Performance of jets at CMS

Matthias Schröder, on behalf of the CMS collaboration
Deutsches Elektronen-Synchrotron (DESY), Notkestraße 85, 22607 Hamburg, Germany
E-mail: matthias.schroeder@desy.de

Abstract. The calibration and reconstruction of jets critically relies on the performance of the calorimeters. Extending out to large pseudorapidities, the measurements depend on the interplay between forward calorimeters, central calorimeters, and the tracking system. The high number of additional pile-up interactions poses further complications. In CMS, these difficulties are overcome using the ‘particle-flow’ approach, which aims at reconstructing individually each particle in the event prior to the jet clustering. Measurements of the jet energy scale and the procedure for jet energy calibration in CMS are reviewed, which are performed with dijet, photon + jet, and Z + jet data collected in proton-proton collisions at a centre-of-mass energy of 8 TeV, corresponding to an integrated luminosity of 19.6 fb$^{-1}$. The effect of pile-up interactions and the state of the art mitigation techniques used in CMS as well as the main sources of uncertainty of the jet energy calibration are also presented.

1. Introduction
Quarks and gluons produced in high-energy processes, such as the proton-proton collisions at the CERN Large Hadron Collider (LHC), manifest as jets. Hence, jets provide experimental access to the underlying partonic process, and the precise understanding of their properties is of crucial importance in many physics analyses.

Here, an overview of the performance of jets at the Compact Muon Solenoid (CMS) experiment [1] at the LHC is given. The central feature of the CMS detector is a 6 m-diameter, superconducting solenoid providing a 3.8 T magnetic field. Within the magnet, closest to the interaction point, a silicon tracking detector is installed that measures the transverse momenta ($p_T$) of charged particles with a relative resolution of $\approx 0.7\%$ at 10 GeV up to pseudorapidities of $|\eta| < 2.5$. Surrounding the tracking volume but still within the solenoid, the calorimetric system covers the region up to $|\eta| < 3$. It consists of a lead-tungstate crystal electromagnetic calorimeter (ECAL) with an energy resolution of $\approx 0.6\%$ for 50 GeV electrons and a brass-scintillator hadronic calorimeter (HCAL) with a resolution of 18% for 50 GeV pions. Steel and quartz-fibre hadron calorimeters extend the coverage to $|\eta| < 5$. Muons are identified in gas-ionisation detectors embedded in the steel flux-return yoke of the magnet.

In the following, the current status of the jet reconstruction and calibration at CMS is discussed, and pile-up mitigation techniques as well as tools to discriminate quark from gluon jets are reviewed. The presented results have been obtained with data collected during 2012 at a centre-of-mass energy of 8 TeV, corresponding to an integrated luminosity of up to 19.6 fb$^{-1}$. 
2. Jet reconstruction at CMS
In most CMS analyses, events are reconstructed using a particle-flow (PF) algorithm, which aims at identifying and reconstructing the individual particles in the event, combining the information from all subdetectors [2]. For example, the energy of photons is directly obtained from the ECAL measurements; conversions in the dense tracker material are accounted for by adding tracking information. The energy of charged hadrons is determined from a combination of the track momenta with associated ECAL and HCAL energy deposits, and the energy of neutral hadrons is determined from the remaining calorimeter energy deposits. In both cases, the measurements are corrected for the non-linear response of the calorimeters to hadrons.

Jets are then clustered from the four-momenta of the reconstructed particles. Commonly, the anti-$k_T$ jet clustering-algorithm [3] with a radius parameter of $R = 0.5$ is used in CMS, but other jet algorithms and reconstruction methods are also supported.

3. Properties of particle-flow jets
The PF-jet performance greatly benefits from two effects: Firstly, approximately 90% of the measured jet energy are reconstructed as charged hadrons and photons, which are measured dominantly with the high-resolution tracker and ECAL, while only the remaining 10% are reconstructed as neutral hadrons, which are measured with the HCAL alone that has a poorer resolution [4]. Secondly, since individual particles are reconstructed, a dedicated calibration can be applied. As a consequence, the impact of the non-compensating calorimeters is mitigated, leading to a good overall jet response with flavour differences less than 4% [4]. Furthermore, the jet energy resolution is greatly improved compared to jets reconstructed from calorimeter-information only [5].

Since tracking and thus vertex information is available, the impact of pile-up (PU) collisions that occur in addition to the hard interaction in the event can be mitigated already at the jet-clustering stage. As shown in Fig. 1, the additional $p_T$ offset per jet due to PU particles amounts to approximately 700 MeV per primary vertex, leading to an average of $\approx 10$ GeV being added to the $p_T$ of each jet\(^1\). The average $p_T$ offset per primary vertex is shown in Fig. 2 as function of $\eta$, split up into the contributions from different PF-particle types. In the central detector region with tracker coverage, approximately 50% of the $p_T$ offset is deposited by charged particles that can be unambiguously associated to PU vertices. These particles can be subtracted from the event prior to the jet clustering, thus effectively reducing the impact of PU.

4. Jet energy corrections
The jet energy corrections (JEC) relate the energy of the reconstructed jets on average to the energy at particle-level, which is independent of the detector response. CMS has adopted a factorised approach to JEC [5]. Given that the jet properties are well described by the simulation, the primary JEC factors are derived from simulated events relative to the generator jets, and only the remaining, small differences between the response in data and simulation are corrected for using data-driven methods. Since the latter corrections are derived as data-to-simulation ratios, potential biases of the methods cancel to first approximation.

The first factor in the JEC chain subtracts the $p_T$ offset due to (the remaining) PU contributions. It is an average correction derived from the global per-event $p_T$ offset density $\rho$ and the jet area [5, 6]. The calibration of $\rho$ is determined as a function of $\eta$ and $p_T$ from two samples of simulated events with identical hard interactions, where one contains in addition PU interactions. Residual differences to the $\rho$ calibration in data are measured as a function of $\eta$ with the random-cone method in zero-bias events [5] and corrected for.

\(^1\) Taking into account an average of 21 PU collisions during the 2012 LHC run and the vertex-reconstruction efficiency of approximately 70%.
The second, primary JEC factor corrects for the \( p_T \) and \( \eta \) dependence of the average jet response due to the calorimeter non-linearities and \( p_T \) thresholds, the different detector properties along \( \eta \), and geometric effects. It is obtained from simulated QCD-multijet events generated with Pythia6 [7] with tune Z2∗ processed through the full Geant4-based [8] CMS detector simulation.

After these corrections, the jet energy scale (JES), i.e. the average response, in simulated events is at 1 within 1% precision [4]. The data JES is validated using the measurements described in the following, and the residual differences to the simulation are converted into the final JEC factor, which is only applied to data.

4.1. Absolute scale from \( Z(\mu\overline{\mu}) + \text{jet} \) events

The JES can be measured in events where the jet is balanced against a well-measured reference object. CMS uses \( Z + \text{jet} \) events with \( Z \rightarrow \mu\overline{\mu} \), where the \( p_T \) of the \( Z \) boson (\( p_T^Z \)) can be measured with high precision (at the 0.1% level) compared to the jet due to the excellent muon-\( p_T \) resolution of the CMS detector. \( Z(\text{e}\overline{\text{e}}) + \text{jet} \) and photon + jet events are used as a cross check [5].

The jet is required to be in the central detector region \( |\eta| < 1.3 \) and back-to-back in azimuth with the \( Z \) boson. Two different estimators of the jet response are defined,

\[
R_{\text{bal}} = \frac{p_T^{\text{jet}}}{p_T^Z}, \quad (1) \quad R_{\text{MPF}} = 1 + \frac{\vec{E}_T \cdot \vec{p}_T^Z}{(p_T^Z)^2}. \quad (2)
\]

For the \( p_T \)-balance method (1), the jet is assumed to fully capture the fragmentation products of the initial parton that balances reference object. For the missing transverse energy projection fraction (MPF) method (2), it is assumed that the missing transverse energy (\( \vec{E}_T \)) in the event solely arises from mismeasurements of the hadronic recoil to the reference object, such that the jet response is given by the projection of \( \vec{E}_T \) on the reference object’s \( \vec{p}_T \).

The absolute JES is defined as the mean of the \( R_{\text{bal}} \) and \( R_{\text{MPF}} \) distributions, respectively. It depends on the amount of additional hadronic activity in the event, which affects the \( p_T \) balance between the jet and the reference object. This is clearly visible when measuring the JES as a function of \( \alpha = p_T^{\text{jet2}}/p_T^Z \), the fractional \( p_T \) of the second-leading jet, cf. Fig. 3 (left). The \( R_{\text{MPF}} \)
is less affected than $R_{\text{had}}$ since it depends less on the association of the particles to the leading jet.

The residual JEC factor is determined as the data-to-simulation response ratio\(^2\). It is measured for different $\alpha$ and extrapolated linearly to $\alpha = 0$ to suppress the influence of the additional hadronic activity on the result. The correction is at the level of 1% in the central detector region, demonstrating the good description of the response by the simulation. A validation of the procedure in Fig. 3 shows excellent agreement of the JES in data and simulation after all corrections.

**Figure 3.** JES in data and simulation (left) and their ratio (right) after all corrections vs. $p_T^\text{jet2}/p_T^Z$ measured in $Z + \text{jet}$ events using the MPF and the $p_T$-balance response estimators. The JES is > 1 due to the different flavour composition compared to QCD-multijet events. [9]

**Figure 4.** Relative residual JEC vs. $\eta$ determined from dijet events using the MPF response-estimator. [9]

### 4.2. Relative scale from dijet events

Due to their relatively small production cross-section, $Z + \text{jet}$ events are only used to measure the absolute JES in the central detector region. The JES in the forward regions is obtained from QCD-dijet events, which occur at a much higher rate [5]. They are selected requiring the two leading jets in the event to be back-to-back in azimuth, with at least one jet being in the central region $|\eta| < 1.3$. Even though there is no well-measured reference object, the two jets are expected to be balanced at parton level, and therefore, the scale relative to the JES of the central jet can be determined using slightly modified versions of the response estimators (1) and (2). The observed JES differences are converted into a relative residual JEC, which is below 5% up to $|\eta| < 2.5$, cf. Fig. 4.

### 4.3. Jet energy correction uncertainty

After application of the described corrections, the JES at CMS is known to percent-level precision even in the difficult forward regions and better than 1% for jets above 100 GeV in the central detector region, cf. Fig. 5. At low $p_T$, the uncertainties are dominated by the PU-correction uncertainty. At higher $p_T$, the dominant uncertainties are due to the fact that the simulation is used to extrapolate the $Z + \text{jet}$-derived JES to different flavour mixtures and $p_T$ regions. In the forward regions with $|\eta| > 2.5$, the largest contribution stems from the uncertainty of the relative residual JEC, which is mostly due to the limited number of available dijet events.

\(^2\) The MPF-based value, which is less affected by additional hadronic activity, is used for the JEC; it agrees with the $p_T$-balance result.
5. Pile-up jet identification

In addition to contributing extra energy to the jets from the hard interaction, particles from PU collisions may also cluster and give rise to jets, referred to as PU jets. Even though these are typically very soft, they can overlap and form jets with sizable $p_T$, which occurs at a rate that grows quadratically with the number of PU collisions [10]. PU jets constitute a significant fraction of all low-$p_T$ jets in the data collected during the 2012 LHC run, e.g. approximately 50% of all 30 GeV jets in $Z +$ jets events are PU jets, cf. Fig. 6, and this contribution will become even larger for the upcoming LHC runs with a much increased instantaneous luminosity.

Figure 5. Relative JEC uncertainty and its components vs. $p_T$ and $\eta$. [9]

Figure 6. Jet $p_T$ spectrum in $Z +$ jets events in data and simulation, for which the contributions from PU and non-PU jets are indicated. [10]

Figure 7. Quark-gluon jet discriminant in $Z +$ jets events in data and simulation, for which the different flavour contributions are shown. [11]

CMS has developed a likelihood-based method to identify and reject PU jets [10]. The likelihood discriminator combines twelve jet-shape and vertex-information related variables that are sensitive to the characteristic substructure of PU jets, which are typically somewhat diffuse and contain many particles from PU vertices. The performance of the discriminator has been evaluated with simulated $Z +$ jets events. A high PU-jet rejection efficiency is achieved of up to
90 – 95% for central jets with $|\eta| < 2.5$ while retaining 99% of the non-PU jets from the hard interaction; the performance degrades somewhat in the forward regions.

6. Quark-gluon jet discrimination

CMS has developed a tool to distinguish between jets initiated by light quarks and by gluons, which makes use of the PF event-reconstruction and exploits the different fragmentation properties of quarks and gluons [11]. In general, gluon jets consist of more and thus softer particles and are less collimated than quark jets. These differences are exploited in a likelihood discriminator, which combines three sensitive variables that are defined at the level of the PF jet-constituents: their multiplicity, the minor axis of the elliptical jet shape spanned by the second moments of the $p_T$-weighted constituents’ directions, and $p_T D = \sqrt{\sum_i p_{T,i}^2 / \sum_i p_{T,i}}$, which measures how the jet $p_T$ is distributed among the constituents $i$. The impact of PU is mitigated by omitting charged particles that cannot be associated to the hard-interaction vertex.

The likelihood is built from jets in simulated QCD dijet events and parametrised as a function of $p_T$, $\eta$, and the average $p_T$-offset density $\rho$ for jets with $p_T > 30$ GeV and $|\eta| < 4.7$. A high quark-jet identification efficiency of up to 80% at a gluon-jet rejection efficiency of 60% is achieved. The discriminator output has been validated in $Z +$ jets events, where good agreement is found between data and simulation as demonstrated in Fig. 7.

7. Summary

CMS commonly uses the particle-flow approach for jet reconstruction in order to optimally benefit from the high-resolution tracking system and electromagnetic calorimeter. Pile-up energy contributions to the jets can be reduced by up to 50% already prior to the jet clustering by omitting charged particles originating from pile-up vertices.

Jet energy corrections are determined in a factorised approach. The dominant factors, which correct for the remaining impact of pile-up and the $\eta$ and $p_T$ dependence of the detector response, are derived from the simulation, which describes the data well. The obtained jet energy scale is validated in data, and small remaining differences to the simulation are converted into a residual correction factor. As a result, the overall jet energy scale at CMS is known to percent-level precision and better than 1% for jets above 100 GeV in the central detector region.

Furthermore, CMS has developed advanced tools to reject pile-up jets and to discriminate between quark- and gluon-induced jets. The procedures exploit likelihood discriminators sensitive to the characteristic substructures of the different jet types.

In summary, jets at CMS are well-understood objects allowing high-precision measurements at the energy frontier, whereby many important final states become accessible for physics analysis. Hence, jets contribute in an essential way to the exploitation of the LHC physics-potential.

References

[1] CMS Collaboration 2008 *JINST* 3 S08004
[2] CMS Collaboration 2009 CMS-PAS-PFT-09-001
[3] Cacciari M, Salam G P and Soyez G 2008 *J. High Energy Phys.* 08 063
[4] CMS Collaboration 2012 CMS-DP-12-012
[5] CMS Collaboration 2011 *JINST* 6 P11002
[6] Cacciari M and Salam G P 2008 *Phys. Lett. B* 659 119 – 126
[7] Sjöstrand T, Mrenna S and Skands P Z 2006 *J. High Energy Phys.* 05 026
[8] Agostinelli S et al. 2003 *Nucl. Instrum. Meth. A* 506 250–303
[9] CMS Collaboration 2013 CMS-DP-13-033
[10] CMS Collaboration 2013 CMS-PAS-JME-13-005
[11] CMS Collaboration 2013 CMS-PAS-JME-13-002