

Object definition and performance at CMS

Carmen Diez Pardos for the CMS Collaboration

Deutsches Elektronen-Synchrotron (DESY), Notkestrasse 85, 22607 Hamburg, Germany

E-mail: carmen.diez@desy.de

Abstract. The performance of the different object reconstruction and identification algorithms in CMS has been studied on data collected in pp collisions at $\sqrt{s} = 7$ TeV and at $\sqrt{s} = 8$ TeV at the LHC. We present measurements of the reconstruction efficiencies, fake rates, and momentum scale and resolution, which are relevant for top physics analyses.

1. Introduction

Top quark physics requires a good understanding of all objects as the products of top decays involve jets, high momentum isolated leptons (in leptonic decay modes), soft leptons from b-jets, E_T^{miss} from leptonic W decays etc. Within the CMS [1] Collaboration, the Particle Flow (PF) approach [2] is used for the event reconstruction and widely used in top analyses. It combines information from all detector components, prior to jet clustering and missing energy calculation, to reconstruct particles (hadrons, photons, muons and electrons). From these particles composite objects, such as jets, taus or E_T^{miss} , are reconstructed. This yields to higher response and a big improvement in energy resolution and tau identification.

2. Leptons

2.1. Muon and electron

Muons are reconstructed using information from the silicon tracking and muon systems. Different algorithms are combined to reach a robust and efficient identification [3]. Electron reconstruction includes information from the tracker system and the calorimeters. Shower shapes, ratio of the energy deposited in the hadronic and electronic calorimeters, and conversion vetoes for electrons [4] are used for its identification. Lepton identification and isolation efficiencies are measured using the tag and probe approach with Z resonances. Fig. 1 shows an example of the reconstruction and identification efficiency for muons (as a function of p_T) and electrons in the central region of the detector (as a function of η). In both cases, there is a remarkable agreement between data and Monte Carlo and the reconstruction efficiencies are very high.

The momentum scale and resolution of muons are studied using $Z \rightarrow \mu\mu$ resonances and cosmic ray data, in the range $20 < p_T < 100$ GeV, where the momentum measurement is provided by the tracker. The average bias in the muon momentum scale was measured with a precision better than 0.2% and was found to be consistent with zero. The relative p_T resolution is between 1.3% to 2.0% for muons in the barrel and better than 6% in the endcaps. The resolution measured from di-electron resonances is around 1–2% in the barrel region and better than 4% in the endcaps.



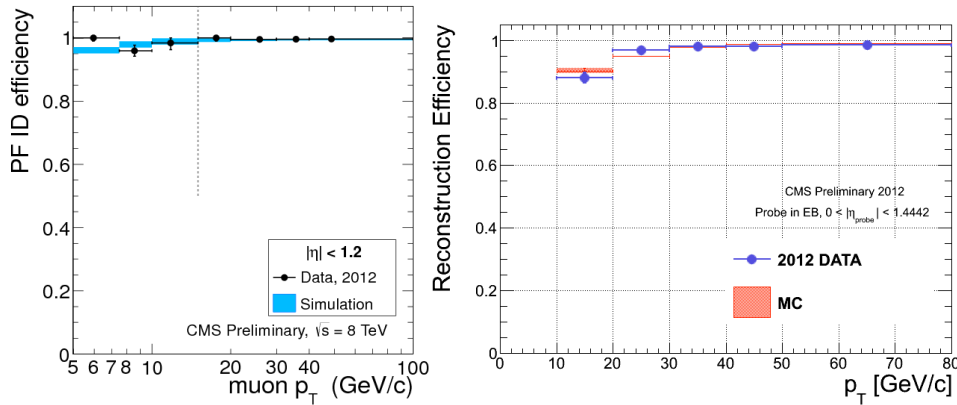


Figure 1. Muon (left) and electron (right) reconstruction efficiency in the central region of the detector obtained with a sample of $Z \rightarrow ll$ at $\sqrt{s}=8$ TeV.

2.2. Taus

τ -leptons decay predominantly hadronically, being their products one or three charged mesons (mostly π^+ and π^-), up to three neutral pions, and τ neutrino. The main algorithm used at CMS to reconstruct taus is the Hadron plus Strips (HPS) algorithm [5]. The HPS combines PF electromagnetic particles in strips, to take into account possible broadening of calorimeter depositions from photon conversions. The neutral objects are then combined with existing charged hadrons to reconstruct the hadronic tau decay products.

To measure the efficiency of its reconstruction and identification in data, a sample of $Z \rightarrow \tau\tau \rightarrow \mu\tau_{\text{had}}$ events enriched in real taus was used with relatively small contribution from other processes. The identification efficiency results approximately 50%, keeping the misidentification rate for jets at the level of $\sim 1\%$, see Fig. 2 left. The tau energy scale measured with the same method was found to be close to unity with the measured energy scale uncertainty better than 3%, see Fig. 2 right.

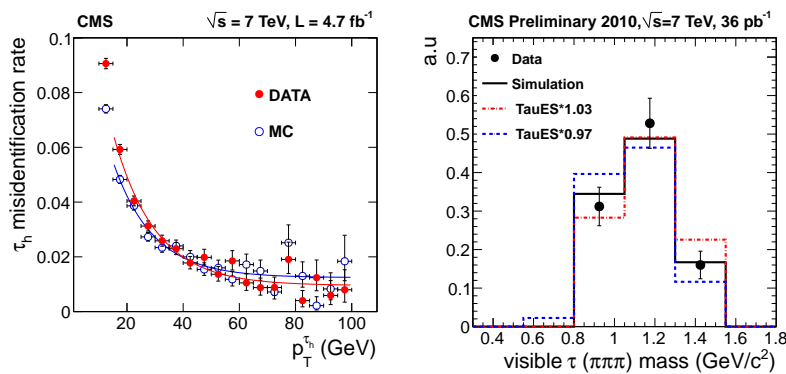


Figure 2. Left: fake-rate measured in data (7 TeV) compared to MC prediction. Right: the reconstructed invariant mass of taus decaying into three charged pions compared to predictions of the simulation.

Isolated electrons passing the identification/isolation criteria of the algorithm, misidentified as pions originating from τ_h are also an important source of background in many analyses with

τ_h in the final state. The misidentification rate, found by applying the tag and probe method with a $Z \rightarrow ee$ sample, is smaller than 2%. For muons, the misidentification rate is below 0.2%.

3. Jets

Jets in CMS are reconstructed using PF candidates with the anti-kT algorithm [6] with a spatial separation parameter of $R = 0.5$. The CMS Collaboration uses a factorised approach for jet calibration to combine the advantages of Monte-Carlo simulation based studies with robust data-driven methods [8]. First, an offset correction is applied, to remove the additional energy inside jets due to pile-up events and electronic noise. Second, a MC calibration factor is used to correct the non-uniformity and non-linearity of energy response in pseudorapidity and jet transverse momentum, respectively. This factor is estimated from MC simulation of multijet events by requiring the energy of the generator jets to match that of the reconstructed jets. Fig. 3 shows the jet energy response before and after applying PU and MC truth corrections.

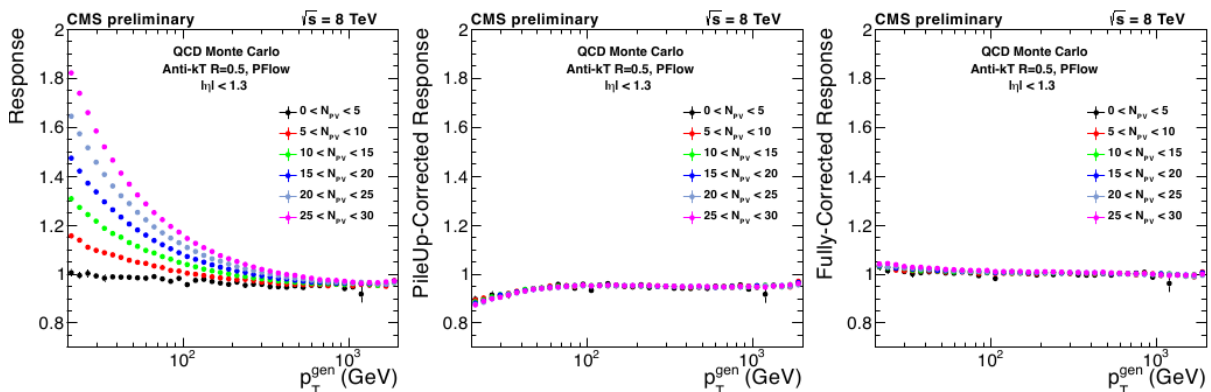


Figure 3. Jet energy response before applying any correction (left), after the pile-up correction (middle) and the corrections to the generator level (right), which show closure over the whole p_T range.

Finally, the residual corrections account for the small differences between data and simulation. The dijet p_T balance technique [8] is used to measure the energy of a jet at any $|\eta|$ relative to the other jet identified in the central region with $|\eta| < 1.3$. The relative jet energy response measured in data is compared to one obtained from the simulation and the correction is derived to correct for the observed small differences. The absolute jet energy response is measured in the reference region of $|\eta| < 1.3$ using γ^*/Z +jet events in data and MC simulation where a jet energy scale is derived by exploiting the balance in transverse momentum between a boson and a jet in the final state.

The uncertainties on the transverse momentum of a corrected jet strongly depend on its p_T and η , see Fig. 4. The jet energy scale is determined with a precision of about 1% in the central part of the detector and better than 3% in the endcaps.

4. B-tagging

The ability to identify the jets from the hadronisation of heavy flavour quarks plays a central role in precision measurements of top quark properties. Flavour tagging algorithms are thus used to identify these jets and reduce the otherwise overwhelming background from processes involving jets from gluon and light quarks fragmentation and from c-quark hadronisation. Analyses also need an accurate estimate of the performance (efficiency and misidentification rate) of the chosen

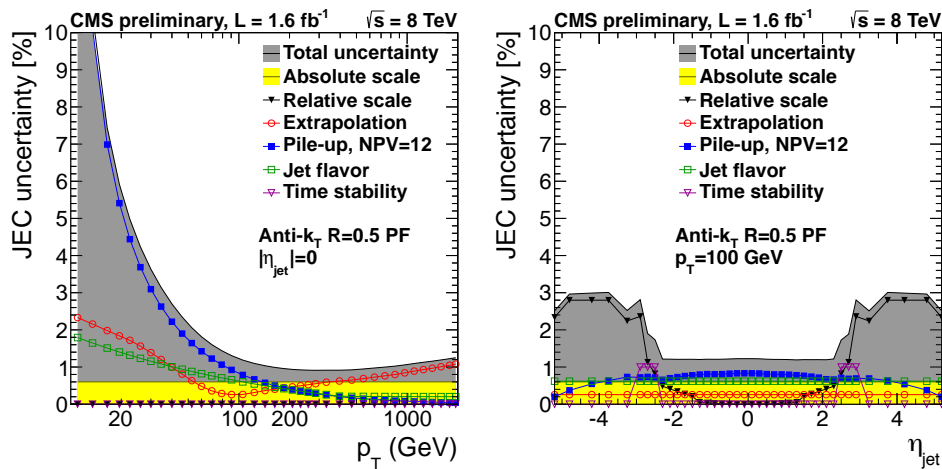


Figure 4. Uncertainty of the final jet energy corrections as a function of p_T (left) and η (right) split up into the considered sources of systematic uncertainties [7].

algorithm. The main features to identify b-quarks are their large lifetime (~ 1.5 ps) and their decay length (~ 1.8 mm).

Different algorithms are used, based on the track impact parameter calculated in three dimensions or/and the presence of a secondary decay vertex and the kinematic variables associated to it. These variables are for instance the flight distance and direction or properties of the system of associated tracks (multiplicity, energy, mass) [10]. Fig. 5 presents the performance of various identification algorithms obtained from MC simulation.

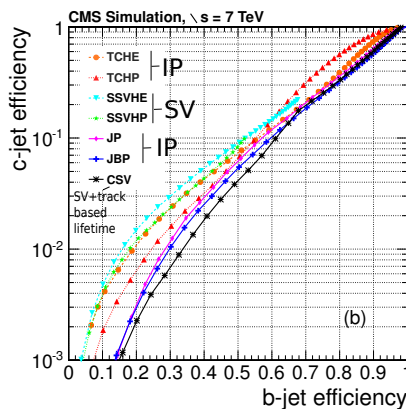


Figure 5. Curves of c-efficiencies as a function of the b-efficiency obtained from simulation for the different algorithms developed by the CMS Collaboration.

The performance of these taggers is studied using multijet events and $t\bar{t}$ events with different methods. Due to a large b-quark mass, the muon momentum component transverse to the jet axis (p_{rel}) or impact parameter (IP) of the muon track is larger for muons from b hadron decays than that for muons in light-flavour jets. Modelling of p_{rel} and IP in MC simulation to represent distributions expected from b-jets is used as a template and that is compared to the distribution observed in multijet data events. The distribution of p_{rel} is used to measure the tagging efficiency for low/average- p_T jets, while that of IP allows to measure efficiency for

high- p_T jets. The measurements from different methods are combined based on the weighted mean taking into account the correlation between uncertainties. Individual results are combined to provide the optimal measurement of data/MC SF for $30 < p_T < 670$ GeV. The results are shown in Fig. 6, left. Various methods using $t\bar{t}$ events [9] in the lepton+jets and dilepton channels provide inclusive results suited for measurements other than the cross section, see Fig. 6 right.

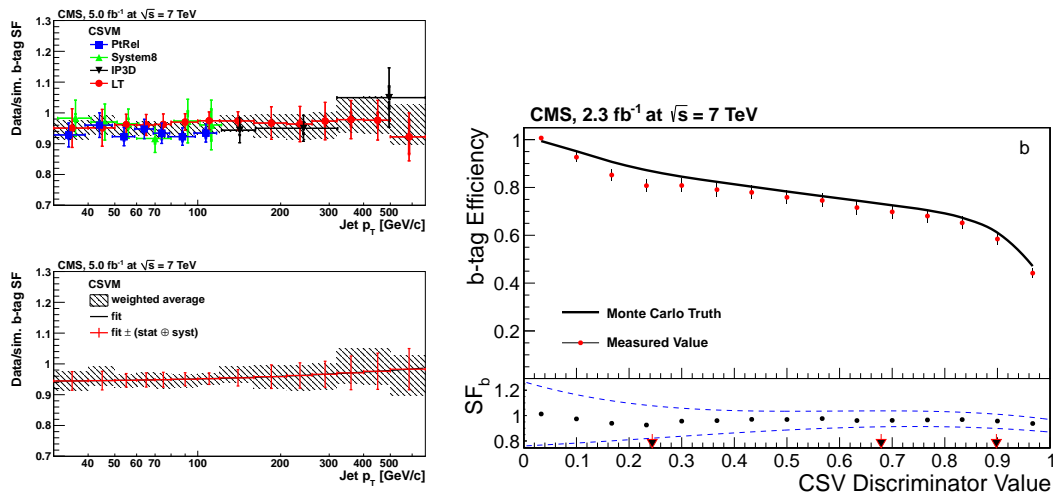


Figure 6. Left: individual (top) and combined (bottom) measurements of the ratio of efficiencies in data and simulation for the CSV method. Thick lines indicate the statistical error, while thin lines correspond to the combined statistical and systematic uncertainties. Top analyses apply those scale factors to correct MC yields. Right: measured b-tagging efficiency as a function of the discriminator threshold for the CSV algorithm from data and predicted from simulation are shown in the upper histogram. The scale factors are shown in the lower histogram. The total uncertainties are represented by a blue band for the scale factor. The arrows indicate the standard operating points.

5. Missing transverse energy

Events are reconstructed with the Particle Flow technique: E_T^{miss} is computed as the negative vectorial sum of all particle candidates, including the JES correction. It is a challenging variable as it is easy to obtain fake E_T^{miss} due to large shower fluctuations, non-linear calorimeter response, instrumental noise or poorly instrumented areas. The missing transverse energy resolution is studied in $Z \rightarrow \mu\mu$ events. In a well-measured event, the transverse momentum of a boson and hadronic activity should be balanced: $q_T^Z + u_T + E_T^{\text{miss}} = 0$ where q_T^Z is the transverse momentum of the boson and u_T is the hadronic recoil, defined as the vector sum of the transverse momenta of all particles except the vector boson. The projection of the latter into the axis parallel to the Z boson momentum, u_{\parallel} , is used to obtain the energy correction to the E_T^{miss} in data. The E_T^{miss} resolution is assessed as the standard deviation of the perpendicular projection of the hadronic recoil to the Z transverse momentum ($\sigma(u_{\perp})$) [11, 12].

The E_T^{miss} resolution as a function of the number of primary vertices is shown in Fig. 7, left. Compared to previous data taking periods, an improved resolution for a fixed number of primary vertices was achieved in 2012 by the use of dedicated pile-up corrections. Furthermore, improved calorimeter reconstruction techniques have reduced the impact of pile-up on the E_T^{miss} measurements. The resolution degradation per additional primary vertex only amounts to 3 GeV. The correction applied to jets are propagated to E_T^{miss} to restore the consistency of

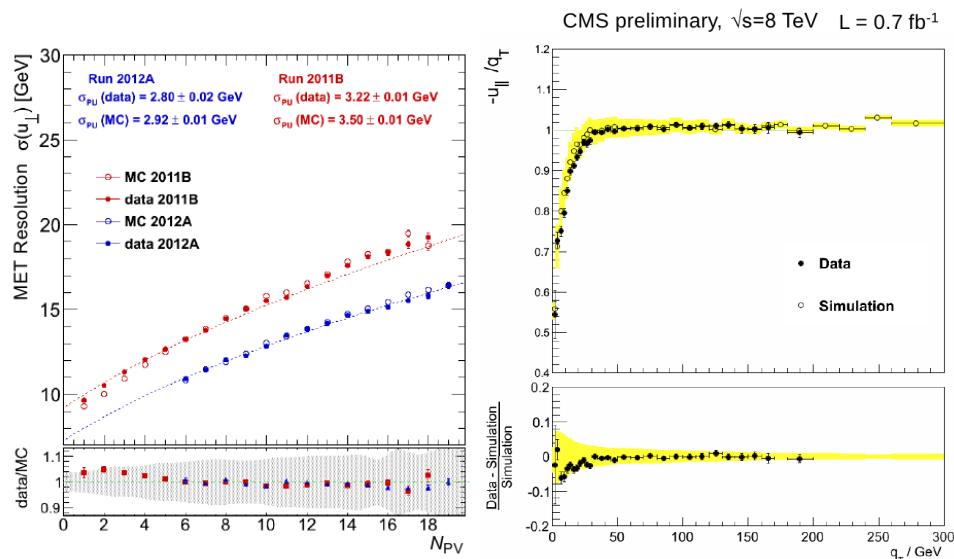


Figure 7. E_T^{miss} resolution, $\sigma(u_{\perp})$, as a function of the number of primary vertices (left) and corrected response, u_{\parallel}/q_T , as a function of transverse momentum of the Z boson q_T (right).

the event. This corrected E_T^{miss} is presented in Fig. 7 right. The response is very close to unity over a wide p_T range and the results in data agree with Monte Carlo simulation.

6. Conclusions

The CMS Collaboration has developed a number of sophisticated algorithms to identify jets, b-jets, leptons, etc. These algorithms are studied with data and simulation, showing a great performance of the object identification both at 7 and 8 TeV and a remarkable agreement between data and simulation. Furthermore, CMS is performing well in the LHC environment, adapting to the increasing luminosity (applying PU dependent corrections, developing customised isolation), which allows for precision measurements of top quark properties.

References

- [1] CMS Collaboration, The CMS experiment at the CERN LHC, JINST 03 (2008) S08004
- [2] CMS Collaboration, Commissioning of the Particle Flow Event Reconstruction with the First Collisions Recorded in the CMS Detector, CMS-PAS-PFT-10-001 (2010).
- [3] CMS Collaboration, Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV, JINST 7 (2012) P10002
- [4] CMS Collaboration, CMS-PAS-EGM-10-004, CMS-PAS-EGM-10-001
- [5] CMS Collaboration, Performance of tau reconstruction algorithms in 2010 data collected with CMS, CMS-PAS-TAU-11-001.
- [6] M. Cacciari, G. P. Salam, and G. Soyez, JHEP 04, 063, (2008).
- [7] CMS Collaboration, Jet Energy Corrections and Uncertainties, CMS DP-12-012.
- [8] CMS Collaboration, Determination of jet energy calibration and transverse momentum resolution in CMS, 2011 JINST 6 P11002.
- [9] CMS Collaboration, Measurement of the b-tagging efficiency using $t\bar{t}$ events, CMS PAS-BTV-11-003.
- [10] CMS Collaboration, b-Jet Identification in the CMS Experiment, PAS-BTV-11-004.
- [11] CMS Collaboration, E_T^{miss} Performance in 2011 CMS Data, CMS DP-12-003.
- [12] CMS Collaboration, E_T^{miss} Performance in 2012 RunA CMS Data, CMS DP-12-013.