Identification of jets, τ leptons, and missing transverse energy at CMS

Ketino Kaadze (for the CMS collaboration)
European Organization For Nuclear Research
CERN - PH Department, E19600
CH 1211 Geneva 23, Switzerland

1 Introduction

The CMS is one of the general purpose high energy physics detectors operating at CERN, collecting pp collision data delivered by the Large Hadron Collider. Its detailed description can be found elsewhere [1].

Data recorded by the CMS detector are successfully used to perform a wide range of physics analyses including both measurements of the standard model (SM) processes and searches for the Higgs boson and beyond SM phenomena. A crucial ingredient to the success of the CMS physics program is an excellent performance of the reconstruction and identification algorithms of final state physics objects. CMS uses the particle-flow (PF) technique [2, 4] to reconstruct all final-state particles (PF candidates) produced in the collision event using information from all CMS sub-detectors. The PF candidates are further classified as charged hadrons, neutral hadrons, electrons, muons, or photons and are used with different algorithms to identify final state physics objects: jets, τ leptons, *etc.* In this article I describe the identification of jets, b jets, τ leptons, and an imbalance of the transverse energy, often referred to as missing transverse energy ($\not\!\!E_{\mathrm{T}}$) or MET. The performance of the object identification is measured both in data and in the Monte Carlo (MC) simulation. The MC samples are passed through a detailed simulation of the CMS detector based on GEANT [3] and both data and MC sample are subject to the same reconstruction chain and selection criteria.

2 Jets

Jets are the experimental signatures of quarks and gluons produced in high energy processes, such as hard scattering of partons in pp collisions. Jets are reconstructed using PF candidates with the anti-kT algorithm [5] with a spatial separation parameter of R = 0.5. Jet reconstruction performance is evaluated in MC: at first, generated jets in the event are identified using generator-level information in the same anti-kT

algorithm; then, to estimate the reconstruction efficiency one checks if the reconstructed jet is spatially matched to the generator-level jet using a separation in $\eta - \phi$ plane $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The ability to measure energy of jets is of of crucial importance for many physics analyses at CMS. Jet energy calibration is performed in order to relate energy measured in the detector jet to the energy of the true particle jet. The total correction is a multiplicative factor and is a product of different components: the offset correction (C_{offset}), the MC calibration factor (C_{MC}) and the residual calibrations, (C_{rel}) and (C_{abs}) for the relative and absolute energy scales, respectively:

$$\mathcal{C} = C_{\mathrm{offset}} \cdot C_{\mathrm{MC}}(p'_T, \eta) \cdot C_{\mathrm{rel}}(\eta) \cdot C_{\mathrm{abs}}(p''_T), \tag{1}$$

where p'_{T} is the transverse momentum of the jet after applying the offset correction and p''_{T} is the p_{T} of the jet after all previous corrections.

The offset corrections are used to subtract the average contribution from particles from other proton-proton collisions in the same beam crossing (pileup). Two different methods are used to estimate offset corrections at CMS: jet area method [6] and average offset method, yielding very similar performance as a function of pileup as shown on Fig. 1 (left). The MC calibration factor is 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 multi-jet events by requiring the energy of the "generator jets" to match that of the reconstructed jets. Finally, the residual corrections account for the small differences between data and simulation. The dijet p_{T} balance technique [7] is used to measure the energy of a jet at any η 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.

Uncertainties from different systematic sources that affect the jet energy scale as a function of p_{T} and η are given in Fig. 1.

3 b jets

Jets arising from b quark hadronization and decay are present in a wide range of physics processes, such as the decay of top quarks, Higgs bosons, and are expected in a wide range of physics beyond the SM, such as supersymmetry. The ability to accurately identify b jets is vital in reducing the otherwise overwhelming background from gluon (g) jets, light-flavour quarks (u, d, s) and from c quark fragmentation.

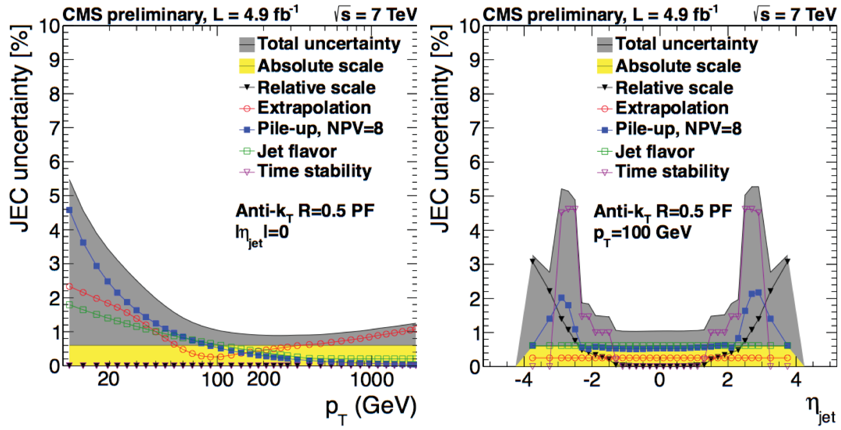

Figure 1: Jet energy scale uncertainties from different sources and the total uncertainty as a function of p_{T} (left) and η (right).

Characteristic features of b quarks are the relatively large mass, high track multiplicity, and high p_{T} of decay products with respect to b-flight direction, and long lifetime of the heavy flavour hadrons. These features are used to develop algorithms that allow distinguishing b jets from those produced by light-flavor quarks and gluons.

Important variables that hold large discriminating power to distinguish decay products of a b hadron from prompt tracks are the impact parameter (IP) of a track with respect to the primary vertex and secondary decay vertex and kinematic variables associated with this vertex. The IP is calculated in three dimensions, taking advantage of the excellent resolution of the pixel detector along the z axis. The variables associated to the secondary vertex are the flight distance and direction, based on the vector between primary and secondary vertex, and various properties of the system of associated secondary tracks such as the multiplicity, the mass or the energy. These variables are used to build simple discriminator variables, such as the *Track Counting* and *Simple Secondary Vertex* algorithms [8]. Besides these, more complex methods are developed that are based on a likelihood estimation using either an IP significance of several tracks in a jet or combining secondary vertices with track-based lifetime information. These yield more complicated identification techniques, such as *Jet Probability* (JP), *Jet B Probability*, or *Combined Secondary Vertex* taggers [8]. Performance of these identification algorithms obtained from MC simulation are shown in Fig. 2.

Performance of these taggers is also measured *in-situ* using various samples and methods. Due to a large b quark mass, the muon momentum component transverse to the jet axis (p_{T}^{rel}) or IP of the muon track is larger for muons from b hadron decays than that for muons in light-flavour jets. Modeling of p_{T}^{rel} and IP in MC simulation to represent distributions expected from b jets is used as a template and that is

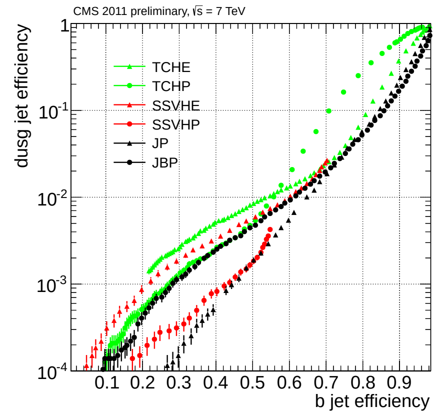

Figure 2: Performance curves obtained from the MC simulation for the algorithms described in the text: light flavour efficiencies as a function of the b efficiency.

compared to the distribution observed in multijet data events. The distribution of p_{T}^{rel} is used to measure 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 weighted mean taking into account the correlation between uncertainties.

A large sample of pair-produced top quark events is also used to measure b-tagging efficiency using several methods. The profile likelihood ratio (PLR) method is used to measure b-tagging efficiency using dilepton final state of $\mathrm{t\bar{t}}$ decay. The method uses 2-dimensional distribution of the jet multiplicity versus the b-tagged jet multiplicity. The Flavor Tag Consistency method (FtCM) requires consistency between the observed and expected number of tags in the lepton+jet events from $\mathrm{t\bar{t}}$ process. Here as well, the results from different methods are combined using weighted mean [9].

4 Hadronically decaying τ leptons

The products of the hadronically decaying τ lepton (τ_{had}) are one or three charged mesons (mostly π^+ and π^-), up to three neutral pions, and tau neutrino. Thus, its signature is similar to that expected from jets. The main reconstruction algorithm for τ_{had} used at CMS is Hadron plus Strips (HPS) algorithm [10]. It searches for PF jets consistent with τ lepton decaying to hadrons. The algorithm employs an optimized π^0 reconstruction method that accounts for a wider energy deposition profile of two spatially-close photons from $\pi^0 \to \gamma\gamma$ decay.

The efficiency of τ_{had} reconstruction and identification is measured in data using a sample of $Z \to \tau\tau$ events, where one of the τ leptons decay hadronically, and the other one as $\tau \to \mu\nu_\mu\nu_\tau$. The events are preselected using a set of requirements to suppress the backgrounds from $Z \to \mu\mu$, W+jets, and multijet events, but without applying the τ_{had} identification algorithms. The $\mu\tau_{\rm jet}$ invariant mass distributions for those events which pass or fail the τ_{had} identification are fit using signal and background templates estimated in Monte Carlo simulation. The efficiency is calculated as: $\varepsilon = N^{Z\to\tau\tau}_{\rm pass}/(N^{Z\to\tau\tau}_{\rm pass} + N^{Z\to\tau\tau}_{\rm fail})$, where $N_{\rm pass}$ ($N_{\rm fail}$) is the number of events, after subtracting background contributions, that pass (fail) the τ_{had} identification criteria. The jet to τ_{had} misidentification rate is measured on a sample of either multijet or W+jets events. The measured misidentification rate as a function of efficiency is shown in Fig. 3.

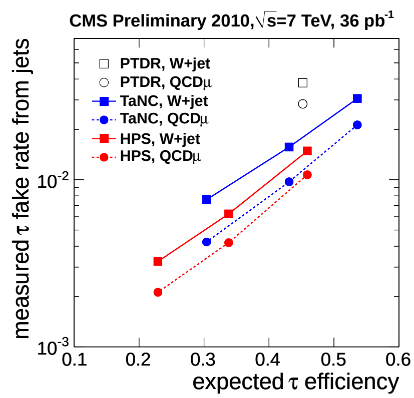

Figure 3: The measured jet to τ_{had} misidentification rate as a function of MC estimated efficiency for all working points for μ-enriched multijet and W+jet data samples. Performance of another algorithm (TaNC) is also shown.

5 Missing Transverse Energy

The missing transverse energy is associated with production of particles that escape detection: neutrinos as well as a number of exotic particles predicted by extensions of the SM. Thus, MET is of crucial importance for a number of physics analyses. The reconstruction of $\not\!\!E_{\rm T}$ is very sensitive to the particle momentum measurements, detector malfunctions, particles impinging on poorly instrumented regions of the detector, beam-halo particles, and cosmic-ray particles.

The $\not\!\!E_{\rm T}$ is computed using all PF candidates in event [11]. The effects due to energy scale correction for jets and other physics objects are taken into account. The missing transverse energy resolution is studied in Z$\to$ $\mu\mu$ events that is a source of clean final state with two energetic, isolated muons and no intrinsic missing transverse energy. Events containing signatures of instrumental noise are rejected. In well-measured event transverse momentum of a boson and hadronic activity should be balanced:

$$\vec{p}_{\mathrm{T}} + \vec{u}_{\mathrm{T}} + \vec{\not{E}}_{\mathrm{T}} = 0 \tag{2}$$

where, p_{T} and u_{T} are transverse momentum of the boson and hadronic recoil, defined as the vector sum of the transverse momenta of all particles except the vector boson, respectively. The projections of u_{T} onto axis, parallel and perpendicular to the boson momentum direction, are noted as $u_{||}$ and $u_{\perp}$, respectively. The former variable is used to obtain the energy correction to the MET in data. The MET resolution is assessed by measuring the RMS spread of $u_{||}$ and $u_{\perp}$ about their mean values, after correcting for the energy scale. The MET resolution as a function of number of pileup vertices is shown in Fig. 4. The resolution worsens in later data of 2011 due to larger pileup contribution.

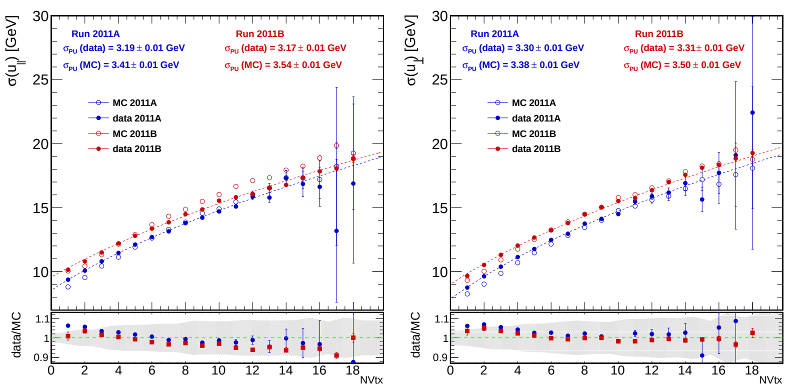

Figure 4: The width of $u_{||}$ and $u_{\perp}$ distributions in data and MC simulation as a function of the number of primary vertices for 2011 different data-taking period.

6 Conclusion

In summary, CMS collaboration has developed a number of sophisticated and well-performing algorithms to identify jets, b jets, τ leptons and missing transverse energy. The performance of these objects is measured using pp collision data and the MC simulation is found to describe the performance in data well. Great performance of the object identification offers unique opportunity to explore a tremendous physics program at CMS.

References

[1] S. Chatrchyan *et al.*, JINST **03**, S08004, (2008).

[2] CMS Collaboration, "Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary, CMS-PAS-PFT-09-001, (2009).

[3] S. Agostinelli *et al.*, "Geant4 a simulation toolkit", Nucl. Instrum. Math. A **506**, 250, (2003).

[4] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and jet Events from pp Collisions at 7 TeV", CMS Physics Analysis Summary, CMS-PAS-PFT-10-002, (2010).

[5] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063, (2008).

[6] M. Cacciari and G. Salam, "Pileup subtraction using jet areas", Phys. Lett. B **659**, 119, (2008).

[7] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", J. Instrum. **6**, P11002, (2011).

[8] CMS Collaboration, "b-Jet Identification in the CMS Experiment", CMS Physics Analysis Summary, CMS-PAS-BTV-11-004, (2011).

[9] CMS Collaboration, "Measurement of btagging efficiency using $t\bar{t}$ events", CMS Physics Analysis Summary, CMS-PAS-BTV-11-003, (2011).

[10] CMS Collaboration, "Performance of τ-lepton reconstruction and identification in CMS", JINST **07**, 01001, (2012).

[11] CMS Collaboration, "Missing transverse energy performance of the CMS detector", JINST **6** 09001, (2011).