Measurement of the jet mass distribution in highly boosted top quark decays in pp collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

A measurement of the jet mass distribution in highly boosted hadronic top quark decays produced in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 13$ TeV is reported. The data were collected with the CMS detector at the LHC, and correspond to an integrated luminosity of 35.9 fb$^{-1}$. The measurement is performed in the $\ell$+jets channel where $\ell$ is an electron or muon. The products of the fully hadronic decay $t \rightarrow bW$ with $W \rightarrow q\bar{q}'$ are reconstructed with a single jet with transverse momentum $p_T > 400$ GeV. The $t\bar{t}$ cross section as a function of the jet mass is unfolded at the particle level and is used to extract a value of the top quark mass of $172.56 \pm 2.47$ GeV. This is the first measurement of the top quark mass in the highly boosted regime with a total uncertainty comparable to those from measurements performed at threshold $t\bar{t}$ production.
1 Introduction

The top quark is the heaviest known elementary particle. Its high mass leads to large contributions from quantum corrections in the electroweak sector. As a result the top quark plays an important role in the mechanism of electroweak symmetry breaking, and precision measurements of the top quark mass \( m_t \) provide a crucial input for consistency tests of the standard model (SM) [1, 2]. Direct measurements of \( m_t \) at the Large Hadron Collider (LHC) reach a precision of around 0.5 GeV [3–9]. However, an ambiguity in the interpretation of the results is caused by the necessity of modeling parton shower dynamics and non-perturbative effects in these measurements. The result may depend on the Monte Carlo (MC) event generator used, the tuning of its free parameters, and in part also on the observables used in the analyses [10]. Precisely relating the obtained value of \( m_t \) to the pole mass is therefore difficult from first principles [11].

As an alternative approach, the value of the pole mass can be obtained from measurements of the total [12, 13] and differential [14, 15] \( t\bar{t} \) production cross sections, with a precision of around 1 GeV. These measurements are dominated by \( t\bar{t} \) threshold production, where uncertainties due to parton distribution functions (PDFs) and higher order strong corrections are important [16–18]. Another possibility are measurements at high Lorentz boosts, where a single jet includes all \( t \rightarrow bW \rightarrow bqq' \) decay products. The location of the peak of the jet mass \( m_{\text{jet}} \) distribution is sensitive to \( m_t \) and can be calculated from first principles [19–25] in soft collinear effective theory [26–29].

A first measurement, reporting the \( t\bar{t} \) cross section as a function of \( m_{\text{jet}} \), has been carried out using proton-proton (pp) collision data at a center-of-mass energy \( \sqrt{s} = 8 \) TeV [30]. This note reports a new measurement of the \( m_{\text{jet}} \) distribution using pp collision data at \( \sqrt{s} = 13 \) TeV with several essential improvements. The use of the exclusive XCone [31] jet clustering algorithm results in a significant improvement of the \( m_{\text{jet}} \) resolution, reduces the effects from the underlying event (UE) and additional inelastic pp interactions within the same or adjacent bunch crossings (pileup) and improves the \( m_{\text{jet}} \) distribution’s sensitivity to \( m_t \). An unfolding of the experimental data with several additional side band regions with high granularity reduces the dependence on the simulation used to correct the data for detector and migration effects. The precise measurement of the \( m_{\text{jet}} \) distribution allows for tests of analytic calculations, the modeling of highly boosted top quark decays in MC event generators, and a determination of \( m_t \) at scales much larger than in other measurements.

2 Trigger and data

This analysis uses data recorded by the CMS detector in pp collisions at \( \sqrt{s} = 13 \) TeV in 2016. The data correspond to an integrated luminosity of 35.9 fb\(^{-1} \) [32]. Events containing the decay of a top quark to a final state including a muon are selected with a high-level single-muon trigger that requires the presence of at least one muon candidate with \( p_T > 50 \) GeV and \( |\eta| < 2.4 \). For events containing a final state with an electron, the high-level trigger requires the presence of at least one isolated electron candidate with \( p_T > 27 \) GeV, or an electron candidate without an isolation requirement with \( p_T > 115 \) GeV and \( |\eta| < 2.5 \), or at least one photon with \( p_T > 175 \) GeV and \( |\eta| < 2.5 \) [33]. The latter requirement ensures events containing electrons with high \( p_T \) are efficiently selected, as the requirements on ECAL shower shapes are less stringent for photons than for electrons.

The \textsc{powheg} [34–39] v2 generator is used for simulating \( t\bar{t} \) production at next-to-leading order (NLO). Alternatively, \( t\bar{t} \) production is also simulated with \textsc{madgraph5_aMC@NLO} v2.2.2 [40,
41] at NLO to test a potential generator dependence of the measured cross sections. Background events resulting from the production of single top quarks in the $s$, $t$ and $tW$-channels are also produced with POWHEG v2 at NLO, where spin correlations are taken into account [42]. The production of a W boson with additional jets are simulated with MADGRAPH5 aMC@NLO at NLO. Events from Drell–Yan (DY) production with additional jets are simulated with MADGRAPH5 aMC@NLO at leading order (LO) and an NLO $k$ factor is applied to the LO DY+jets production cross section. The simulation of the production of two heavy gauge bosons with additional jets is performed at LO with PYTHIA 8.212 [43]. Events in which jets are produced through the strong interaction only, referred to as quantum chromodynamic (QCD) multijet events, are also simulated with PYTHIA at LO.

In the simulated MADGRAPH5 aMC@NLO samples the matrix element calculations at NLO and LO accuracy are matched to parton showers with the FxFx [44] and MLM [45] algorithms, respectively. The parton shower, hadronization process, and multiple-parton interactions (MPI) are simulated with PYTHIA. The NNPDF3.0 [46] PDFs at LO and NLO are used for processes simulated at LO and NLO, respectively. The UE tune CUETP8M2T4 [47] is used for the simulation of $t\bar{t}$ and single top quark production, all other processes are generated using CUETP8M1 [48, 49]. The detector response is simulated with the GEANT4 package [50, 51]. Simulated events are processed through the software chain used for collision data and are reweighted to match the observed distribution of the number of pileup interactions in data.

### 3 Event reconstruction

Physics objects are reconstructed using the particle-flow (PF) algorithm [52], which combines information from various subdetectors of the CMS experiment. In the process, PF candidates are either identified as a photon, electron, muon, charged hadron or a neutral hadron. The reconstructed vertex with the largest value of summed physics-object $p_T^2$ is taken to be the primary pp interaction vertex. The physics objects are the jets, clustered using the jet finding algorithm [53, 54] with the tracks assigned to the vertex as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the $p_T$ of those jets.

In this analysis hadronic jets are reconstructed from PF candidates using the anti-$k_T$ [53] or XCone algorithm [31] as implemented in the FASTJET software package [54]. In the jet clustering procedure, charged PF candidates associated with nonprimary vertices are excluded. Jets clustered with the anti-$k_T$ algorithm and a distance parameter of 0.4 are used for the identification of jets originating from the hadronisation of b quarks. They are required to have $p_T > 30$ GeV, $|\eta| < 2.4$, and to pass the tight working point of the combined secondary vertex v2 [55] algorithm in order to be identified as a b jet. The XCone jets are obtained by a two-step jet clustering [56]. First, the exclusive XCone algorithm is run with $R_{\text{jet}} = 1.2$ and requiring exactly two jets, $N = 2$. Using the constituents of these two large jets as input, XCone is run again with $N_{\text{sub}} = 3$ and $R_{\text{sub}} = 0.4$, resulting in exactly three subjets inside both large-radius jets. Subjets are only considered if they fulfill $|\eta| < 2.4$.

Lepton candidates (electrons or muons) need to have $p_T^l > 55$ GeV, $|\eta| < 2.4$. Electrons with $p_T < 120$ GeV have to be isolated as a consequence of the high-level trigger, where the isolation [57] is defined as the $p_T$ sum of charged hadrons and neutral particles in a cone with radius $\Delta R = 0.3$ around the electron, where $\Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2}$ describes the angular distance between two objects and $\phi$ is the azimuthal angle in radians. Electrons with $p_T > 120$ GeV and muons with $p_T > 55$ GeV are required to pass a two-dimensional selection of either $\Delta R(\ell, j) > 0.4$ or $p_{T, \text{rel}}(\ell, j) > 40$ GeV, where j is the anti-$k_T$ jet with minimal angular separation $\Delta R$ from the lepton $\ell$, and $p_{T, \text{rel}}(\ell, j)$ is the component of the lepton momentum
orthogonal to the jet axis [58, 59]. Each selected event is required to contain exactly one lepton.

The four-momentum of the lepton candidate overlapping with an anti-$k_T$-jet or XConesubjet is subtracted from the four-momentum of this jet. Jet energy corrections [60] are applied to anti-$k_T$-jets and XConesubjets accounting for additional energy deposits from pileup, non-linearities in $\eta$ caused by the detector design and a $p_T$ dependent detector response. In all simulated samples the jet energy resolution is smeared to match the resolution in data. An additional correction applied to the XConesubjets is derived from t$\bar{t}$ simulation in the all-jets channel to account for the differences between the subjet momenta obtained with the XCones algorithm and the momenta of anti-$k_T$-jets. This correction is parametrized as function of subjet $p_T$ and $|\eta|$, and has an average size of 2% with an average uncertainty of 0.3%. It is tested in the $\ell$+jets channel and in single top quark production in the tW-channel. Good agreement between the generated and reconstructed subjet momenta is found.

The four-momenta of the three XConesubjets are combined to form the final XCones jet, which is referred to as ’jet’ in the following if not otherwise specified. The jet used to perform the measurement is the one with the largest distance $\Delta R$ to the selected lepton. Each of the three subjets in this jet have to fulfill $p_T > 30$ GeV, otherwise the event is discarded. The jet mass $m_{\text{jet}}$ is defined as the invariant mass of the sum of the four-momenta of all PF candidates clustered into the three subjets.

The missing transverse momentum vector $\vec{p}_T^{\text{miss}}$ 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 $p_T^{\text{miss}}$ [61]. The $\vec{p}_T^{\text{miss}}$ is modified to account for corrections to the energy scale of the reconstructed anti-$k_T$-jets in the event.

4 Differential cross section measurement

The fiducial region chosen for this measurement is studied through simulations at the particle level, defined by all particles with lifetimes longer than $10^{-8}$ s. The fiducial region is defined as t$\bar{t}$ events including one lepton from the decay of a W boson with $p_T > 60$ GeV, where lepton
The measurement at the particle level uses a regularized unfolding procedure based on a least-squares fit, implemented in the TUnfold [62] framework. The optimal regularization strength is determined through a minimization of the average global correlation coefficient of the output bins [63]. The response matrix is evaluated by using $t\bar{t}$ events simulated with POWHEG, which
pass either the particle or reconstruction level selections. Prior to the unfolding, contributions from background processes are subtracted from data. Sideband regions are constructed to constrain migrations in and out of the measurement phase space. For every selection step resulting in relevant migrations a sideband region is defined: $55 < p_T^\ell < 60$ GeV, $350 < p_T^{\text{jet}} < 400$ GeV, no requirement on the $p_T$ of the subjets, $m_{\text{jet}}$ being smaller than the mass of the second jet and the lepton, and a looser b tagging criterion on anti-$k_T$ jets. Additionally, the measurement and sideband regions are divided into three and two bins in $p_T^{\text{jet}}$, respectively. Except for the b tag sideband, all sideband selections have corresponding selections on particle level in the evaluation of the migration matrix. In the migration matrix the bin size in $m_{\text{jet}}$ at the particle level is smaller than the bin size of the final measurement, which helps to reduce the dependence on variations in the signal modeling through a more precise determination of migration effects. The muon and electron channels are combined before the unfolding in order to increase the statistical precision, but are also unfolded separately to verify their consistency.

Experimental uncertainties are estimated in simulation and propagated through the unfolding process. We consider uncertainties in trigger, lepton identification and b tagging [55] efficiencies. Furthermore, uncertainties related to the jet energy scale [60], jet energy resolution, additional XCon jet corrections and pileup reweighting [64] are taken into account. Other uncertainties considered are related to the integrated luminosity measurement [32] and the production cross sections of all relevant background processes [65–70]. In terms of modeling uncertainties, variations of the parton shower (PS) modeling and its matching to the matrix element (ME) calculation, the UE, renormalization and factorization scales $\mu_R$ and $\mu_F$, the PDF set and the choice of $m_t$ used in the simulation are considered. Uncertainties in the PS modeling include variations of the initial- and final-state radiation (ISR and FSR) scales by factors of 2 and $\sqrt{2}$ [47], respectively. The matching of the ME calculation to the PS is controlled by the model parameter $h_{\text{damp}} = 1.58^{+0.66}_{-0.59}$ [47], which is varied within its uncertainties. The uncertainty related to modeling the UE is estimated by varying the parameters used to derive the CUETP8M2T4 tune in signal events. The model of color reconnection in PYTHIA based on MPIs with early resonance decays switched off is changed to three other models: the MPI-based scheme with early resonance decays switched on, a gluon-move scheme [71], and a QCD-inspired scheme [72]. Uncertainties due to the modeling of b quark fragmentation and the semileptonic branching fractions of b hadrons are found to be negligible.

The corresponding uncertainties from the signal modeling are estimated by unfolding alternative simulations and comparing the results to the true particle level distributions. The measured differential cross section in data is shown in Fig. 3 (left) and compared to the predictions from POWHEG and MC@NLO. In the peak region the total relative uncertainty is between 16 and 36%. It is dominated by uncertainties related to the jet energy scale which amount to 12–31%. The largest model uncertainties are due to variations in the modeling of final state radiation which result in uncertainties of 4–18%. The statistical uncertainties amount to 6–7%. The total measured $t\bar{t}$ cross section in the fiducial region with $112 < m_{\text{jet}} < 232$ GeV is

$$\sigma = 526.8 \pm 15.2 \text{(stat)} \pm 38.7 \text{(exp)} \pm 28.7 \text{(model,unfold)} \text{ fb}$$

$$= 526.8 \pm 50.5 \text{(tot)} \text{ fb}.$$ 

The cross section predicted by POWHEG is $679.5 \pm 109.3$ fb, where the theoretical uncertainties are obtained by simultaneous up or down variations of both scales $\mu_R$ and $\mu_F$ by factors of two, changes of the ISR and FSR PS scales by 2 and $\sqrt{2}$, respectively, variations in the parameter $h_{\text{damp}}$, and changes in the UE modeling. A smaller cross section is observed in data compared to simulation, in agreement with previous high-$p_T$ top quark measurements [30, 73–76].
Figure 3: The particle level $t\bar{t}$ differential cross section in the fiducial region as a function of the jet mass (left). The measurement is compared to predictions from POWHEG and aMC@NLO. Theoretical uncertainties are shown as colored bands for the predictions from POWHEG. The normalized differential cross section (right) is compared to predictions from POWHEG with different values of $m_t$. The vertical bars represent the statistical (inner) and the total (outer) uncertainties. The horizontal bars show the bin width.

Figure 3 (right) shows the normalized differential cross section as a function of $m_{jet}$, which is obtained by dividing the differential cross section by the total cross section in the fiducial region. The normalized differential cross section benefits from a partial cancellation of systematic uncertainties and shows good agreement with the prediction from POWHEG with a value of $m_t = 172.5 \text{ GeV}$.

5 Determination of the top quark mass

As illustrated in Fig. 3 (right), the measured $m_{jet}$ distribution shows high sensitivity to $m_t$ and the normalized differential cross section measurement can be used to extract a value of $m_t$. A fit is performed based on the $\chi^2$ evaluated as $\chi^2 = d^T V^{-1} d$, where $d$ is the vector of differences between the measured normalized cross sections and the predictions obtained from POWHEG with different values of $m_t$, and $V$ is the covariance matrix, which includes the statistical, experimental systematic, signal modeling in the unfolding, and theoretical uncertainties. The latter are calculated by changing the scales $\mu_R$ and $\mu_F$, the ISR and FSR PS scales, the parameter $h_{damp}$, and the UE modeling in the simulation. The result is

$$m_t = 172.56 \pm 0.41 \text{ (stat)} \pm 1.58 \text{ (exp)} \pm 1.55 \text{ (model,unfold)} \pm 1.02 \text{ (theo)} \text{ GeV}$$

$$= 172.56 \pm 2.47 \text{ GeV}.$$  

(1)

The fit converges at a minimum of $\chi^2 = 0.59$ for three degrees of freedom. Equation (1) represents the first determination of $m_t$ from highly boosted top quark decays with an uncertainty similar to the ones obtained from $t\bar{t}$ measurements at threshold production. The large improvement compared to the measurement at 8 TeV [30] can be attributed to the larger data set, the increased $\sqrt{s}$, and the novel jet reconstruction with XCon, which result in a reduction of the statistical uncertainty by a factor of about 14. The systematic uncertainties are also reduced.
through the XCon jet reconstruction, which allows for a more precise calibration of the subjet energies, results in a better resilience against effects from pileup, and an improvement of a factor two in the $m_{\text{jet}}$ width and resolution. Modeling uncertainties are reduced by enlarging the migration matrix with additional side band regions with finer binning. This measurement shows for the first time the importance of highly boosted $t\bar{t}$ production for the precision determination of a SM parameter.

6 Summary

In conclusion, we have presented a measurement of the differential cross section in highly boosted top quark decays as a function of the jet mass $m_{\text{jet}}$. The measurement relies on a novel method to reconstruct highly-boosted top quark decays with the XCon jet algorithm, which results in a large improvement of the $m_{\text{jet}}$ resolution and reduced systematic uncertainties. The shape of the unfolded distribution is well described by the simulation of $t\bar{t}$ production and shows high sensitivity to the top quark mass $m_t$. A determination of $m_t$ from the normalized $m_{\text{jet}}$ distribution results in a value of $m_t = 172.56 \pm 2.47$ GeV, which has an uncertainty similar to the ones from measurements at threshold production. This measurement can be compared directly to precise analytical calculations, feasible only in the highly-boosted regime. This result comprises an important step in understanding the ambiguities arising between the top quark pole mass and $m_t$ measurements at hadron colliders.

References


