as input objects [32]. Selected jets are required to have a jet-energy-scale-corrected $p_T > 30 \text{ GeV/c}$ and $|\eta| < 2.4$. At least four jets have to be present in the event. One of the jets is required to be tagged as a $b$ jet by an algorithm which uses the significance of the impact parameter of the second track of the jet as discriminator value. The tracks are ordered in descending IP significance. The mistag rate of the chosen working point of this $b$ tagging algorithm is about 1%.

4 Data-driven Background Estimation

Applying the described event selections to the recorded dataset of $1.09 \text{ fb}^{-1}$ we select 12757 events, 5665 in the electron+jets channel and 7092 in the muon+jets channel. The aim of the analysis documented in this note is to measure the charge asymmetry in $t\bar{t}$ events and a data-driven estimation of the background contributions to the selected dataset is therefore a crucial prerequisite.

We make use of the discrimination power of two kinematic variables, which have already been used in a previous $t\bar{t}$ cross section measurement [36]: the missing transverse energy and $M_3$, the invariant mass of the combination of three jets that yields the largest vectorially summed transverse momentum. We separate the electron+jets and muon+jets selected datasets each in two subsets, events with $E_T < 40 \text{ GeV}$ and events with $E_T > 40 \text{ GeV}$. We fit simultaneously the $E_T$ distribution in the low-$E_T$ dataset and the $M_3$ distribution in the high-$E_T$ dataset to obtain the numbers of events for the different processes in the whole dataset ($E_T > 0$).

In order to be less dependent on lepton selection and trigger efficiencies we perform the background estimation individually for the two lepton types. Since $W$+jets production is asymmetric at the LHC and more $W^+$ bosons are produced than $W^-$ bosons, we measure the $W^+$+jets contributions from $W^+$ and $W^-$ individually using two different fit parameters and fit templates.

The number of available events after the full event selection for the different simulated QCD samples is also very limited and it is furthermore very difficult to model this specific background properly. Therefore a QCD model from data is used in order to simulate the QCD background. For this purpose the requirements of the default event selection are changed in order to enrich the selected dataset in QCD events. In both channels this is done by inverting the cut on the relative isolation. Instead of $I_{\text{Rel}}^\ell < 0.125$ we require $0.3 < I_{\text{Rel}}^\ell < 0.5$. The purity of the datasets selected with these altered event selections in terms of QCD events has been determined using simulated samples and has been found to be 92% for the muon case and 87% in the electron+jets channel.

Figure 1 shows the different fit templates for the various processes, in which the discrimination power of the two variables becomes clearly visible. The fits are performed as binned likelihood fits using the theta framework [37].

In total eight distributions are used for the fit: $E_T(e^+ \text{+jets})$, $E_T(e^- \text{+jets})$, $M_3(e^+ \text{+jets})$, $M_3(e^- \text{+jets})$, $E_T(\mu^+ \text{+jets})$, $E_T(\mu^- \text{+jets})$, $M_3(\mu^+ \text{+jets})$, $M_3(\mu^- \text{+jets})$. Each fit template is normalized to the predicted number of events for this process obtained from MC simulation. Since we fit the electron+jets and the muon+jets data individually, consequently different fit parameters are
Figure 1: $E_T$ (left) and M3 (right) shape-comparison between the different processes in the e+jets (upper) and mu+jets (lower) selection. For W+jets we show the two templates for positive and negative charged leptons.

Table 1: Fit results for the background estimations in the electron+jets and the muon+jets channel together with their statistical uncertainties. The last column presents the sum of both channels, where the uncertainties have been added in quadrature. In the last row the number of observed events in each channel can be found.

<table>
<thead>
<tr>
<th>process</th>
<th>electron+jets</th>
<th>muon+jets</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}$</td>
<td>$4401 \pm 165$</td>
<td>$5835 \pm 199$</td>
<td>$10236 \pm 258$</td>
</tr>
<tr>
<td>single top (t + tW)</td>
<td>$213 \pm 58$</td>
<td>$293 \pm 81$</td>
<td>$507 \pm 99$</td>
</tr>
<tr>
<td>$W^+$/jets</td>
<td>$313 \pm 84$</td>
<td>$404 \pm 106$</td>
<td>$718 \pm 135$</td>
</tr>
<tr>
<td>$W^-$/jets</td>
<td>$299 \pm 90$</td>
<td>$245 \pm 109$</td>
<td>$544 \pm 141$</td>
</tr>
<tr>
<td>Z+jets</td>
<td>$81 \pm 24$</td>
<td>$85 \pm 26$</td>
<td>$165 \pm 35$</td>
</tr>
<tr>
<td>QCD</td>
<td>$355 \pm 71$</td>
<td>$232 \pm 79$</td>
<td>$587 \pm 106$</td>
</tr>
<tr>
<td>total fit result</td>
<td>$5663 \pm 226$</td>
<td>$7094 \pm 276$</td>
<td>$12757 \pm 357$</td>
</tr>
<tr>
<td>observed data</td>
<td>$5665$</td>
<td>$7092$</td>
<td>$12757$</td>
</tr>
</tbody>
</table>

Table 1 summarizes the fit results for the electron+jets and the muon+jets channel together with their statistical uncertainties. Figure 2 shows the measured $E_T$ and M3 distributions in the two channels together with the Monte Carlo simulation normalized to the fit results. It is clearly visible that data and simulation scaled to the fit results agree very well.
W boson mass equals the pole mass of 80.4 GeV/c^2. The masses of the reconstructed top quarks and the hadronically decaying W boson are taken into account as well as the tag information of the four jets assigned to the four final state quarks. Since these masses, especially the mass of the hadronically decaying W boson and the mass of the hadronically decaying top quark, are correlated, in a first step the masses are linearly decorrelated into new masses m_1, m_2, and m_3. We study simulated t\bar{t} events and build for each of the three masses the quadratic constraint using the four-jets decay channel through the four-momenta of their decay products. In each event there exist numerous hypotheses for the reconstruction of the t\bar{t} pair from the assignment of jets to the top quark decay products. The four-momentum vector of the neutrino is derived in the leptonically decaying top quark, while the hadronically decaying top quark is then given by the combination of three of the selected jets with this hypothesis. Adding the resulting four-momentum of the neutrino and the four-momentum of the charged lepton leads to the four-momentum of the leptonically decaying W boson. Combining one of the selected jets with this W boson leads to the four-vector of the leptonically decaying top quark, while the hadronically decaying top quark is then given by the comination of three of the remaining jets.

From the list of the N^{hyp} possible reconstructions, one hypothesis has to be chosen in each event which matches best the assumption of a t\bar{t} event. For this purpose we define a criterion based solely on reconstructible quantities. The masses of the reconstructed top quarks and the hadronically decaying W boson are taken into account as well as the b tag information of the four jets assigned to the four final state quarks. Since these masses, especially the mass of the hadronically decaying W boson and the mass of the hadronically decaying top quark, are correlated, in a first step the masses are linearly decorrelated into new masses m_1, m_2, and m_3. We study simulated t\bar{t} events and build for each of the three masses m_i the ratio L(m_i) between the distribution of the best possible hypotheses (defined by a matching of the remaining reconstructible quantities) and the distribution of the best possible hypothesis (defined by a matching of all reconstructible quantities).

Figure 2: Data-MC comparison for the E_T (left) and M3 (right) distributions for the e+jets (upper) and mu+jets (lower) selection. The simulation has been scaled to the fit results.

5 Reconstruction of t\bar{t} Pairs

The measurement of the t\bar{t} charge asymmetry is based on fully reconstructing the top quarks in the lepton+jets decay channel through the four-momenta of their decay products. In each event there exist numerous hypotheses for the reconstruction of the t\bar{t} pair from the assignment of jets to the top quark decay products. The four-momentum vector of the neutrino is derived from the missing transverse energy. To calculate the z-component of the neutrino momentum, a quadratic constraint using the W boson decay kinematics is used, with the assumption that the W boson mass equals the pole mass of 80.4 GeV/c^2. This leads in general to two solutions for the neutrino momentum. If the solution of the equation is complex, the x and y components are varied such that the imaginary part vanishes. Comparison studies in simulated events show that the three spatial components of the neutrino vector reconstructed with this treatment are closer to the generated values yielding a better resolution of the neutrino direction of motion. Adding the resulting four-momentum of the neutrino and the four-momentum of the charged lepton leads to the four-momentum of the leptonically decaying W boson. Combining one of the selected jets with this W boson leads to the four-vector of the leptonically decaying top quark, while the hadronically decaying top quark is then given by the comination of three of the remaining jets.
reconstructed and the generated top quarks and W bosons) and all hypotheses, which is a measure for the probability that a hypothesis with a given value of the three masses is the best possible hypothesis.

In addition, also the $b$ tagger discriminator values for the jets assigned to the two $b$ quarks and to the two light quarks are considered. The probability that a jet with a certain $b$ tag discriminator value $x$ is assigned to one of the $b$ quarks is denoted $P_b(x)$.

Now, having all ingredients at hand, we can calculate the final estimator $\psi$ which is used for the purpose of choosing one single hypothesis in each single event. $\psi$ is given by the product of the three ratios of mass distributions and the four $b$ jet probabilities, where for the light quarks for obvious reasons $(1 - P_b(x_{q_1}))$ has been used:

$$\psi = L(m_1)L(m_2)L(m_3)P_b(x_{b,lep})P_b(x_{b,had})(1 - P_b(x_{q_1}))(1 - P_b(x_{q_2})) . \quad (3)$$

Implementing the observed values of the three reconstructed masses and the $b$ tagger discriminator values of the used jets for each hypothesis a value of $\psi$ can be calculated. The hypothesis with the smallest value of $-\ln \psi$ is then chosen in each event for further consideration. Studies on simulated events show that in about 29% of all events we choose the best possible hypothesis using this criterion. Studying only events in which all jets corresponding to one of the four final-state quarks are present in the event, the best solution is found in 51% of the events.

6 Measurement of the $t\bar{t}$ Charge Asymmetry

![Figure 3: $\Delta(\eta)$ and $\Delta(y^2)$ distributions for the combined lepton+jets channel. The simulation has been normalized to the prediction.](image)

The distributions of the sensitive variables obtained from the reconstructed top and antitop quark four vectors are shown in figure 3. These distributions can be used to calculate an uncorrected raw charge asymmetry $A_{C,raw}$ by simply counting the numbers of events with negative and positive values.

Using the definition in equation 1 we find $A_{C,raw}^{\eta} = -0.004 \pm 0.009$ and $A_{C,raw}^{y^2} = -0.004 \pm 0.009$, where in both cases the combined electron+jets and muon+jets dataset has been used.

These values are not directly comparable with any theoretically motivated prediction, since several effects bias the measurement at this stage. First of all, 15% of events used to measure $A_{C,raw}$ come from background processes, although we apply a relatively tight $t\bar{t}$ event selection.
Therefore we subtract the predicted amount of background events from the measured distributions. In this step a Gaussian error propagation is performed taking the measured uncertainties on the background rates as input.

An even larger effect is due to the event selection efficiency and to the imperfect reconstruction method. Depending on the true value of $\Delta(|\eta|)$ (or $\Delta(y^2)$) the probability for an event to survive the event selection varies, which leads to distortions in the reconstructed distributions. Furthermore, in many events not exactly the best possible jet parton assignment and neutrino reconstruction is chosen. But even when the best possible reconstruction hypothesis is chosen, this does not guarantee that the top quark four-momenta are reconstructed correctly. Due to the finite energy resolution of the calorimeters and the jet reconstruction, and due to the possibility that jets from the $t\bar{t}$ decay can lie outside the detector acceptance, a perfect reconstruction is a priori not possible. In order to investigate all these effects one can analyze simulated $t\bar{t}$ events, where one can compare the four momenta of the generated top quarks to the four vectors reconstructed from measured objects like jets and leptons. From these studies one finds that the distortion of the distributions can be factorized into effects due to the event selection efficiency and due to the event reconstruction procedure. Figure 4 shows as an example the migration matrix between true and reconstructed distribution and the selection efficiency as a function of the true $\Delta(|\eta|)$ value.

To correct for these effects, we apply a regularized unfolding procedure [38]. The unfolding algorithm corrects the measured spectrum for migration and efficiency effects by applying a generalized matrix inversion method. The measured spectrum, denoted as vector $\vec{w}$, is divided into 12 bins, where the different widths of the bins have been chosen such that they contain approximately equal numbers of events. For the corrected (unfolded) spectrum $\vec{x}$ six bins have been used. Using twice the number of bins for the uncorrected spectrum than for the corrected spectrum is a standard recommendation for the applied unfolding technique [38].

The smearing matrix $A$, which accounts for migration effects and efficiency losses, is derived from simulated $t\bar{t}$ events. Mathematically this transition matrix is the product of the migration matrix (visualized in figure 4 (left)) and a diagonal matrix with the efficiencies for the different bins on the diagonal and all other elements equal to zero. It defines the translation of the true spectrum $\vec{x}$ into the measured spectrum $\vec{w}$:

$$\vec{w} = A \vec{x}.$$  (4)
We solve equation 4 in terms of the true spectrum \( \vec{x} \) writing it as a least-square (LS) problem looking for the \( \vec{x}_{LS} \) which minimizes the LS by introducing a generalized inverse matrix of the smearing matrix \( A \).

In general, this solution is unstable and shows huge fluctuations for small changes in \( \vec{w} \). This is not a numerical problem but is inherited from the properties of the matrix \( A \) [38]. Usually, the singular values of \( A \) are of different orders of magnitude. The generalized inversion of \( A \) will therefore be dominated by the smallest singular values which belong to highly fluctuating eigenmodes of \( \vec{w} \). Consequently, the solution \( \vec{x} \) will be dominated by small and insignificant fluctuations of \( \vec{w} \). To regularize the problem and to avoid unphysical fluctuations two additional terms, a regularization term and a normalization term, are introduced [39, 40]. The regularization parameter is chosen such that the solution is neither dominated by the unregularized solution from minimizing just the LS problem nor from the regularization term. For both variables we use independent unfolding procedures, based on the respective observable.

The performance of the unfolding algorithm is tested in sets of pseudo experiments. For each pseudo experiment we randomly generate a pseudo data distribution. First we determine the numbers of events of the different processes by throwing random numbers from Gaussian distributions centered around the measured event rates in table 1 with a width corresponding to the respective uncertainties. To account for statistical uncertainties the number of events is further randomly varied for each process according to a Poisson distribution around the previously thrown Gaussian number. Then these numbers of events are randomly drawn from the respective MC templates to generate pseudo data distributions.

Each pseudo data distribution is then subject to the unfolding method described above. For all pseudo experiments we subtract the same amount of background events as done in the real measurement on data.

In a first step we draw the signal part of the pseudo data from the default \( t\bar{t} \) sample and perform 50,000 pseudo experiments. In each pseudo experiment we compare the unfolded spectrum with the generated distribution and calculate bin-by-bin the relative differences between the two as well as the corresponding pull distributions. The average asymmetry from these pseudo experiments agrees very well with the true asymmetry in the used sample and also the pull distributions look as expected, indicating that the error-treatment is done correctly. In a second step we re-weight the events of the default \( t\bar{t} \) according to their \( \Delta(|\eta|) \) value for the purpose of generating a non-zero asymmetry and perform 50,000 pseudo experiments for each of the re-weighted distributions. The true asymmetry of the re-weighted distributions is varied between \(-0.21 \) and \(+0.23 \) in ten steps. We find a linear behavior of the average fit results as a function of the true values. While for the \( \Delta(|\eta|) \) variable the agreement is almost perfect, for \( \Delta(y^2) \) a small deviation in the order of 0.1% has been observed for which the measured asymmetry has to be corrected.

The measurement of the charge asymmetry \( A_C \) might be affected by several sources of systematic uncertainties. In principle, only systematic uncertainties influencing the direction of the reconstructed top quark momenta can change the value of the reconstructed charge asymmetry. The overall selection efficiency and acceptance will not change the measured asymmetry. For each source of systematic we draw pseudo data distributions from systematically shifted samples and perform the unfolding with the migration and efficiency matrix taken from the standard \textsc{madgraph} sample.

To determine the systematic effect due to the imperfect knowledge of the jet energy scale (JES), we vary simultaneously all jet four momenta by either \(+1\sigma \) or \(-1\sigma \) in their \( \eta \) and \( p_T \) depen-
dent uncertainties. Jet asymmetry measurements suggest that jet $p_T$ resolutions are about 10% worse in data compared to simulation. Therefore all jets in the simulated samples are scaled such that the jet energy resolution (JER) in the simulation equals the resolution in data. The corresponding uncertainty is estimated by varying the JER within its uncertainties of $\pm 10\%$. To account for effects due to the uncertainty on the $Q^2$ scale to use for the strong coupling constant $\alpha_s$, we use two different $t\bar{t}$ Monte Carlo samples, in which the $Q^2$ scale has either been multiplied with 4 or 0.25. Effects due to extra hard parton radiation are estimated by varying the jet matching threshold for the MLM matching scheme for the simulated $t\bar{t}$ sample by a factor 0.5 or 2 from its default. The impact of initial-state- and final-state-radiation (ISR and FSR) is estimated using two alternative signal samples, where the PYTHIA parameters for additional parton radiation have been varied to produce more or less ISR/FSR compared to the default configuration. We evaluate the systematic uncertainty on the measured asymmetry induced by the imperfect knowledge of the parton distribution function (PDF) of the colliding protons using the CTEQ6.6 [41] PDF set and the LHAPDF [42] package. For this purpose, a re-weighting procedure is applied to all generated samples, in which each CTEQ6.6 PDF parameter is independently varied by its positive and negative uncertainties, with a new weight assigned to each variation. The resulting templates are used to estimate the impact of variations in the PDFs on our measurement. The overall scale factor of the $b$-tagging efficiency does not affect the result on the measured charge asymmetry. Since only an $\eta$ dependent variation of the $b$-tagging scale factor can lead to a potential change of the result, we re-weight the simulated events according to the $\eta$ dependent uncertainties on the $b$-tagging scale factor given in reference [43]. Potential effects due to different lepton efficiencies for positively and negatively charged leptons are estimated by re-weighting simulated events depending on the charge of the selected lepton. The re-weighting is performed such, that we end up with maximally different efficiencies for negatively and positively charged leptons within the overall uncertainties. We estimate an uncertainty arising from the QCD model derived from data by either using only the template for negatively charged leptons or the template for positively charged leptons instead of the standard mixture of both for the QCD pseudo data. The used MC samples are re-weighted such, that the number of simulated pileup events matches the number of pileup events in data. We apply a systematic uncertainty arising from the uncertainties from this pileup re-weighting procedure. Therefore, all simulated events gain additional weight factors which correspond to a variation of the average number of pileup events by $\pm 0.6$.

The impact on the charge asymmetry of all systematic uncertainties is summarized in table 2. The largest systematic uncertainties arise from the variation of the $Q^2$ scale and matching threshold and from the variation of initial- and final-state-radiation in the used $t\bar{t}$ signal Monte Carlo sample.

### 7 Results

We apply the described unfolding procedure to the measured $\Delta|\eta|$ distribution as well as to the distribution of the second variable, $\Delta(y^2)$. Table 3 gives an overview of the raw asymmetries and the asymmetries after the background subtraction and the final unfolding and correction for both variables. Figure 5 shows the unfolded spectra used for computing the asymmetries together with the SM prediction at NLO. In the unfolded $\Delta|\eta|$ distribution we measure an asymmetry of

$$A_C^{\eta} = -0.016 \pm 0.030 \text{ (stat.)}^{+0.010}_{-0.019} \text{ (syst.)},$$

while in $\Delta(y^2)$ we measure an unfolded and corrected (divided by 0.94) asymmetry of
The measured asymmetries for both variables at the different stages of the analysis from the raw value to the background subtracted value and to the final unfolded result.

<table>
<thead>
<tr>
<th>Observable</th>
<th>Raw $A_C$</th>
<th>BG-subtracted $A_C$</th>
<th>Unfolded (and corrected) $A_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta</td>
<td>\eta</td>
<td>$</td>
<td>$-0.004 \pm 0.009$</td>
</tr>
<tr>
<td>$\Delta(y^2)$</td>
<td>$-0.004 \pm 0.009$</td>
<td>$-0.007 \pm 0.010$</td>
<td>$-0.013 \pm 0.026^{+0.019}_{-0.021}$</td>
</tr>
</tbody>
</table>

Both measured values are within the uncertainties in agreement with the theory predictions of $A_C^{(\text{theo.})} = 0.013 \pm 0.001$ [24] and $A_C^{(\text{theo.})} = 0.011 \pm 0.001$ [24]. One can also measure the background subtracted asymmetry as a function of the reconstructed invariant mass of the $t\bar{t}$ system to investigate whether one can see a dependence of the asymmetry on $m_{t\bar{t}}$. Figure 6 shows the results for the two variables, where no increase of the asymmetry for increasing $m_{t\bar{t}}$ can be seen. However, these studies allow only for a qualitative statement, while for a quantitative statement a proper simultaneous unfolding in the sensitive variable as well as in

$$A_C^{\gamma} = -0.013 \pm 0.026 \text{ (stat.)}^{+0.026}_{-0.021} \text{ (syst.)}. \quad (6)$$
the invariant mass of the $t\bar{t}$ system has to be performed.

![Graph](image)

Figure 6: Raw asymmetries for $\Delta(|\eta|)$ (left) and $\Delta(y^2)$ (right) for different regions in the reconstructed invariant $t\bar{t}$ mass.

8 Conclusion

An update of the measurement of the charge asymmetry in $t\bar{t}$ events using a dataset corresponding to an integrated luminosity of $1.09 \text{fb}^{-1}$ has been reported. After an event selection procedure the selected events from the lepton+jets channel have been reconstructed under the $t\bar{t}$ event assumption and from the reconstructed top and antitop quark four vectors the distributions of the sensitive variables have been constructed. These distributions have then been subject to a regularized unfolding method yielding finally the asymmetry values corrected for acceptance and reconstruction effects. As opposed to the Tevatron measurements, where the measured asymmetry has been found to be significantly larger than predicted, we measure a small negative asymmetry which is within the uncertainties still very well compatible with the SM predictions and shows no tendency to large deviations from the prediction. Also the background subtracted asymmetry as a function of the invariant mass of the $t\bar{t}$ system shows no tendency to larger values for large invariant masses. For a quantitative statement of this behavior a proper simultaneous unfolding in both the sensitive variable and the invariant mass of the $t\bar{t}$ system will be performed in the future in order to confirm or to refute the Tevatron findings.

9 Acknowledgment

We would like to thank German Rodrigo for fruitful discussions, for the close collaboration and for providing us with theoretical predictions on the top quark charge asymmetry.

References


[10] D0 Collaboration, “Measurement of the forward-backward production asymmetry of $t$ and $\bar{t}$ quarks in $p \bar{p} \to t \bar{t}$ events”, DO note 6062-CONF (2010).


