CMS Physics Analysis Summary

Contact: cms-pag-conveners-top@cern.ch 2015/09/16

Measurement of the top-quark mass from the b jet energy spectrum

The CMS Collaboration

Abstract

The top-quark mass is measured using the peak position of the energy distribution of b jets produced from top-quark decays. The analysis is based on a recent theoretical proposal. The measurement is carried out selecting $\mathrm{t\bar{t}}$ events with one electron and one muon in the final state in proton-proton collision data at $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 19.7 fb^{-1}. The fitted peak position of the observed energy distribution is calibrated using simulated events and translated to a top-quark mass measurement using relativistic kinematics, with the result $m_{\mathrm{t}} = 172.29 \pm 1.17\,(\mathrm{stat.}) \pm 2.66\,(\mathrm{syst.})$ GeV.

1 Introduction

The mass of the top quark (m_t) is a parameter that plays an important role in the standard model (SM) of particle physics. It is well known that m_t is connected to the masses of the W, Z, and Higgs bosons through radiative corrections. As a consequence, an improved understanding of m_t tests the internal consistency of the SM and places constraints on beyond standard model (BSM) physics scenarios.

Many measurements of m_t have been performed using $t\bar{t}$ events produced at hadron colliders. The 2014 world average mass of the top quark was determined to be $m_t = 173.34 \pm 0.27\,(\text{stat.}) \pm 0.71\,(\text{syst.})$ GeV from a combination of measurements performed at the Fermilab Tevatron and the CERN LHC [1]. At the Compact Muon Solenoid (CMS) detector, the top-quark mass has been measured with great accuracy in the fully hadronic, lepton+jets, and dilepton channels. The combined CMS result from Run 1 of the LHC is $m_t = 172.44 \pm 0.13\,(\text{stat.}) \pm 0.47\,(\text{syst.})$ GeV [2].

In the SM, top quarks decay almost exclusively to a W boson and a b quark through electroweak interactions. We study $t\bar{t}$ events where both W bosons decay leptonically, one to an electron and a neutrino and the other to a muon and a neutrino. Hence, the final state under study consists of two b jets, an electron, a muon, and missing tranverse energy due to the two neutrinos.

This note presents the first measurement of m_t using the position of the peak of the energy spectrum of the b jets in the laboratory frame. Under the hypothesis that top quarks produced via QCD interactions are not polarized, this observable is independent of the boosts of the top quarks. Therefore, it can be related to the energy of the b quark in the rest frame of the top quark, which in turn depends on the top-quark mass [3]. This measurement is complementary to methods that involve a direct kinematic reconstruction of the invariant mass of the top quark or its daughters.

2 Data Samples, Simulation, and Event Selection

The analysis is performed on the full 2012 CMS dataset of proton-proton collisions at $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 19.7 fb^{-1}. A more detailed description of the CMS detector, together with a definition of the coordinate system and the relevant kinematic variables, can be found in [4].

Simulated $t\bar{t}$ signal events are generated using the leading order (LO) MADGRAPH matrix element generator (v5.1.3.30), with MADSPIN for the decay of heavy resonances, and are passed to the PYTHIA (v6.426) parton showering framework using its Z2* tune [5–8]. The τ leptons are decayed with the TAUOLA package (v27.121.5) [9]. The LO CTEQ6L1 PDF set is used in the generation [10]. The $t\bar{t}$ signal events are generated for seven different top-quark mass values ranging from 166.5 GeV to 178.5 GeV.

The largest background component is expected to come from single top-quark production in association with a W boson (tW). All single top-quark events are modeled using POWHEG (v1.0, r1380) [11] with the CTEQ6M PDF set and with three top-quark mass values ranging from 166.5 GeV to 178.5 GeV. The remaining backgrounds from W+jets, Drell–Yan, diboson, and $t\bar{t}$ +boson production are generated with MADGRAPH. The parton showering and fragmentation for all background events is done using PYTHIA 6.

Next-to-leading-order (NLO) or next-to-next-to-leading-order (NNLO) cross sections are used to normalize all simulated event samples [12–15]. Additional proton-proton collisions in the

same and neighboring beam crossings (pileup) are simulated with PYTHIA and superimposed on the hard collisions, using a pileup multiplicity distribution that reflects the luminosity profile of the analysed data. The CMS detector response is simulated using a detector model implemented in GEANT4 [16].

Dilepton triggers are used to select data events with a muon and an electron with a minimum p_T requirement of 17 GeV for the leading and 8 GeV for the sub-leading lepton.

Events are reconstructed using the CMS particle-flow (PF) algorithm [17], which provides a global event description that optimally combines the information from all sub-detectors to reconstruct and identify all individual particles. Reconstructed electron and muon candidates are required to have a minimum transverse momentum of 20 GeV and $|\eta| \leq 2.4$. The event selection requires exactly one muon candidate and one electron candidate with opposite charge and invariant mass larger than 12 GeV. Lepton relative isolation (I_{rel}) is defined as the ratio between the scalar sum of the transverse momenta of all reconstructed particle candidates neighboring the electron (muon) track within a cone of size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3\ (0.4)$ and its tranverse momenta. I_{rel} is corrected for contributions from pileup interactions using event-by-event information. Electrons are required to have $I_{rel} < 0.15$ and muons to have $I_{rel} < 0.12$.

Hadronic jets are constructed by clustering PF candidates using the anti-k_T algorithm [18] with a distance parameter of $R = 0.5$. Selected jets are required to have $p_T > 30$ GeV and $|\eta| \leq 2.5$. The Combined Secondary Vertex (CSV) b-tagging algorithm is used in order to identify jets that originate from the hadronization of a b-quark [19], using the loose operating point described in [19]. The expected efficiency for the identification (misidentification) of b (udsg) jets is 84% (13%). Selected events are required to have one or two b-tagged jets. Events with three or more b-tagged jets are excluded. The CSV loose operating point is found to be optimal for this analysis. A tighter requirement would slightly increase the fraction of correctly identified b jets in the sample, but significantly decrease the number of selected events.

In Table 1 the observed yields in the data and those predicted from simulation after the full event selection in the $e\mu$ dilepton channel are shown. The observed number of events is in agreement with the predictions when the total uncertainty (statistical and systematic) is taken into account.

Table 1: Observed and predicted event yields for events with 1 or 2 b-tagged jets. Statistical uncertainties are quoted.

Sample	1 b-tag	2 b-tags
$t\bar{t}$ (dileptonic)	13500±30	18710±40
W+jets	51.4±6.5	3.4±1.2
Diboson	308.7±4.6	53.6±1.9
Single top	952.3±3.5	636.9±2.9
Drell-Yan	458.5±10.9	78.4±3.7
$t\bar{t}$V (V=W,Z)	43.2±1.1	51.6±1.2
Total simulation	15320±40	19540±40
Data	14336	18518

3 Top Quark Mass Measurement from b jet Energy Peak

The top-quark is the only fermion heavy enough to produce an on-shell W boson in its decay, hence allowing a 2-body electroweak decay. Based on relativistic kinematics, in the top quark rest frame we can derive a simple relation between the masses of the top quark and its two daughters, the W boson and the b quark:

$$m_{\mathrm{t}} = E_b^{\mathrm{Rest}} + \sqrt{m_W^2 - m_b^2 + E_b^{\mathrm{Rest}^2}}. \tag{1}$$

A direct measurement of the b-quark energy in the top-quark rest frame (E_b^{Rest}) would require a full reconstruction of the top quark. However, assuming the top to be unpolarized, the peak position of the b-quark energy distribution in the laboratory frame can be shown to be independent of the boosts of the top quarks and hence related to the peak position in the top rest frame. Furthermore, it has been demonstrated that the distribution of the logarithm of the energy is symmetric about its peak [3]. These features are verified in the simulation.

By modeling the top-quark decay as described above, we are able to extract the top-quark mass from data by a measurement of the peak position of the energy (E) distribution of b-tagged jets in a very pure sample of $\mathrm{t\bar{t}}$ events. In this analysis, the energy spectrum is analyzed on a logarithmic scale, in view of the fact that it is more symmetric and its peak position is easier to determine. The distribution $1/E\ \mathrm{dN_{bjets}}/\mathrm{dlog}(E)$ is shown in Figure 1. In the following we assume that E is measured in units of GeV.

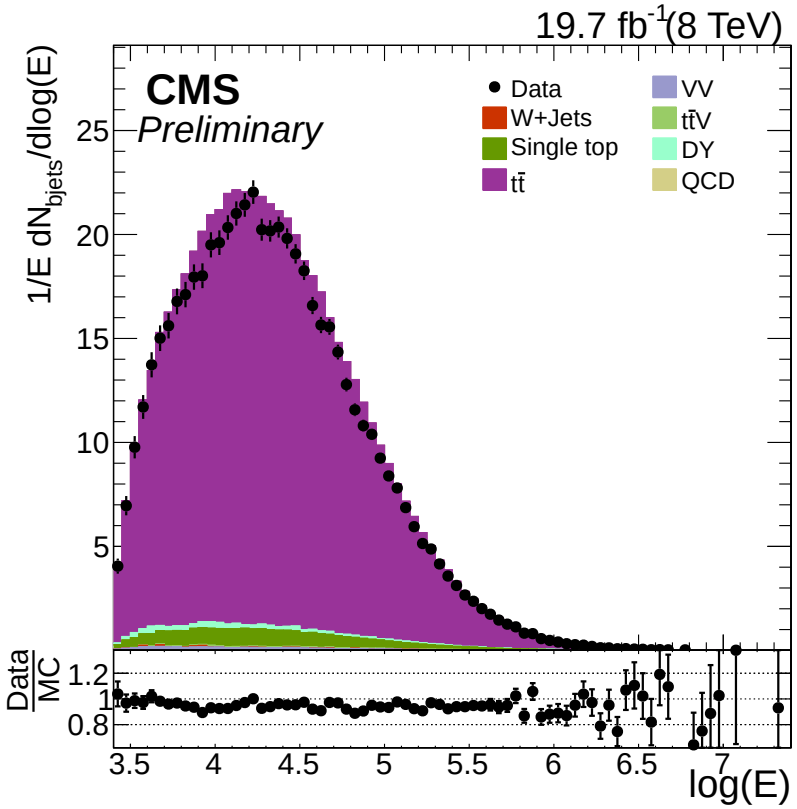


Figure 1: Logarithmic energy distribution of b-tagged jets in the 2012 dataset in events with one or two b-tagged jets. Simulated events are normalized to the prediction.

The shape of the logarithmic energy distribution around the peak position is described well by a Gaussian, as shown for simulated $\mathrm{t\bar{t}}$ events in Figure 2. We extract the peak position

using a Gaussian fit close to the peak, within a range given by twice the standard deviation of the Gaussian. With this procedure, there is an improved stability of the fit to the peak using simulation-based pseudo-experiments.

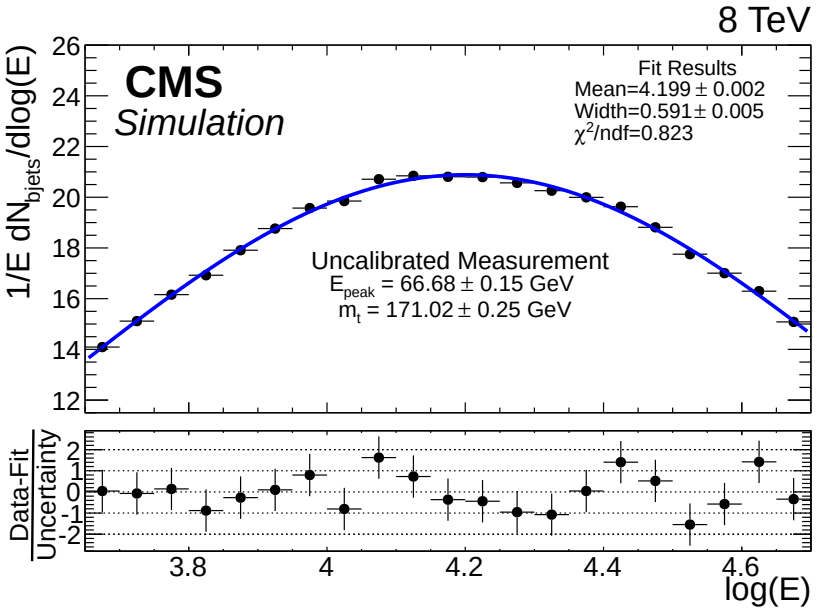


Figure 2: Fitted $\log(E)$ distribution in the whole simulated sample of $\mathrm{t\bar{t}}$ events with a mass hypothesis of 172.5 GeV. The Gaussian fit yields a $\log(E)$ peak position of 4.199 ± 0.002, corresponding to an uncalibrated value of $m_{\mathrm{t}} = 171.01 \pm 0.25$ GeV using Eq. 1.

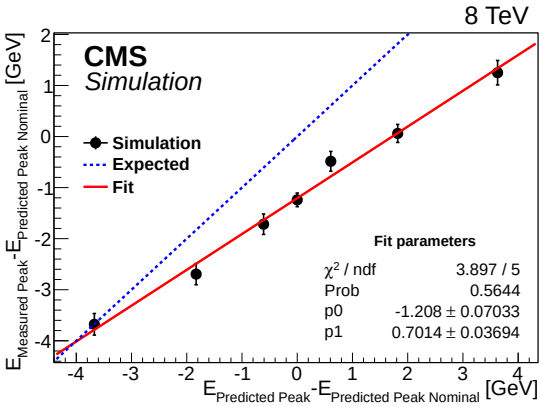
To test the method and to calibrate the measurement, pseudo-experiments are generated for each of the different top-quark mass hypotheses ranging from 166.5 GeV to 178.5 GeV. The signal and background events are combined to produce a $\log(E)$ spectrum template for generating pseudo-experiments.

The background component is divided in two parts: single-top-quark backgrounds (tW, t and s channels) and other backgrounds (DY, W+jets, VV and $\mathrm{t\bar{t}V}$). In the case of single-top-quark backgrounds, templates are created using a linear interpolation of histograms [20] and samples with different top-quark mass hypotheses. The event yields of the other backgrounds are assumed to be independent of the top-quark mass and are kept fixed (the $\mathrm{t\bar{t}V}$ yield is small and its mass dependence can be neglected). The $\mathrm{t\bar{t}}$ component of the template is then scaled such that the total template yield matches the yield observed in data.

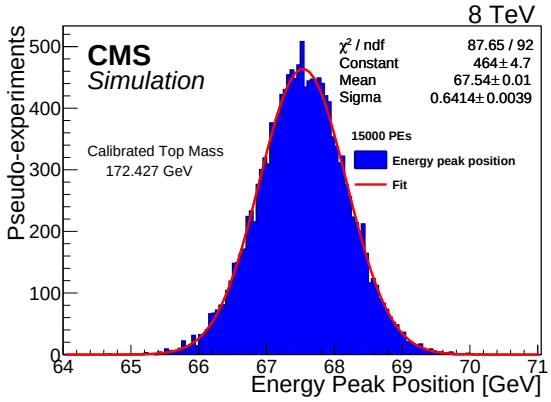
The pseudo-experiments are constructed using a Poisson distribution of the number of events in each bin of the $\log(E)$ spectrum template. The results of the pseudo-experiments are used to derive a calibration function to correct the measured energy peak position to the expected b-quark energy peak position given by Eq. 1, see Figure 3. a).

The calibration accounts for three main sources of bias in the measured energy peak: event selection cuts, reconstruction effects, and purity (which includes effects from both misidentified b jets and from background events). The event selection cuts (particularly jet p_{T} and b-tagging) generate a bias towards higher values of the peak position, whereas the reconstruction tends to shift the peak towards lower values due to the unmeasured energy of neutrinos from semileptonic b-quark decays [21]. Jets not originating from a top-quark decay and background events tend to lie in the lower region of the energy spectrum, causing a negative shift in the position of the peak.

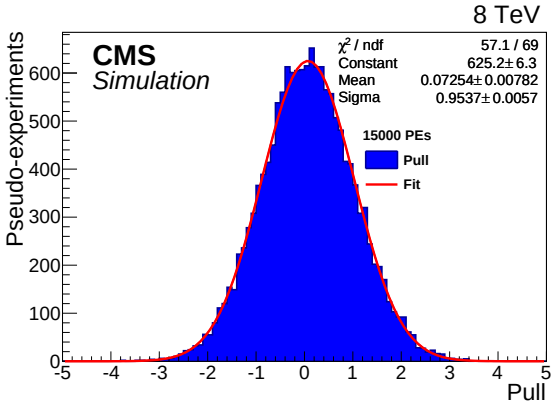
The distribution of the measured energy in pseudo-experiments for the nominal mass point of 172.5 GeV is presented in Figure 3. b). The shape of the distribution is well described by a Gaussian function. As shown in Figure 3. c), the pull distribution indicates that the fitting method overestimates the statistical uncertainties by approximately 5%. In spite of that, no correction is applied given the small effect in the total uncertainty.



(a) Calibration



(b) Energy peak position



(c) Pull

Figure 3: Measurement and calibration test on pseudo-experiments. (a) Calibration curve, (b) Distribution of calibrated energy peak position, and (c) Pull distribution. The blue line in (a) shows the calibration curve that we would expect to fit without bias effects in the measured energy peak position.

4 Systematic Uncertainties

The sources of systematic uncertainties considered for this measurement are listed in Table 2. The calculations are performed using pseudo-experiments. In Figure 4, the $\log(E)$ distributions of the dominant systematic variations are illustrated and compared to data, in order to see the systematic uncertainty coverage of the shape discrepancies between data and simulation.

- Jet energy scale (JES): The nominal jet energy scale factor used to correct for the difference between the jet energy scale in the simulation with respect to the data is varied according to its p_{T}- and η-dependent uncertainties [22]. The b-quark flavor component (bJES) is determined separately using a b+Z balancing method [21].
- Jet energy resolution (JER): The jet energy resolution in the simulation is corrected to match that of the data. The uncertainty in the top quark mass is calculated by

a variation of the correction to the resolution within one standard deviation of its uncertainty [22].

- Pile-up: This uncertainty is evaluated by generating two alternative b-jet energy distributions after changing the average number of pileup interactions in the simulation by $\pm 5\%$.
- Lepton selection efficiency: The uncertainties on the lepton selection (trigger, isolation and identification) are expected to have a small impact on the measurement as they impact the selection of the events. They are estimated by variation of the respective scale factor within their uncertainties.
- b-tagging efficiency: Two components are considered in this calculation and added in quadrature. The first one is the variation of the b-tagging efficiency within its uncertainty and the second one is a variation of the mis-tag rates within their uncertainties [23].
- Fit calibration: The calibration fit has an intrinsic systematic uncertainty due to the statistical uncertainty of its parameters.
- Backgrounds: The background components are scaled up and down within their uncertainties, defined as a certain percentage of the background cross section. These percentages are divided in three groups: 25% for single top, 100% for W+jets, and 50% for the remaining backgrounds (DY, VV, $t\bar{t}V$). The total uncertainty is the sum in quadrature of the three components.
- Modeling of the hard scattering process:
  - Parton distribution functions (PDFs): This uncertainty is taken from the envelope systematic of different variations for the PDF CT10 PDF [24].
  - Renormalization and factorization scale (Q^2): Separate samples are generated with Q varied up and down by a factor of 2.
  - Matrix element/parton shower matching scale: Separate samples are generated with the matching scale varied up and down by a factor of 2.
  - ME Generator: The uncertainty considers the difference between the results obtained by using the POWHEG generator and the MADGRAPH generator [11]. The encountered discrepancy in both generators comes mainly from the differences in allowed top quark and W boson virtualities at the generator-level.
  - Top quark p_T modeling: Recent CMS measurements suggest a mismodeling of the top transverse momentum distribution in simulated $t\bar{t}$ events [25]. The top quark p_T distribution in simulation is varied to match the observed distribution. Although the energy peak position is independent of the top quark p_T distribution in the absence of cuts, the widening of the energy distribution with harder top quark p_T gives a systematic bias once cuts are applied.
- Modeling of non-perturbative QCD:
  - Underlying event (UE): PYTHIA tunes with varied underlying event activity are compared. Perugia 2011 (P11) is compared to P11 mpiHi and P11 Tevatron tunes [26].
  - Color reconnection (CR): P11 is compared to the P11 noCR set of parameters [26].

Table 2: Sources of systematic uncertainties and their contributions to the total uncertainty. The bJES estimated uncertainty covers uncertainties related to b-quark fragmentation.

Source of uncertainty	δE_{peak} (GeV)	δm_t (GeV)
Experimental uncertainties		
Jet energy scale	0.74	1.23
b jet energy scale	0.13	0.22
Jet energy resolution	0.18	0.30
Pile-up	0.02	0.03
b-tagging efficiency	0.12	0.20
Lepton efficiency	0.02	0.03
Fit calibration	0.14	0.24
Backgrounds	0.21	0.34
Modeling of hard scattering process		
Generator modeling	0.91	1.50
Renormalization and factorization scales	0.13	0.22
ME-PS matching threshold	0.24	0.39
Top p_T reweighting	0.91	1.50
PDFs	0.13	0.22
Modeling of non-perturbative QCD		
Underlying event	0.22	0.35
Color reconnection	0.38	0.62
Total	1.62	2.66

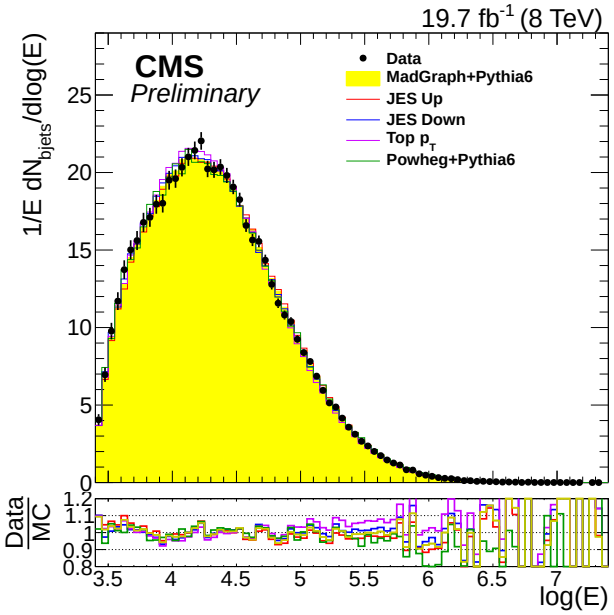
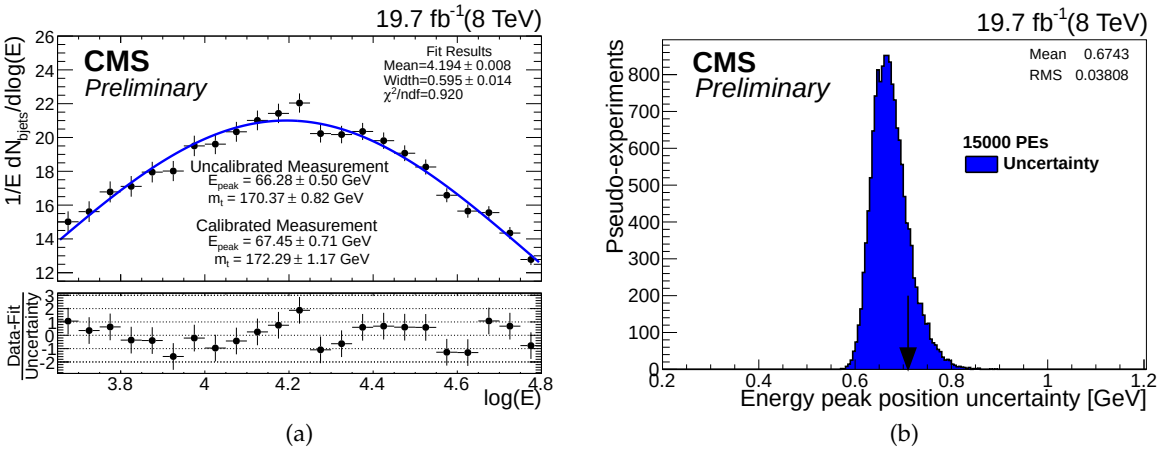


Figure 4: Comparison of nominal simulation and dominant systematics variation shapes with data. The templates are normalized to the event yields in the data. The background contributions are included.

5 Results

The top-quark mass is extracted by a measurement of the peak position of the energy spectrum of b-tagged jets as described in Section 3. The raw energy peak position in data is measured to be $E_{peak} = 66.28 \pm 0.50$ GeV. The calibration yields a value of $E_{peak} = 67.45 \pm 0.71$ GeV, which translates to a measured top-quark mass of $m_t = 172.29 \pm 1.17$ GeV (See Fig 5).



(a) (b)

Figure 5: (a) Fitted $\log(E)$ distribution in data. The calibrated mass measurement yields a value of $m_t = 172.29 \pm 1.17\,(\text{stat.}) \pm 2.66\,(\text{syst.})$. (b) Calibrated statistical uncertainty on the energy peak position in pseudo-experiments (the arrow indicates the value measured in data).

The dominant sources of systematic uncertainty for this novel technique originate from the jet energy scale and modeling of the hard scattering process. In Figure 6, the measurement is compared with the most precise top-quark mass measurements from the CMS experiment and other combinations.

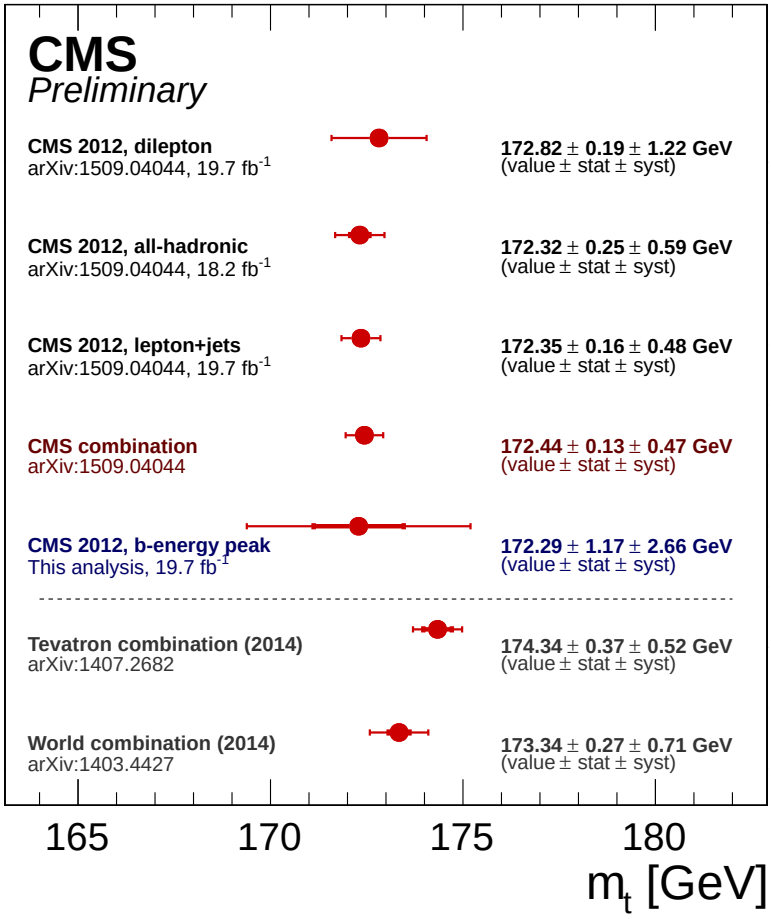


Figure 6: Comparison between the m_t measurement presented in this note and measurements at the CERN LHC and Fermilab Tevatron. In this comparison are considered the 2014 world combination [1], the LHC Run 1 m_t measurements and their combinations [2], and the 2014 Tevatron combination [27].

6 Summary

The first top-quark mass measurement using only two-body decay kinematics has been presented. The top-quark mass yields a value of $m_t = 172.29 \pm 1.17\,(\text{stat.}) \pm 2.66\,(\text{syst.})$ GeV. This measurement is perfomed selecting $t\bar{t}$ events with eμ final states in proton-proton collision data at $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity of 19.7 fb^{-1}. The measured top-quark mass is consistent with both the world average from measurements at the Tevatron and LHC of $m_t = 173.34 \pm 0.27\,(\text{stat.}) \pm 0.71\,(\text{syst.})$ GeV and the combined CMS result from Run 1 of the LHC of $m_t = 172.44 \pm 0.13\,(\text{stat.}) \pm 0.47\,(\text{syst.})$ GeV.

References

[1] ATLAS, CDF, CMS and D0 Collaborations, "First combination of Tevatron and LHC measurements of the top-quark mass", arXiv:1403.4427.

[2] CMS Collaboration, "Measurement of the top quark mass using proton-proton data at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV", arXiv:1509.04044.

[3] K. Agashe, R. Franceschini, and D. Kim, "Simple invariance of two-body decay kinematics", *Phys. Rev. D* **88** (2013) 057701, doi:10.1103/PhysRevD.88.057701.

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

[5] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[6] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations", *JHEP* **03** (2013) 015, doi:10.1007/JHEP03(2013)015, arXiv:1212.3460.

[7] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[8] CMS Collaboration, "Measurement of the underlying event activity at the LHC with $\sqrt{s} = 7$ TeV and comparison with $\sqrt{s} = 0.9$ TeV", *JHEP* **09** (2011) 109, doi:10.1007/JHEP09(2011)109, arXiv:1107.0330.

[9] S. Jadach, Z. Was, R. Decker, and J. H. Kuhn, "The tau decay library TAUOLA: Version 2.4", *Comput.Phys.Commun.* **76** (1993) 361–380, doi:10.1016/0010-4655(93)90061-G.

[10] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis", *JHEP* **07** (2002) 012, arXiv:hep-ph/0201195.

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

[12] N. Kidonakis, "Differential and total cross sections for top pair and single top production", doi:10.3204/DESY-PROC-2012-02/251, arXiv:1205.3453.

[13] N. Kidonakis, "Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution", *Phys.Rev.* **D82** (2010) 114030, doi:10.1103/PhysRevD.82.114030, arXiv:1009.4935.

[14] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through $O(\alpha_S^2)$", *Phys. Rev.* **D74** (2006) 114017, doi:10.1103/PhysRevD.74.114017, arXiv:hep-ph/0609070.

[15] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC", *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10–15, doi:10.1016/j.nuclphysbps.2010.08.011, arXiv:1007.3492.

[16] S. Agostinelli et al., "Geant4 - a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250 – 303, doi:http://dx.doi.org/10.1016/S0168-9002(03)01368-8.

[17] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in minimum bias and jet events from pp collisions at 7 TeV", CMS Physics Analysis Summary PFT-10-002, 2010.

[18] 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.

[19] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[20] A. L. Read, "Linear interpolation of histograms", *Nucl. Instrum. Meth.* **A425** (1999) 357–360, doi:10.1016/S0168-9002(98)01347-3.

[21] CMS Collaboration, "Calculation of Residual Energy Correction for b Jets Using Z+b Events in 8 TeV pp Collisions", CMS Physics Analysis Summary JME-13-001, 2014.

[22] 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.

[23] CMS Collaboration, "Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, ttbar and boosted topology events", CMS Physics Analysis Summary CMS-PAS-BTV-13-001, 2013.

[24] S. Alekhin et al., "The PDF4LHC Working Group Interim Report", arXiv:1101.0536.

[25] CMS Collaboration, "Measurement of the Differential Cross Section for Top Quark Pair Production in pp Collisions at $\sqrt{s}$ = 8 TeV", arXiv:1505.04480.

[26] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes", *Phys.Rev.* **D82** (2010) 074018, doi:10.1103/PhysRevD.82.074018, arXiv:1005.3457.

[27] CDF and D0 Collaborations, "Combination of CDF and D0 results on the mass of the top quark using up to 9.7 fb^{-1} at the Tevatron", arXiv:1407.2682.