Available on CMS information server **CMS IN 2007/xxx**

CMS Internal Note

The content of this note is intended for CMS internal use and distribution only

October 29, 2007

Version 0.7

Plans for Jet Energy Corrections at CMS

S. Esen, G. Landsberg
Brown, Providence, RI, USA

M. Vazquez Acosta
CERN, Geneva, Switzerland

G. Dissertori
ETH, Zurich, Switzerland

D. Elvira, R. Harris, K. Kousouris
Fermilab, Batavia, IL, USA

A. Nikitenko
Imperial College, London, England

S. Petrushanko, O. Kodolova
Moscow State University, Moscow, Russia

R. Cavanaugh
U. of F, Gainesville, FL , USA

N. Varelas
U. of Illinois at Chicago, Chicago, IL , USA

A. Bhatti
U. Rockefeller, New York, NY, USA

J. Cammin, M. Zielinski
U. Rochester, Rochester, NY, USA

B. Hirosky
U. of Virginia, Charlottesville, VA , USA

J. D'Hondt, P. Van Mulders
Vrije Universiteit Brussel, Brussels, Belgium

Abstract

We present a plan for Jet Energy Corrections at CMS. Jet corrections at CMS will come initially from MC tuned on TB data, directly from collision data when available, and ultimately from a MC tuned on collision data. The corrections will be factorized into a fixed sequence of sub-corrections associated with different detector and physics effects. The following three factors are minimum requirements for most analysis: offset corrections for pile-up, noise, and effects of thresholds; correction for the response of the calorimeter as a function of jet pseudorapidity relative to the barrel; correction for the absolute response as a function of transverse momentum in the barrel. The required correction gives a jet four vector equivalent to the sum of particles in the jet cone emanating from a QCD hard collision. The following four factors will be provided for use if desired by the analysis: dependence on the fraction of jet energy in the Ecal; dependence on the flavor of the final state jet; removal of underlying event; correction back to the parton level. We discuss the status of these corrections, the planned data-driven techniques for their derivation, and their anticipated evolution with the stages of the CMS experiment.

1 Introduction

2 A Multi-Level Jet Correction

Contributor : R. Harris

We believe that CMS would benefit from a factorized multi-level jet correction. In such an approach the jet correction is decomposed into (semi) independent factors applied in a fixed sequence. The levels we currently envisage are listed below and pictured in Fig. 1:

1. **Offset**: correction for pile-up, electronic noise, and jet energy lost by thresholds.
2. **Relative** (η): correction for variations in jet response with pseudo-rapidity relative to a control region.
3. **Absolute** (p_T): correction to particle level versus jet p_T in the control region.
4. **EMF**: correction for variations in jet response with electromagnetic energy fraction.
5. **Flavor**: correction to particle level for different types of jet (light quark, b, τ, etc.)
6. **Underlying Event**: luminosity independent spectator energy in jet removed. (**FIXME: GD : it is not only spectator energy**)
7. **Parton**: correction to parton level

We caution the reader that the order of these corrections and what each includes is still being debated. (**FIXME: GD : instead of "debated" rather "studied"?**)

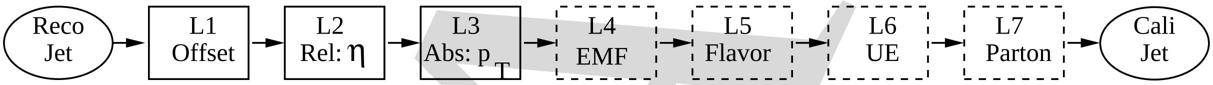

Figure 1: Schematic picture of a factorized multi-level jet correction, in which corrections to the reconstructed jet are applied in sequence to obtain the final calibrated jet. Required correction levels are shown in solid boxes and optional correction levels are shown in dashed boxes.

Here we factorize to better understand the jet energy scale and reduce the systematic uncertainty. In this approach each level can be separately determined, refined and understood. Systematic uncertainties can then be determined (semi) independently for each level, providing a better overall understanding of the origins of systematic uncertainty in the jet energy scale. The Tevatron experiments found that factorization into multiple levels was needed in order to measure the jet energy scale using in-situ collider data [2, 4]. It was easier to develop physics based methods to understand each level separately. For example, methods were developed to understand the absolute scale as a function of p_T in a single control region independent of the issue of calorimeter uniformity with η. Other methods, sometimes using different physics channels, were developed to understand the response of the calorimeter in η with respect to this control region. This allowed the Tevatron experiments to utilize large collider data samples to determine jet corrections, and reduce the systematics. Breaking the problem into it's natural components allowed measurement and better understanding of each component.

At CMS corrections for each level can be developed in progressive stages: with MC truth now, with data-driven techniques applied to MC data samples over the next year, and with actual collision data roughly a year from now. Breaking up the correction into multiple levels allows for collaborative work and continuity. In contrast to monolithic corrections, which must be replaced completely in order to make an improvement on any part, multi-level corrections can gradually evolve and improve.

Factorization facilitates prioritization, evolving corrections and scope. Addressing problems one-by-one, factorization allows us to use more people to solve the jet energy correction problem quicker. We can address the most essential corrections first. CMS must with high priority correct the jets for significant amounts of pile-up, flatten the jet response versus η, and correct the jet response as a function of p_T to the particle level. This will be done in the required correction levels numbered 1 to 3. We then provide a framework for including less pressing but needed corrections at higher levels, such as electromagnetic fraction (EMF) dependent corrections, flavor dependent corrections or model dependent corrections to the parton level. Further, these higher level corrections do not repeat the prior work of the lower levels, and therefore also do not potentially introduce different answers for these low level jet corrections. The correction levels can then evolve (semi) independently on their own natural time scales. Level 2 relative corrections versus η can be redone every year from new data samples as needed. Level 3 absolute corrections versus p_T can evolve as new data-driven techniques emerge. Scope can gradually expand to include sophisticated jet definitions. Corrections for the jet plus track algorithm, or full particle flow, can borrow the levels that apply to their jet definitions and re-evaluate the levels that have unique answers for those types of jet definition.

The multi-level approach also seeks to factor out the hard problem and move it away from the easier problems. The hardest problem, which introduces the largest systematic uncertainty, is the absolute jet energy scale as a function of p_T in the control region. We want to know the systematic error on this independent of the other corrections. The easier problems are the jet correction as a function of η relative to the control region, and the estimation of pile-up energy in the jet.

In this approach, once the correction levels are provided, the user could specify what level of corrected jet they want, but will most likely not be able to skip a prior level. Users wanting just the minimally required corrections would specify level 3, and obtain jets corrected for levels 1 through 3 back to the particle level. Those wanting jets corrected all the way to the parton level would specify level 7, and obtain jets corrected for levels 1 through 7.

3 Level 1: Offset

Contributors : O. Kodolova, R. Cavanaugh

3.1 ECAL Selective Readout and HCAL Zero Suppression

Because only a limited amount of data per event has been allocated to ECAL and to HCAL, neither the full number of ECAL channels nor the full number of HCAL channels can be recorded. The reduction of data size is accomplished via a "selective readout" (SR) in the case of ECAL and via "zero suppression" (ZS) in the case of HCAL. In the barrel, the ECAL selective readout modularity corresponds to the 5×5-crystal trigger towers.[1)] If a tower has $E_T > 5$ GeV, then the tower and all neighboring towers are readout without any zero-suppression (225 total crystals). If a tower has $E_T > 2.5$ GeV, then only that tower is readout without any zero-suppression (25 total crystals). If a tower has $E_T < 2.5$ GeV (and does not neighbor a tower with $E_T > 5$ GeV), then the crystals in that tower are readout with zero-suppression thresholds of about $3\sigma_{\rm noise}$ [5]. All HCAL cells (which correspond to a 5×5-crystal trigger tower) are readout with zero-suppression thresholds. The exact thresholding scheme, which will be employed for data-taking, is still being evaluated.

3.2 Effect on jet energy scale

The energy of a jet, measured in the calorimeter within a certain cone-size area, is the result of two additive contributions: (1) the energy of the event (sum of energy deposited from the hard interaction $E_{ij}^{\rm HI}$, and energy deposited from underlying event $E_{ij}^{\rm UE}$, energy deposited from pile-up $E_{ij}^{\rm PU}$), and (2) the contribution from the electronic noise of the detector. Hence, the mean measured energy of a jet, sampled over several events, corresponds to:

$$\langle E_{\rm jet}^{\rm meas} \rangle = \frac{1}{N_{\rm events}} \sum_{i=1}^{N_{\rm events}} \sum_{j=1}^{N_{\rm towers}} (E_{ij}^{\rm event} + E_{ij}^{\rm noise}) \quad (1)$$

where $N_{\rm events}$ is the number of events in the sample, $N_{\rm towers}$ is the number of towers belonging to a jet, and $E_{ij}^{\rm event} = E_{ij}^{\rm HI} + E_{ij}^{\rm UE} + E_{ij}^{\rm PU}$. The noise term, $E_{ij}^{\rm noise}$, is assumed to be distributed according a gaussian, with a mean value (pedestal) of 1.2 GeV, X GeV, Y GeV (**FIXME: GD : Numbers needed**) and corresponding widths of 0.125 GeV, x GeV, y GeV (**FIXME: GD : Numbers needed**), in the barrel, endcaps, and HF respectively. After the pedestal is subtracted from each cell, the mean jet energy becomes:

$$\langle E_{\rm jet}^{\rm meas,ped} \rangle = \frac{1}{N_{\rm events}} \sum_{i=1}^{N_{\rm events}} \sum_{j=1}^{N_{\rm towers}} (E_{ij}^{\rm event} + E_{ij}^{\rm noise} - \langle E_{j}^{\rm noise} \rangle) \quad (2)$$

[1)] Because of the (x,y) geometry, in the endcaps the situation is more complex.

where $\sum_{i=1}^{N_{\rm events}} \sum_{j=1}^{N_{\rm towers}} E_{ij}^{\rm noise} - \langle E_j^{\rm noise} \rangle = 0$. Hence, Eq. 2 may be rewritten so that the noise dependence is seen to be removed:

$$\langle E_{\rm jet}^{\rm meas,ped} \rangle = \frac{1}{N_{\rm events}} \sum_{i=1}^{N_{\rm events}} \sum_{j=1}^{N_{\rm towers}} E_{ij}^{\rm event} = \langle E_{\rm jet}^{\rm HI} \rangle + \langle E_{\rm jet}^{\rm PU} \rangle + \langle E_{\rm jet}^{\rm UE} \rangle \tag{3}$$

where the sum is explicitly performed over all cells within the jet area. We note that the contributions from the different sources (HI, PU, and UE) are currently assumed to be independent from each other. This is strictly true for PU effects, however not strictly true for the effects of UE, which have a dependence on the hard-interaction. If zero suppression (ZS) is applied, the energy of a read-out tower, $E_{ij}^{\rm ZS} = E_{ij}^{\rm event} + E_{ij}^{\rm noise} - \langle E_j^{\rm noise} \rangle$, must be above some threshold cut, $E_{ij}^{\rm ZS} > E_{\rm cut}$. Hence, applying ZS, one obtains:

$$\begin{aligned} \langle E_{\rm jet}^{\rm meas,ped,ZS} \rangle &= \frac{1}{N_{\rm events}} \sum_{i=1}^{N_{\rm events}} \sum_{E_{ij}^{\rm ZS} > E_{\rm cut}}^{N_{\rm towers}} E_{ij}^{\rm ZS} \\ &= \frac{1}{N_{\rm events}} \sum_{i=1}^{N_{\rm events}} \sum_{E_{ij}^{\rm ZS} > E_{\rm cut}}^{N_{\rm towers}} (E_{ij}^{\rm HI} + E_{ij}^{\rm UE} + E_{ij}^{\rm PU} + E_{ij}^{\rm noise} - \langle E_j^{\rm noise} \rangle) \end{aligned} \tag{4}$$

where the sum over towers is not performed over all cells in the jet area, but only over those cells which are above the zero suppression threshold cut. Hence, whether a cell contributes to the measured jet energy after zero suppression or not, depends on the hardness of the interaction, the effects due to the underlying event, the effects of pile-up, and noise fluctuations. In addition, the measured jet energy will also depend upon other variables such as the jet energy itself, the jet shape, and the jet type (flavour). Because one can not experimentally disentangle the effect of ZS from these other jet-algorithm dependent variables, one can not simply determine a universal contribution of PU and UE in areas free of jet activity and then subtract it from the jet energy without introducing biases. Several effects are worth noting:

- Due to the calorimeter non-linearity, the effective response of hadrons inside a (dense environment) jet and outside a jet (sparse environment) are different.
- PU and UE contributions inside and outside a jet are different and depend on the jet-algorithm
- There is no possibility to experimentally distinguish, on an event-by-event basis, the different contributions to the jet energy. Hence, PU subtraction will depend on all other effects and can not be universally determined in an independent fashion.
- Due to the high magnetic field, the loss of signal energy inside the cone is essential for low-E_T calorimeter jets

Hence, in order to correct for pile-up and underlying event (i.e. subtract it from the measured jet energy) one must first restore the energy and area of the jet. This in turn requires that the jet have no noise and no ZS threshold cuts are applied. We propose to estimate a combined PU + UE + ZS correction with real data using special events with non-zero suppression and without selective readout. The offset due to pile-up and underlying event must then be subtracted on top of the zero suppression correction.

3.3 Run Strategy without Zero Suppression and Selective Readout

It is vital to determine, from data, the effects that ZS and SR have on the calibration of reconstructed calorimeter objects, such as jets. There are several possible strategies which could be employed. For all strategies, a special pre-scaled event trigger and DAQ configuration must be defined which, when fired, reads out all ECAL and HCAL channels. Such a non-ZS trigger could be (1) randomly sampled throughout an LHC fill, (2) devoted to a single luminosity block within an LHC fill, (3) periodically devoted to an entire LHC fill. There are advantages and disadvantages of all three strategies. Strategy (1) accounts for the effect of varying luminosity conditions, but requires a complex DAQ re-configuration which is undesirable during stable physics running. Strategy (2) requires only a single DAQ re-configuration, but does not take into account the differences in occupancy due to varying luminosity (and thus pile-up) conditions inside a single fill. Hence, strategy (3) is the preferred choice. Such a dedicated "calibration" run (over entire LHC fill) would be conducted a few times per month, thereby having a minimal impact on physics, while providing a time depending track of the effect of ZS and SR on the Jet Energy Scale.

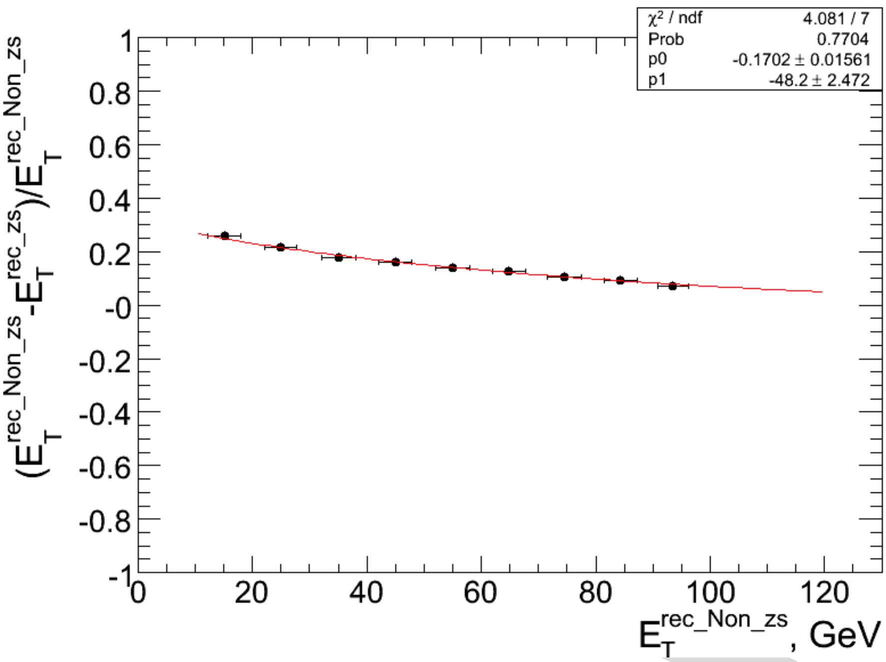

Figure 2: The relative ratio of the difference between the jet energy with and without cuts corresponding to Schema B [5] depending on the energy of jets without cuts

3.4 Corrections to Jets for zero suppression and pile-up subtraction

We propose to provide the correction function for the different algorithm types and different cut conditions based on runs without zero suppression and selective readout. The outline for a proposed procedure is, for each jet-algorithm and ZS condition over a run taken without ZS:

- Perform the nominal pedestal subtraction for all jet triggers
- Perform the jet-finding algorithm with ZS condition using data with triggered jet(s)
- Find the area that the reconstructed jet occupied in the case of no noise in the calorimeter cells and when no ZS threshold is applied to the calorimeter cells (ideal calorimeter). In general, this will depend on the jet-algorithm.
- Calculate the energy of the jet without ZS thresholds ($E_{\rm jet}^{\rm No-ZS}$) by summing the energy depositions over all cells within the jet area
- Calculate the correction ($E_{\rm jet}^{\rm ZS} - E_{\rm jet}^{\rm No-ZS}$) (so-called Type I curves designated as CorJet_1):

$$E_{\rm jet}^{\rm cor} = E_{\rm jet}^{\rm ZS} \times \mathrm{CorJet}_1(E_{\rm jet}^{\rm No-ZS}) \tag{5}$$

RC: Since this is an offset correction, why is the correction applied as a multiplicative factor? Why not use the following:

$$E_{\rm jet}^{\rm cor} = E_{\rm jet}^{\rm ZS} + \mathrm{CorJet}_1(E_{\rm jet}^{\rm No-ZS}) \tag{6}$$

A similar procedure is used in DØ, but due-to absence of the jet triggers data without zero-suppression, Monte-Carlo estimation is used in addition for zero-suppression effects [6]. The example of the Type I corrections is presented in Fig.2.

3.4.1 Cone Algorithm

The definition of the area for a cone-like algorithm is straightforward and stable. It is the cone with a radius corresponding to the radius used while the jet is found.

3.4.2 k_T Algorithm

Unlike cone-like algorithms, k_T Jets have not a well defined jet area. A few possibilities can be foreseen to define the area:

- use of the FASTJET internal procedure (**FIXME: GD: Reference needed**) to define the area. One has to check if this procedure works properly in the calorimeter environment.
- add very small energy to each of the calorimeter cells (10^{-9} GeV) assuming its infrared safety also in the calorimeter environment and refind the jet. The jet area increases twice with this procedure. The jet energy and direction in general are not changed. But, in some very rare cases, this procedure significantly changes the energy of a jet. One has to check the stability of this procedure relative to the calorimeter cells occupancy.

3.5 Corrections to Particle Flow Clusters

The essential features of the Particle Flow (PF) Clustering Algorithm [8] are described here. (Since the algorithm is slightly different, but conceptionally identical for ECAL crystals and HCAL cells, only the generic algorithm is described.) Because the particle-flow clustering algorithm is still being optimized, some of the details given in this section are subject to change.

First, noise supression thresholds are applied to all calorimeter cells. Next, simple topological clusters are determined using all cells above the noise threshold: each cell is assigned to the same topological cluster if that cell neighbors at least one other cell in that cluster. Following that, cluster seeds are identified: a cell is defined to be a seed if it has energy greater than some seed threshold and if it has energy greater than its four direct neighbors (or possibly eight direct neighbors). Such a seed cell is then designated as the starting position of a particle-flow cluster. The energy of any given cell in a topological cluster is then shared between particle-flow clusters according to a cell-to-cluster distance (assuming an expected Gaussian transverse profile). The position of every particle-flow cluster is then iteratively recalculated using a barycentre energy weighting algorithm, in which energy is re-shared between a given cell and the new updated particle-flow cluster positions until convergence is achieved for all particle-flow clusters belonging to a topological cluster.

Particle-flow ECAL clusters are formed independently from particle-flow HCAL clusters. The ECAL clusters and HCAL clusters can be linked to a given track (or linked together), depending on the spatial distance of the clusters and tracks, and thus form a "particle-flow block". For example, the block for a non-interacting single charged hadron, would be formed by one track, one ECAL cluster and one HCAL cluster. Particle-flow blocks are expected to be calibrated using isolated tracks, according to the strategy, briefly outlined below.

In the case of a block which contains only an HCAL cluster linked to a track, the procedure [9] is to form a two dimensional histogram of the reconstructed HCAL cluster energy E_{HCAL} versus the linked reconstructed track momentum p_{track} (assumed to be the true cluster energy). For each HCAL cluster energy bin, the track momentum distribution is fit with a Gaussian of mean μ and width σ (and where μ is constrained to be above zero). The Gaussian mean μ is then fit with an ad-hoc function $\mu = f(E_{\mathrm{HCAL}})$. The Gaussian width σ is also fit with an ad-hoc function $\sigma = g(E_{\mathrm{HCAL}})$. (The fit function f will also depend on the pseudo-rapidity of the HCAL cluster.) The corrected HCAL cluster energy is then $E^{\mathrm{cor}} = f(E_{\mathrm{HCAL}})$, with expected energy resolution $\sigma(E^{\mathrm{cor}}) = g(E_{\mathrm{HCAL}})$.

In the case of a block which contains both an ECAL cluster and an HCAL cluster linked to a track, the procedure [9] is similar and extended by one dimension: a three dimensional histogram is formed from the reconstructed ECAL cluster energy E_{ECAL} versus the linked reconstructed HCAL cluster energy E_{HCAL} and versus the linked reconstructed track momentum p_{track} (assumed to be the true particle-flow block energy). For each ECAL cluster energy bin and each HCAL cluster energy bin, the track momentum distribution is fit with a Gaussian of mean μ and width σ. The Gaussian mean μ is then fit with a plane $\mu = f(E_{\mathrm{ECAL}}, E_{\mathrm{HCAL}}) = a + bE_{\mathrm{ECAL}} + cE_{\mathrm{HCAL}}$. (The fit parameters b and c will depend on pseudo-rapidity and may have a residual energy dependence, due to the non-linear response of the HCAL.) The Gaussian width σ is also fit with an ad-hoc function $\sigma = g(E_{\mathrm{ECAL}}, E_{\mathrm{HCAL}})$. The corrected particle-flow block energy is then $E^{\mathrm{cor}} = a + bE_{\mathrm{ECAL}} + cE_{\mathrm{HCAL}}$, with expected enegy resolution $\sigma(E^{\mathrm{cor}}) = g(E_{\mathrm{ECAL}}, E_{\mathrm{HCAL}})$.

Finally, the expected energy deposited in the ECAL-only from a hadron track needs to be determined. The procedure [10] is to form a two dimensional histogram of the reconstructed ECAL cluster energy E_{ECAL} versus the linked reconstructed charged hadron track momentum p_{track}. For each track momentum bin, the resulting ECAL energy distribution is fit with a Gaussian of mean μ and width σ. The Gaussian mean μ is then parameterised according to the form: $\mu = f(p_{\rm track}) = a + b p_{\rm track}$. (The fit parameters a and b will depend on pseudo-rapidity.) The Gaussian width σ is also fit with an ad-hoc function $\sigma = g(p_{\rm track})$. The expected ECAL cluster energy due to a hadron is then $E^{\rm had}_{\rm ECAL} = b + a p_{\rm track}$, with an expected energy spread $\sigma(E^{\rm had}_{\rm ECAL})$.

The hadron energy spread map for ECAL $\sigma(E^{\rm had}_{\rm ECAL})$ is used to identify photons which are merged with charged hadrons. If the difference between the expected ECAL energy $E^{\rm had}_{\rm ECAL}$ and the measured ECAL energy $E_{\rm ECAL}$ is significantly higher than the expected hadron energy spread $\sigma(E^{\rm had}_{\rm ECAL})$, then a photon is created with an energy, for example, $E_\gamma = E_{\rm ECAL} - E^{\rm had}_{\rm ECAL}$. (Note that there are other possibilities for the photon energy, such as determining the energy from a multivariate analysis involving $p_{\rm track}$, $E_{\rm ECAL}$, $E_{\rm HCAL}$, cluster shapes, cluster-to-track distance, etc, and the exact algorithm estimating the photon energy is still being optimized.) Next, the resolution map $\sigma(E_{\rm ECAL}, E_{\rm HCAL})$ is used to confirm the identification of a photon in the previous step. If the difference between $(E^{\rm cor} - E_\gamma) - p_{\rm track}$ is not significantly negative, compared with $\sigma(E_{\rm ECAL}, E_{\rm HCAL})$, then the photon identificaton is confirmed; otherwise it is rejected.

Finally, the resolution map $\sigma(E_{\rm ECAL}, E_{\rm HCAL})$ is used to identify neutral hadrons which are merged with charged hadrons: if the difference between the calibrated calorimeter energy $E^{\rm cor}$ and the measured track momentum $p_{\rm track}$ significantly higher than the expected resolution of the corrected calorimeter energy resolution $\sigma(E_{\rm ECAL}, E_{\rm HCAL})$, then a neutral hadron is created with energy $E_{\rm N0} = E^{\rm cor} - p_{\rm track}$.

Because the track momentum provides an absolute scale which is neither affected by ZS nor by SR in the calorimetery, the **isolated** particle-flow cluster calibration procedure automatically includes the corrections due to any ZS or SR readout of the calorimeters.

The situation may be different in the case of **non-isolated** particle-flow clusters (as in a jet) where the effects of ZS (or possibly SR) could be reduced by the presence of nearby clusters. In such a case, the calibration maps derived from isolated particle-flow clusters from ZS data might be biased if applied to non-isolated particle-flow clusters. However, the effect of any such bias (due to the ZS readout) can be estimated (and corrected) by applying the above calibration method to isolated tracks with non-ZS data (taken from special calibration runs), and then re-performing the calibration *a posteriori* to the same data, but with an emulation of the nominal ZS. The magnitude of such an effect is still to be studied and quantified.

4 Effect of Pile-up

Pile-up events (in pp collisions) and high multiplicity events (HI collisions) deposit additional energy in the jet area. The additional energy amounts to 2.5 GeV/10 GeV/200 GeV in a cone of radius 0.5 in the barrel for low luminosity pile-up, high luminosity pile-up and heavy ion collisions, respectively.

Assuming that the correction for the ZSP in jet area is done and using runs without ZSP with min.bias trigger one can provide the estimation of the additional energy in jet. The estimation can be done on event by event basis in two different ways:

- Calculate the mean energy deposition of pile-up events depending on the pseudorapidity of the calorimeter cell for either different luminosities or as a function of the number of vertices ($N_{\rm vtx}$) in the event (Type II curves - Cor2) in runs without ZSP and with min. bias trigger [7]. This method is preferable for pp collisions.
- Calculate the mean energy deposition of pile-up events depending on the pseudorapidity of the calorimeter cell for different multiplicity ranges and create a correction curve as function of pseudorapidity and multiplicity. This method is preferable for heavy ion collisions (Type II curves - Cor2).

4.1 Corrections to Jets

In general, the correction will be applied in steps. First, a jet is found with a cut and the Type I curve is used to find the jet energy without cuts (Fig.3).

The effect of the cuts is completely compensated.

Second, the Type II correction is subtracted:

$$E^{\rm cor}_{\rm jet} = E^{\rm cut}_{\rm jet} \times {\rm CorJet}_1(E^{\rm cut}_{\rm jet}) - {\rm CorJet}_2(N_{\rm vtx}, \eta) \quad (7)$$

However, for the case of heavy ion collisions, the pile-up subtraction procedure is included already at the step of the jet finder [11]. The correction may need to be included as part of the jet finder. This procedure can also be implemented for the high luminosity conditions in pp collisions [12].

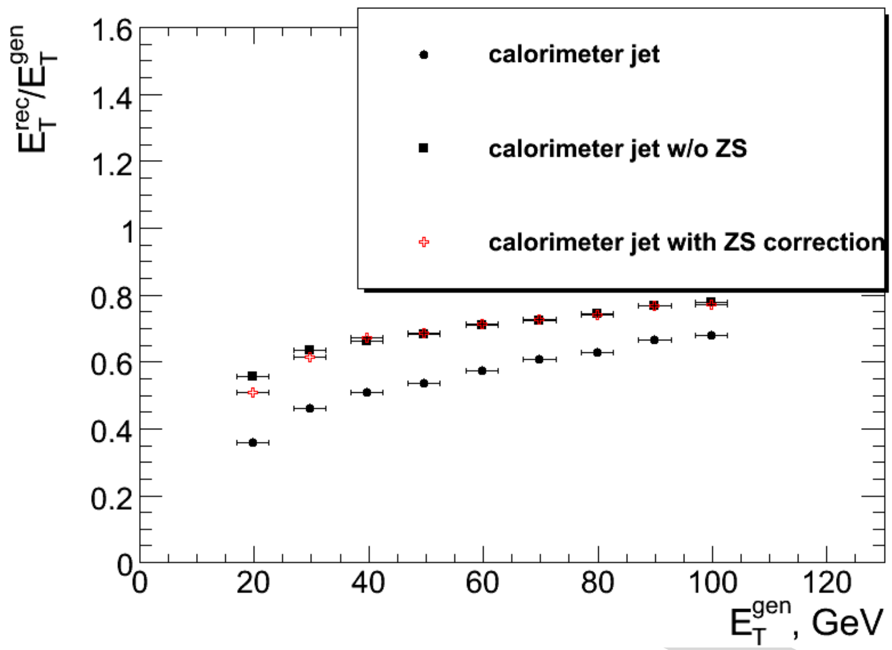

Figure 3: The relative jet response for jets found with cuts on the calorimeter readouts corresponding to Schema B (full circles), for jets found without cuts (full square) and for jets found with Schema B cuts after applying the Type I corrections of Fig.2 (cross)

This step-wise procedure can directly be applied to jets found with cone-type algrithms. For k_T - jets an intermediate step is needed in order to determine the jet area.

4.2 Corrections to Particle Flow Clusters

The particle flow algorithm will attempt to identify charged pile-up particles via a primary vertex constraint. Nevertheless pile-up subtraction may still be required in the test for, and the determination of, the neutral component of each calorimeter cluster. Such a correction will depend upon pseudo rapidity and instantaneous luminosity. Since, for low luminosity running, pile-up effects will only sparsely populate the detector, care must be exercised not to over-correct.

5 Level 2: η dependence

Contributors: A. Bhatti, D. Elvira, R. Harris, K. Kousouris

In this section we first define the scale and the correction and then we discuss how these are determined: first with MC truth and later on data-driven methods.

5.1 Definition

We define the relative jet energy scale as the η dependence of the jet energy scale with respect to the control region. The barrel is the preferred control region because it is probably the easiest to calibrate in absolute terms and it also contains the largest statistics and highest p_T reach given the $|\eta|$ dependence of the inclusive jet production cross section. A simplified defining equation for the relative jet energy scale is

$$F(\eta) = \frac{\langle \text{CaloJet } p_T \text{ / GenJet } p_T \rangle \text{ at } \eta}{\langle \text{CaloJet } p_T \text{ / GenJet } p_T \rangle \text{ control}} \quad , \tag{8}$$

which is simply the ratio of the mean scale, $\langle$ CaloJet p_T / GenJet p_T $\rangle$, at a particular value of η to the mean scale in the control region. The relative jet energy correction is then just $1/F$, the multiplicative correction that flattens the measured energy as a function of pseudo-rapidity. At a higher level of detail, the relative scale is determined after the offset correction is applied, so we must replace CaloJet p_T in eq. 8 with CorJet$_1$ p_T, where

$$\text{CorJet}_1\ p_T \;=\; \text{CaloJet}\ p_T \;-\; \text{Offset} \tag{9}$$

is a level 1 (offset) corrected CaloJet. Also, as we will see in the next section, the relative scale is weakly dependent on jet p_T at low $|\eta|$, and highly dependent on jet p_T at high $|\eta|$. We therefore re-write Eq. 8 including both the offset correction and the p_T dependence as

$$F(p_T, \eta) = \frac{\langle \text{CorJet}_1\ p_T \text{ / GenJet } p_T \rangle \text{ at } \eta}{\langle \text{CorJet}_1\ p_T \text{ / GenJet } p_T \rangle \text{ control}} \quad . \tag{10}$$

Here we have defined the relative scale independent of the method for its determination. Whether we proceed with MC truth or use a data driven method we seek to find the quantity defined by eq. 10 for the relative scale. We then have a well defined level 2 correction for the η dependence of the jet energy scale. In the case of data driven methods, we will try to define the best possible estimators for eq. 10. Very often, however, estimators are biased or correlated with other sub-corrections. While we try to define clear boundaries for all sub-corrections of the factorized JES, this is not always possible and we may have to accept these boundaries to blur in the form of overall bias corrections or common factors as a result of a finer factorization of the total JES.

5.2 MC Truth Determination of the Relative Scale

The scale is already completely defined in terms of MC truth. Determining the relative scale is straight forward once a matching between reconstructed and generated level jet objects have been clearly established:

- Produce a QCD MC sample
- Measure $\langle \text{CorJet}_1 p_T / \text{GenJet} p_T \rangle$ in fine bins of η and coarse bins of p_T.
- Divide by $\langle \text{CorJet}_1 p_T / \text{GenJet} p_T \rangle$ in the control region in the same p_T bin.
- Define a smooth spline for the relative scale as a function of η and p_T.

An example of the MC truth determination of the relative scale is shown in Fig. 4 using the control region $|\eta| < 1.3$ and the Spring07 QCD sample. This definition of $F(p_T, \eta)$ is available now **(FIXME: GD : should we specify better what it means to be "available now"?)**, and will be used to provide a correction soon.

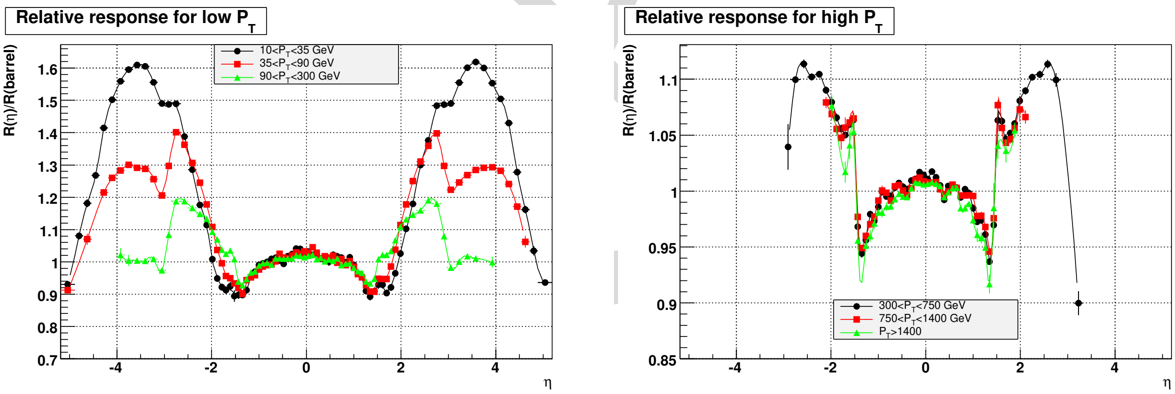

Figure 4: The relative jet energy scale determined from MC truth as a function of CaloJet η (points) compared to a smooth spline (curves). The relative scale is shown separately for three bins of low GenJet p_T (left) and three bins of high GenJet p_T (right). Iterative cone algorithm with cone size $R = 0.5$.

5.3 Data Driven Determination of the Relative Scale

Data driven methods to derive the η dependence of the jet correction will have to be investigated in CMS. Two methods were developed at the Tevatron: one based on direct dijet p_T balance (DB) and the other one on p_T balance of the full event (called Missing E_T Projection Fraction, MPF). DB is a simple approach to obtain a single number to calibrate jets reconstructed with a given algorithm back to the particle level. MPF is based on the determination of the calorimeter response to the energy cluster forming the jet by means of projecting the missing E_T into the direction of the probe jet, in case of dijet events, or photon, in case of photon+jets events. The calorimeter response measured with MPF is, to first order, independent of the jet algorithm. It is binned in terms of a well measured object (for example a photon) and mapped onto jet P_T space. An additional showering correction, for energy lost/gained through the jet boundaries during calorimeter showering has to be added as part of the η dependent correction. Data based methods are, unfortunately, affected by biases or lack of statistics in some regions of phase space. They rely on a set of cuts to ensure the necessary physics properties for the dijet system or the event to balance. The resulting corrections are not easy to transport to all samples used in physics analysis: light jets or heavy quark jets, gluon jets or quark jets, multijets events, high p_T dijet events, top events, W+jet events, etc. It is therefore critical to develop a data driven Monte Carlo simulation that models the data in such a way that it can be used to transport the JES corrections from those that are only valid for a very specific sample to the ones associated to the many exclusive samples. A third method, direct determination of the JES correction from a data driven Monte Carlo, may be the method to be used on day one when data based methods are still not well understood, as well as toward the end of the run when the Monte Carlo is well tuned to data. These three methods should be explored since they will play roles at different stages of the CMS experiment.

Based on the discussion following in the subsections below, we recommend, at the present stage of the CMS experiment, to investigate the systematics associated with dijet and event balance as well as showering in the presence of a high magnetic field as the one we have in CMS.

5.3.1 Dijet p_T Balance

Dijet p_T balance uses conservation of momentum in a $2 \to 2$ dijet process to measure the jet energy scale as a function of η relative to a control region. This technique was proposed to CMS previously [1] as a method of determining the relative jet energy scale $F(p_T, \eta)$. This method is also used by D0 [3, 4] to perform the closure tests of the full jet energy scale derivation. To briefly summarize dijet balance, the two jets with highest p_T in the event are found and for each range of dijet p_T, defined by

$$p_T^{\mathrm{dijet}} = (p_T^{\mathrm{probe}} + p_T^{\mathrm{barrel}})/2, \tag{11}$$

we divide the data into bins of probe jet η and measure the distribution of the dijet balance

$$B = (p_T^{\mathrm{probe}} - p_T^{\mathrm{barrel}})/p_T^{\mathrm{dijet}} \quad .$$

The relative jet energy scale from dijet balance is then given by

$$F_{\mathrm{DB}} = \frac{2 + \langle B \rangle}{2 - \langle B \rangle}$$

where $\langle B \rangle$ is the mean value of B. To insure we have a dijet system where balance is expected, we may employ cuts on the 3rd jet p_T and require the leading two jets be back-to-back in ϕ to suppress events with hard radiation.

The relative jet energy scale measured with dijet balance, $F_{DB}(p_T, \eta)$, should be a good estimator of the true relative scale, $F(p_T, \eta)$. Although the binning in terms of the average p_T of the two jets minimizes resolution bias effects, there is still a bias due to the different jet resolutions in the central and forward regions revealed by the steeply falling nature of the inclusive jet production cross section. Performing dijet balance in the MC and comparing to MC truth will give us a MC based estimator of this bias provided scale and resolution effects are well modeled in the MC. Showering corrections are not needed since the dijet balance η-dependent correction is algorithm dependent and accounts for showering losses of forward (probe) jets with respect to the central (control) jets. The assumption is, of course, that the central jets can be absolutely calibrated.

An example of dijet balance in the MC is shown in Fig. 5 using the control region $|\eta| < 1.3$ and the Spring07 QCD sample. Here we required the 3rd jet p_T be less than 10% of the dijet p_T and required the ϕ separation of the two jets be greater than 2.7 radians. The relative jet energy scale from dijet balance, in a bin of dijet p_T, is compared to the relative jet energy scale from MC truth. The MC truth is shown for the appropriate bin of GenJet p_T given the bin of dijet p_T and the absolute energy scale we determined in the control region (see next section). There is good agreement between the relative jet energy scale from dijet balance and MC truth, indicating that for this bin of dijet p_T the bias is small. Figure 5 also compares dijet balance with both weighted and unweighted events from the QCD sample. While the cross section weights increase the statistical uncertainties, they do not change the relative response found by dijet balance, indicating that there is little QCD spectrum dependence to dijet balance in this bin of dijet p_T.

5.3.2 Event Balance Method

Let us consider a two-body process X+jet, where X(=γ, Z or jet) is referred to as the “tag object”, and the jet is the “probe object” whose response we are interested in estimating. As we will see, the MPF method can be used

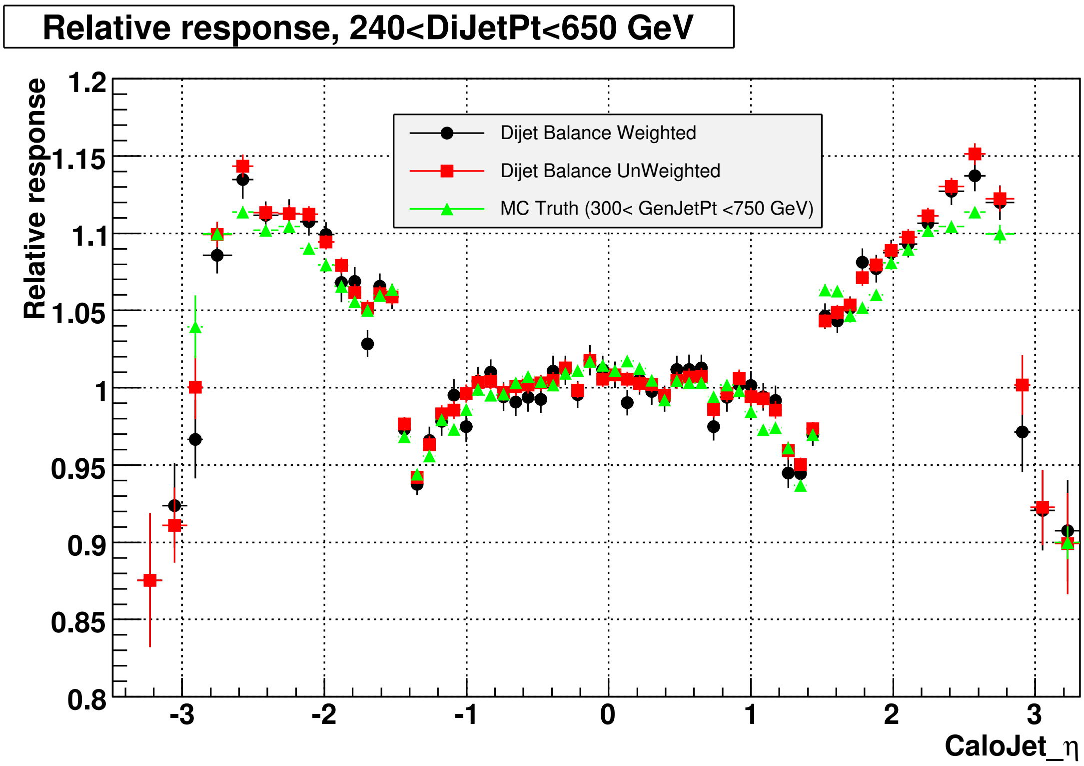

Figure 5: The relative jet energy scale determined from dijet balance $F_{\mathrm{DB}}(p_T, \eta)$ is compared to the relative jet energy scale from MC truth $F(p_T, \eta)$. Dijet balance is shown for both a true QCD jet spectrum (weighted) and for a flatter spectrum (unweighted).

to estimate the calorimeter response of the probe jet relative to the response of the tag object. This fact can be exploited to inter-calibrate the response of different calorimeter regions. In case the absolute response of the tag object is known, it will then be possible to estimate the absolute response of the probe jet. At the particle level, the transverse momenta of the tag object (p_{Ttag}) and of the hadronic recoil (p_{Trecoil}) are balanced:

$$\vec{p}_{\mathrm{Ttag}} + \vec{p}_{\mathrm{Trecoil}} = 0 \tag{12}$$

Please note that the probe jet is part of the hadronic recoil but may not constitute all of it. In a real calorimeter the response of the tag object (R_{tag}) and of the hadronic recoil (R_{recoil}) might be different (an obvious case is when the tag object is a photon), which results in a transverse momentum imbalance as measured by the calorimeter:

$$\vec{p}^{\,\mathrm{meas}}_{\mathrm{Ttag}} + \vec{p}^{\,\mathrm{meas}}_{\mathrm{Trecoil}} = \vec{\not{E}}_T \quad , \tag{13}$$

where $\vec{p}^{\,\mathrm{meas}}_{\mathrm{Ttag}} = R_{\mathrm{tag}}\, p_{\mathrm{Ttag}}$ is the measured transverse momentum of the tag object, $\vec{p}^{\,\mathrm{meas}}_{\mathrm{Trecoil}} = R_{\mathrm{recoil}}\, p_{\mathrm{Trecoil}}$ is the measured transverse momentum of the hadronic recoil, and $\vec{\not{E}}_T$ is the measured missing E_T in the event. From Eqs. 12-13 it is possible to derive the following expression:

$$\frac{R_{\mathrm{recoil}}}{R_{\mathrm{tag}}} = 1 + \frac{\vec{\not{E}}_T \cdot \vec{n}_{\mathrm{Ttag}}}{p^{\mathrm{meas}}_{T\mathrm{tag}}} \,, \tag{14}$$

which shows that the response of the hadronic recoil relative to the response of the tag object can be estimated from the projection of $\vec{\not{E}}_T$ onto the tag object direction in the transverse plane $\vec{n}_{\mathrm{Ttag}}$ and $p^{\mathrm{meas}}_{\mathrm{Ttag}}$. This quantity is often called the Missing E_T Projection Fraction (MPF). In the ideal case where the probe jet is identical to the hadronic recoil, then we can replace in Eq. 14 R_{recoil} by R_{jet}. However, among other effects, the presence of additional jets in the event (some of which might not even be reconstructed), make this idealized situation impossible to achieve in practice. By requiring exactly two reconstructed objects (tag and probe) back-to-back in azimuth, it will be possible to improve the approximation that $R_{\mathrm{jet}} = R_{\mathrm{recoil}}$. Residual effects, however, remain at the percent level which need to be corrected for. MPF is, as the DB method, affected by resolution bias effects. However, this is minimized by deriving the response correction from photon-jet events (photon resolution can be neglected with respect to that of the jet). Dijet events may be used to extend the η-dependent corrections to the highest p_T range. MPF also depends on a good understanding of $\vec{\not\!\!E}_T$ in dijet events. $\vec{\not\!\!E}_T$ is a challenging quantity to understand in day one, until spurious "hot" calorimeter regions are fixed and tools for filtering problematic events are developed and tuned. These factors, however, also seriously affect the ability to remove multijet events from the dijet sample for the direct balance method.

Finally, an algorithm-dependent showering correction is needed to compensate for the net energy flow through the jet cone boundary occurring during the shower development in the calorimeter. It is important to note that this correction does not account for parton radiation and it is a purely instrumental effect. DØ derives the showering correction from a MC template fit to the data measured jet energy profile in $\eta - \phi$ space measured from the jet centroid [4]. The template is a scalable linear combination of the contribution from underlying event, pileup, and the MC derived energy profiles from particles belonging and not belonging to the particle-jet. Showering effects are small and have little energy dependence at DØ : $\approx 1\%$ (7%) at η =0 (3) for R =0.7 cone jets. Segmentation and a strong magnetic field could make a difference in CMS.

Although more complex, the event balance and showering correction approach offers the advantage of a "physics based" JES factorization which allows to extrapolate/interpolate the corrections to regions with low statistics. In other words, it is expected (and verified at the Tevatron) for calorimeter response to have a logarithmic dependence with energy and for showering effects to be fairly flat in energy and grow with η. Physics based factorizations also allow to better understand and motivate systematic uncertainties as well as their correlations point-to-point in p_T and η. Dijet balance could then be used as a method to verify, after the full correction is applied, closure of the full correction to within its claimed uncertainties.

6 Level 3: p_T dependence

Contributors : ?

A particle jet is reconstructed from all stable particles produced in an event, using a clustering algorithm with a specific set of parameters. A particle jet may include particles produced in the hard interaction and the underlying event but will not include any contribution from additional $p\bar{p}$ interactions in the same or near bunch crossings. Experimentally a particle jet is generally obtained by correcting the calorimeter jet. Below we describe the technique to determine the particle jet energy from a calorimeter jet which has been corrected for any detector related noise and pile up contributions. The techniques described are general but will only be used in the central region, $|\eta| < 1.3$. It is expected that the absolute energy scale can be determined more precisely in this region as it has a good tracker coverage, largest P_T range, and at a given P_T it will have highest statistics. In addition, this region has maximum reach the new physics.

The calorimeter response to a jet depends on the jet fragmentation which depends on the flavor of the jet. For CMS, the absolute corrections will be derived for a flavor mixture given by either dijet or photon-jet events. In a photon-jet sample with $30 < P_T^{\gamma} < 80$ GeV, and $|\eta| < 1.3$, $\approx 90\%$ of the jets balancing the photons are either a q or a $\bar{q}$ jet where as at low P_T, dijets are mainly gluon jets. The relative contribution of quark and gluon jets changes with the transverse energy of the jets. Flavor dependent corrections must be applied if one requires a smaller jet energy scale uncertainty, provided the flavor of the jet is known.

Two different procedures have been used to correct the calorimeter jets to particle jets at the Tevatron. The CDF technique[2] relies on the detailed simulation of the single particle response and corrections are determined from the simulation and used for both real and simulated data. The DØ method decouples [3] the energy scale of real and simulated data. The energy scale corrections for these two types of data are determined independently by the same technique using photon-jet events. In the former case, an accurately simulated detector response is the primary source of calibration, whereas in the latter case the simulation is used only to evaluate any bias in the procedure.

Although a data-driven photon-jet balancing technique may require less effort, an accurate simulation of the detector is essential for the physics analyses, especially the ones relying on a precise simulation of (rare) signal events which have to be separated from a large background. Thus every effort should be made to ensure that the detector response can be accurately simulated. Moreover, it is less difficult to determine JES corrections for different algorithms after simulated data accurately represent the real data.

At CMS, the use of these techniques will evolve over time. At the start of data taking, we will used the Monte Carlo based corrections derived from simulated dijet data. As the photon-jet data is accumulated and understood, the photon-jet based corrections will replace the Monte Carlo corrections if the overall uncertainty on the jet energy scale is smaller. At some later stage, after the detector simulation has been tuned to pp collision data, the simulated data can be used to improve upon or replace the data based corrections.

6.1 γ-Jet and Z-jet Balancing

In the balancing methods, the jet energy scale is obtained by comparing the hadronic event with a photon measured in the EM calorimeter or a Z boson reconstructed from either electrons or muons. The photon-jet events have larger statistics but suffer from a large dijet background, where one of the jet satisfies all the photon identification requirements. The contamination depends on the details of photon identification requirements. The Z-jet sample is less contaminated but has smaller statistics. Actual differences in the statistics will depend on the pre-scale used during data taking. Below we discuss the photon-jet samples.

Two variations of the photon-jet balancing technique commonly used are described below.

6.1.1 Transverse Momentum Balancing

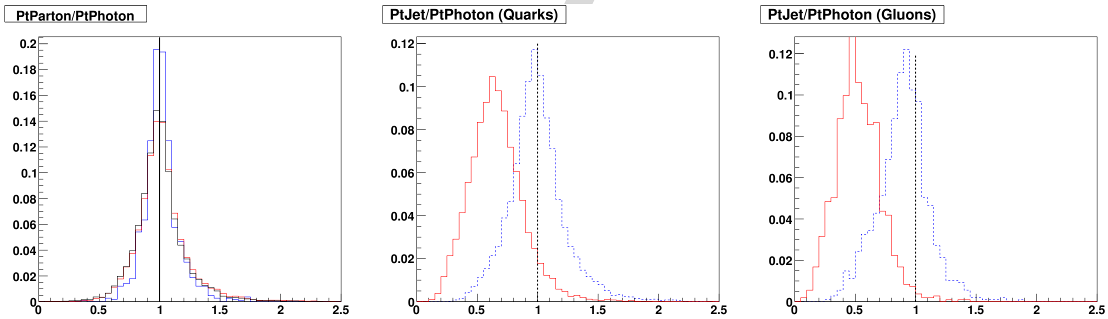

Figure 6: Ratio of the parton P_T (left), calorimeter jet and particle jet P_T to the photon P_T. In the figure on the left, the three curves correspond to gluons, quarks and all partons. The middle and right figures are for calorimeter and particle jets, quark-initiated (middle) and gluon-initiated (right). The calorimeter jets energy will be scaled such that the peak of the $P_T^{\mathrm{CaloJet}}/P_T^{\gamma}$ distribution lies at 1.0.

In the absence of any radiation, the ${P_T}^{\gamma}= P_T^{\mathrm{parton}}$ relation can be used to determine P_T^{parton}, provided P_T^{γ} is accurately known. In practice, because of initial state radiation, the mean of the $P_T^{\mathrm{parton}}/P_T^{\gamma}$ distribution may not be 1.0. Dijet events, where one jet mimics a photon, may modify the distribution. The exact contribution of dijet events depends on the photon identification cuts and the P_T of the photons. The $P_T^{\mathrm{parton}}/P_T^{\gamma}$ distributions for $50 < P_T^{\gamma} < 80$ GeV for $q\gamma$ (red) , $g\gamma$(blue) and all events (black) are shown in Fig. 6(left). From this figure, we see that for the events at the peak the parton energy is equal to the photon energy. Thus the jet energy scale at parton level can be determined by positioning the peak of the observed response ($P_T^{\mathrm{CaloJet}}/P_T^{\gamma}$) at 1.0 [14].

For analyses like resonance mass measurements the parton energy is of primary importance and is directly determined by this procedure. However, some analyses, e.g. the jet cross section measurement, require the particle jet information. The difference in the parent parton energy and the particle jet is modeled by showering/hadronization models such as PYTHIA. The ratio of particle (calorimeter) jet energy to the photon energy for quark jets is shown in solid (dashed) histograms in Fig. 6(middle). The mean of the ratio $P_T^{\mathrm{particle}}/P_T^{\gamma} = 0.996 \pm 0.002$ and the mean of the ratio $P_T^{\mathrm{CaloJet}}/P_T^{\gamma} = 0.660 \pm 0.002$. The same distributions for gluon jets are shown in Fig. 6(right) and the mean values are 0.896 ± 0.005 and 0.527 ± 0.004. Thus a gluon initiated particle jet with cone size $R = 0.5$ contains only 89.6% of the gluon energy compared to 99.6% for a quark initiated particle jet. The calorimeter response to a gluon initiated particle jet is also $\sim$ 10% lower than the calorimeter response to a quark initiated particle jet. In this scheme, the particle jet energy will be derived using generator-level Monte Carlo information. PYTHIA describes the energy profile of the jet and hence the leakage outside the cone at $\sqrt{s} = 1.96$ TeV reasonably well [15], but this has to be verified at $\sqrt{s} = 14$ TeV.

6.1.2 Missing Transverse Momentum Fraction (*MPF*) Balancing

In an event, the transverse momentum of the photon ($\vec{p}_{T\,\gamma}$) should balance the transverse momentum of the recoiling system ($\vec{p}_{T\,\mathrm{recoil}}$). The recoil system includes both the energy clustered into a jet, the energy from the hard interaction which is not included in the cluster and the energy of particles produced in the underlying event.

$$\vec{p}_{T\,\gamma} + \vec{p}_{T\,\mathrm{recoil}} = 0. \tag{15}$$

(**FIXME: GD : the following can be removed, since already defined earlier**). The difference in the calorimeter response to photon and recoiling system leads to imbalance in transverse energy, E_T^{meas}

$$\vec{p}_{T\,\gamma}^{\,1,meas} + \vec{p}_{T\,recoil}^{\,meas} = -\vec{E}_T^{meas} \tag{16}$$

Assuming the response of calorimeter is given by $\vec{p}_{T\,\gamma}^{\,meas} = R_\gamma \vec{p_T}_\gamma$ and $\vec{p}_{T\,recoil}^{\,meas} = R_{recoil}\vec{p}_{T\,recoil}$, and using Eqn. 15, the calorimeter response is given by

$$\frac{R_{recoil}}{R_\gamma} = 1 + \frac{\vec{E}_T^{meas} \cdot \hat{n_\gamma}}{p_{T\,\gamma}^{meas}} \tag{17}$$

where $\hat{n_\gamma}$ is the unit vector along photon direction.

In an EM calorimeter well calibrated for photons we expect $R_\gamma = 1.0$. Thus, the jet response R_{jet} is the same as R_{recoil} in cases where the probe particle jet is identical to the hadronic recoil. This condition can be approximately achieved by requiring one and only one jet in the event and and requiring that this jet is in the opposite direction to the photon. Even after these requirements, small biases may remain. These biases can be determined by performing the same analysis in the simulated data and comparing the results with the expectations based on generator-level information. The DØ collaboration has achieved an accuracy of $< 2\%$ on the particle jet energy scale.

6.1.3 Comparison of the two methods

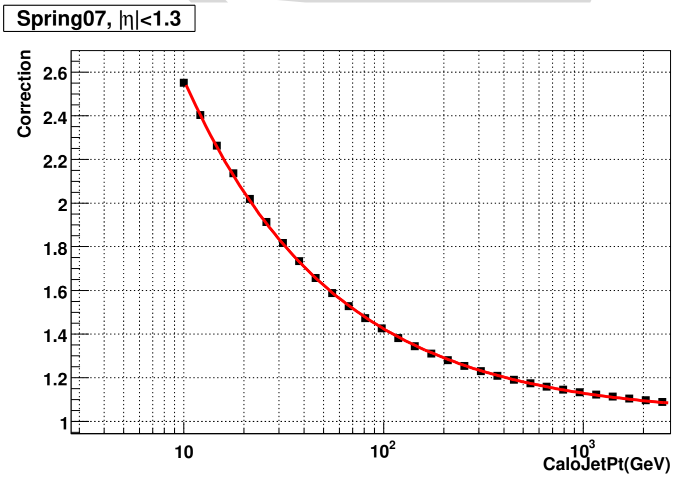

Figure 7: *This figure is a place holder for Jet energy scale correction factor as determined from photon-jet data.* Correction factors derived for P_T balanceing and MPF methods.

By P_T balancing, one determines the parton transverse momentum, whereas with the MPF technique the particle jet energy is determined. The MPF technique depends on the accurate measurement of the missing transverse energy which may not be available at the start of the run. In any case, we plan to use both methods and compare the results. A comparison of the corrections using Spring07 samples is shown in Fig. 7 as a function of P_T^γ. For simplicity we have used a pure photon-jet sample. Any response determined using photons in the data would require an additional correction factor to account for the effects of fake photons.

6.2 Monte Carlo based Energy Corrections

It is highly desirable to have a detector and physics simulation which reproduces the collider data very accurately. A good simulation simplifies the data analysis. CMS has spent a considerable amount of effort in measuring the detector response in the test beam environment and implementing the measured response in detector simulation. After a good simulation is available, the jet corrections can be derived for specific physics samples for jets reconstructed with different clustering algorithms and parameters.

Current JES corrections are derived from simulated di-jet data. A particle jet is matched to nearest calorimeter jet in $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Particle jets where the nearest calorimeter jet has $\Delta R > 0.3$ are not used. The ratio $P_T^{\rm CaloJet}/P_T^{\rm ParticleJet}$ defines the calorimeter response to a jet. For the particle jets with fixed P_T, the most probable value of the response is determined by iteratively fitting the ratio distribution by a Gaussian in a narrow range around the maximum. The resulting response is parameterized as a function of calorimeter jet P_T[2] and is used to scale the calorimeter jet 4-vector. The correction factors for CMSSW 1_3_1 as a function of calorimeter jet P_T for the Iterative cone algorithms with R=0.5 is shown in Fig. 8.

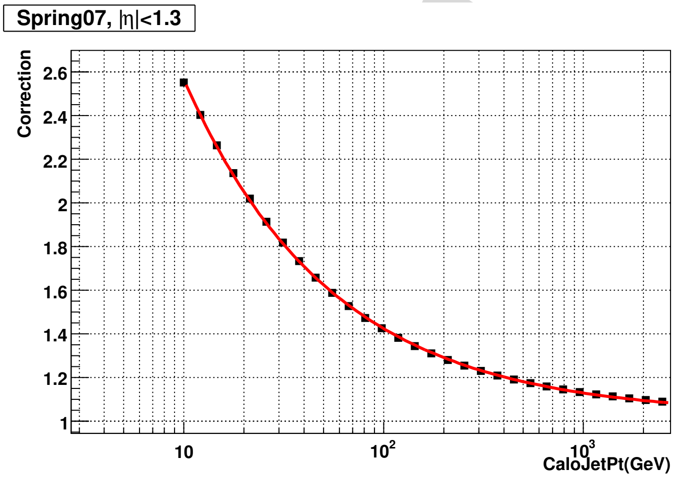

Figure 8: Monte Carlo based jet corrections as a function of Calorimeter jet P_T

The current calorimeter response simulated with GEANT4 describes the CMS test beam data reasonably well, but shows a small discrepancy for low momentum charged particles. In the actual running conditions, the calorimeter response to the single particles will be different from what is observed in test beam data due to the presence of the magnetic field and the extra material in front of the calorimeter. These differences must understood and modeled to allow for an accurate determination of the jet energy scale using Monte Carlo based methods.

7 Level 4: EMF dependence

Contributors: S. Esen, B. Hirosky, G. Landsberg

In this section we propose a jet electromagnetic energy fraction (EMF) dependent correction to be applied in addition to the above uniformity and absolute response corrections. We outline the steps necessary to determine this correction using both MC truth and data when sufficient samples are available.

7.1 Motivation

Non-uniformities in detector elements and intrinsic non-linearities in their response for particles of different energies necessitate the development of corrections for the observed response of the detector to jets as a function of p_T and $|\eta|$. These methods are described in detail in the previous sections. However, applying corrections as a function of only these variables integrates over all other characteristics of the jets, possibly limiting the overall resolution for measuring the energies of individual jets.

Consider a model that describes the response of a detector for jets as a linear combination of the response for various particle types as illustrated by eq. 18, where $R_{\frac{e}{h}}$ represents the ratio of the response of the detector for electrons and hadrons. In a non-compensating calorimeter, we expect $R_{\frac{e}{h}}$ to depend on energy, $R_{\frac{e}{h}}(E)$. For simplicity, we can take $EM\,particles$ to be mainly π^0's, $HAD\,particles$ to be mainly $\pi^\pm$'s, and $invisible\,particles$

[2] The previous procedure where the correction was determined by inverting the response versus particle jet P_T while applying the corrections has been updated.

to represent anything that deposits little or no energy in the calorimeter (e.g. μ, ν, etc.). Equation 18 also implicitly describes a detector with an ideal response for $EM\,particles$. It is then evident that the response of the detector for jets is dominated by the $R_{\frac{e}{h}}$ ratio and the fraction of $invisible\,particles$. Fortunately this latter category of particles is not expected to make up a large fraction of the energy of a typical jet, thus highlighting the importance of $R_{\frac{e}{h}}$.In particular if $R_{\frac{e}{h}}$ is greatly different from unity, the response will be very sensitive to fluctuations in the hadronization of a jet favoring the creation of more or fewer energetic π^0's.

$$\langle R_{\mathrm{jet}} \rangle = \frac{\displaystyle\sum_{i\in\, EM\,ptcls} E_i + \sum_{j\in\, HAD\,ptcls} \frac{E_j}{R_{\frac{e}{h}}(E_j)}}{\displaystyle\sum_{i\in\, EM\,ptcls} E_i + \sum_{j\in\, HAD\,ptcls} E_j + \sum_{k\in\, \text{``}invisible\,ptcls''} E_k} \tag{18}$$

While we do not have direct access to this information on a jet-by-jet basis, we can easily measure the jet EMF. As illustrated in eq. 19, the EMF is strongly correlated with the $R_{\frac{e}{h}}$ as well as the amount of energy carried by EM particles.

$$\langle EMF_{\mathrm{jet}} \rangle = \frac{\displaystyle\sum_{i\in\, EM\,ptcls} E_i + \sum_{j\in\, HAD\,ptcls} \frac{E_j}{R_{\frac{e}{h}}(E_j)} \cdot EMF_j(E_j)}{\displaystyle\sum_{i\in\, EM\,ptcls} E_i + \sum_{j\in\, HAD\,ptcls} \frac{E_j}{R_{\frac{e}{h}}(E_j)}} \tag{19}$$

EMF_j represents the fraction of energy deposited by a hadron in the EM calorimetry. In addition to energy, $R_{\frac{e}{h}}$ and EMF_j will, in general, depend on details of the detector elements and the incident angles of particles, since these can modify the sampling of showers in the calorimetry. See for example Fig. **??** showing the average EMF versus η, the effects of crossing detector boundaries are clearly visible.

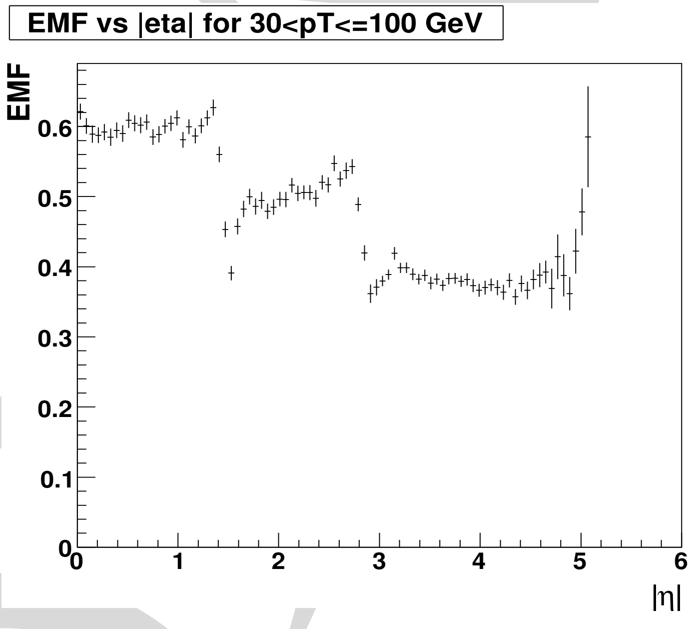

Figure 9: EM fraction as a function of $|\eta|$.

It has been shown that including an EMF-dependent correction in three parameters (p_T,η,EMF), a 3D-correction, can improve the observed jet resolution [16] in CMS. In [16] a 3D-correction was applied to jets produced in MC and the observed resolution for p_T measurements was studied. Figure 10 illustrates the observed change in response as a function of EMF, integrating over p_T and η. Using CMSSW_1_2_0 QCD jet samples, an EMF-dependent 3D correction was derived up to 300 GeV RecJet p_T within the $|\eta| < 2.5$ region. The correction is defined as the difference between GenJet p_T and CaloJet p_T, for a given CaloJet p_T, $|\eta|$, and EMF bin, C(p_T,$|\eta|$,EMF):

$$C(p_T, |\eta|, EMF) = \langle \text{GenJet } p_T \text{ - CaloJet } p_T \rangle \tag{20}$$

After application of the EMF-dependent 3D-correction, jet resolutions are compared (Fig. 11) to those for uncorrected jets and jets corrected with separate 1D corrections for $|\eta|$ and p_T. For both central and forward jets an improvement of order 5-10% is observed in the resolution. We describe plans for the derivation and inclusion of this correction below.

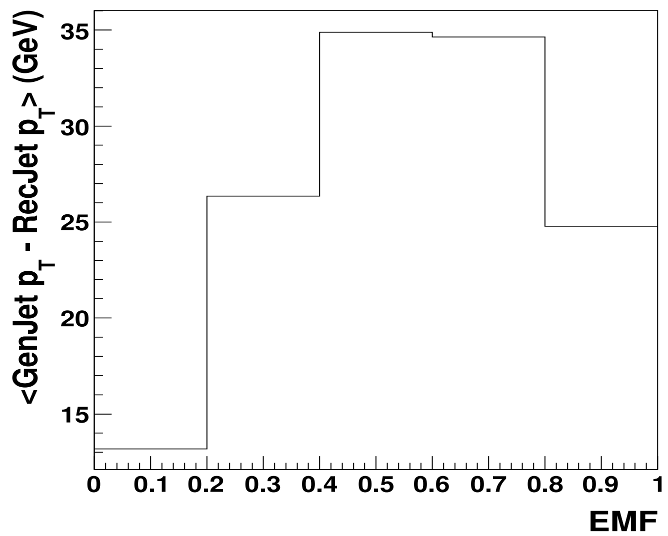

Figure 10: Correction, GenJet p_T - CaloJet p_T, as a function of EMF.

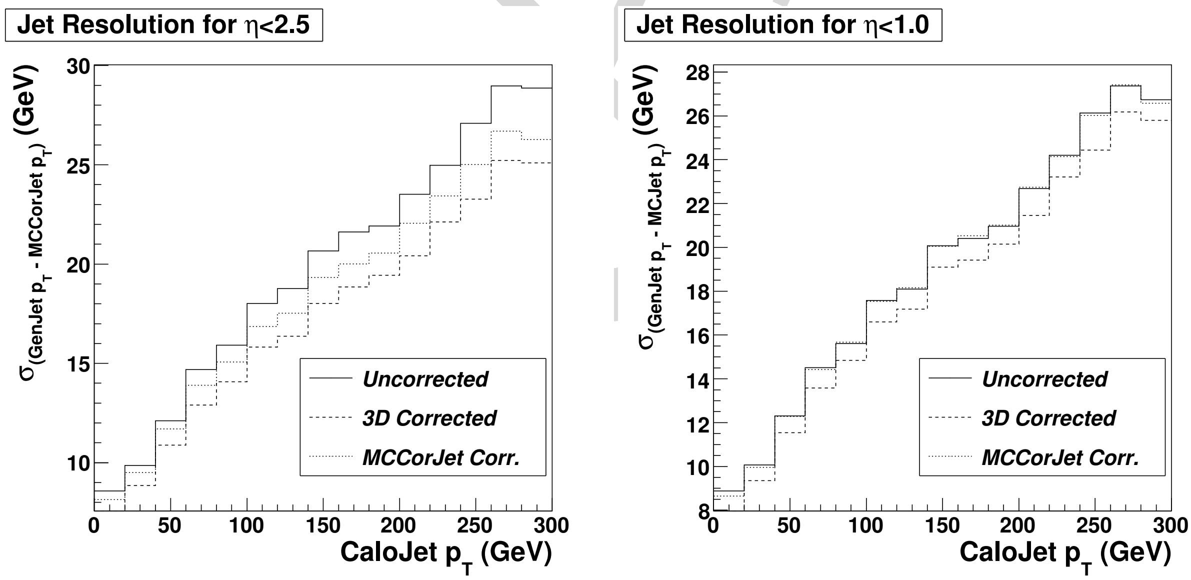

Figure 11: Jet resolution after C(p_T,η, EMF). Left: $|\eta| < 2.5$ Right: $|\eta| < 1.0$

7.2 JES EMF Correction

After applying the standard MCJet corrections [17] a p_T and $|\eta|$ dependence remains for the application of the EMF correction. This is most easily illustrated by integrating over one of the variables:

$$C(\text{EMF}, p_T) = \langle \int d|\eta|(\text{GenJet}p_T - \text{MCCorJet}p_T)(\text{EMF}, |\eta|)\rangle \quad (21)$$

$$C(\text{EMF}, \eta) = \langle \int dp_T(\text{GenJet}p_T - \text{MCCorJet}p_T)(\text{EMF}, p_T)\rangle \quad (22)$$

where MCCorJet p_T is the corrected p_T which comes from the standart MCJet corrections. Here "corrected" refers to $|\eta|$ and p_T corrections.

Figure 12 (left) shows C(EMF) for various fixed p_T bins integrating over the entire $|\eta|$ range. A significant p_T dependence is evident. The right plot in Fig. 12 shows C(EMF) for $0. < |\eta| < 1.0$ (HB), $1.8 < |\eta| < 2.5$ (HE) and $3.5 < |\eta| < 4.5$ (HF), and demonstrates that there is significant $|\eta|$ dependence in C(EMF) after MCCorJet corrections, that decreases as we go forward; more in HB than in HE, more in HE than in HF. p_T and $|\eta|$ dependence must therefore be included in the EMF-based corrections to maximize their the performance.

We propose to apply an EMF-dependent correction after offset (L1), relative (L2) and absolute (L3) corrections. The plan is to derive this correction as a function of p_T, $|\eta|$ and EMF for HB, HE and HF separately. The definition of the EMF dependent correction is the difference between GenJet p_T and CorJet p_T at a given reconstructed jet CorJet p_T, $|\eta|$ and EMF bin, S(p_T,$|\eta|$,EMF):

$$\text{S}(p_T,|\eta|,\text{EMF}) = \langle \text{GenJet } p_T \text{ - CorJet } p_T \rangle \quad (23)$$

where CorJet is a L3 (absolute) corrected CaloJet, and p_T, $|\eta|$, EMF are L3 jet variables.[3] We will have 3 equations for HB, HE and HF separately and will then use smooth interpolations between them to provide a single correction.

7.2.1 MC Truth Determination

Determining the EMF scale by comparison with MC-truth is straight forward:

- Produce a QCD MC sample
- Measure $\langle$GenJet p_T - CorJet $p_T\rangle$ as a function of CorJet p_T in each of η and EMF bins for HB, HE and HF.
- Find a smooth fit to as a function of p_T, $|\eta|$ and EMF for each of the three subdetectors.
- Interpolate them smoothly to combine in a single correction

In practice the samples used must be consistent with those used to derive the Levels 1-3 corrections otherwise the closure test results may be adversely affected.

7.2.2 Determination of the Correction in Data

In data the correction can be derived from the same γ-jet samples used to determine the absolute jet energy scale, though it will require significantly larger data samples than will be necessary for early passes at the data-derived absolute scale, one or two orders of magnitude larger at the very least. This will not cause a great difficulties in the long run, since enormous data sets are expected and most corrections will quickly become systematics limited. It may also be possible to paramaterize the EMF scale dependence using dijet data, where data samples will be significantly larger. Possible biases in this approach should be determined in the MC by comparing with both γ-jet and EM-enriched di-jet samples. Another alternative in early running is the possible application of an MC-derived EMF correction on top of data-driven corrections at Levels 1-3. This would require quatitively excellent agreement between data and the Monte Carlo for jet fragmentation and detector showering. While the MC should be used to study the EMF correction and its systematics, direct application of these corrections to the data may prove difficult.

[3] η and EMF do not change as the result of jet corrections. So they are the same for uncorrected and L3 corrected jets.

7.3 Discussion

The EMF correction has been shown to improve jet resolution at up to 10% levels. The EMF correction will be derived as an additional factor to be applied after the absolute response. In this way the EMF may be turned on or off to study its affect on resolutions and closure at Level 3. We note that deriving benefits from the EMF correction requires sufficiently fine binning of the data to prevent "integrating away" too much information. Also biases in the EMF measurement (especially as a result of minbias pileup and zero suppression effects/corrections) must be understood.

Sufficient sample sizes are only a matter of time. Part of the Monte Carlo efforts to study the JES methods, should include an estimation of how much γ+jet data are required to derive an adequate EMF correction. The average values of jet EMF's may be affected by pileup at high luminosities (possibly related to zero suppression thresholds), by choices in techniques to subtract pileup energy (e.g. will different energies ultimately be subtracted from EM and HAD portions of the calorimeter?), etc. As part of the offset studies, we should seek methods that remove as much luminosity dependence as possible from the EMF determinations. Finally, the EMF correction must either be constructed to preserve η-uniformity either by applying an explicit residual correction factor or by developing a two-step algorithm, which eliminates any residual η-dependent effects during the last step of the EMF correction determination.

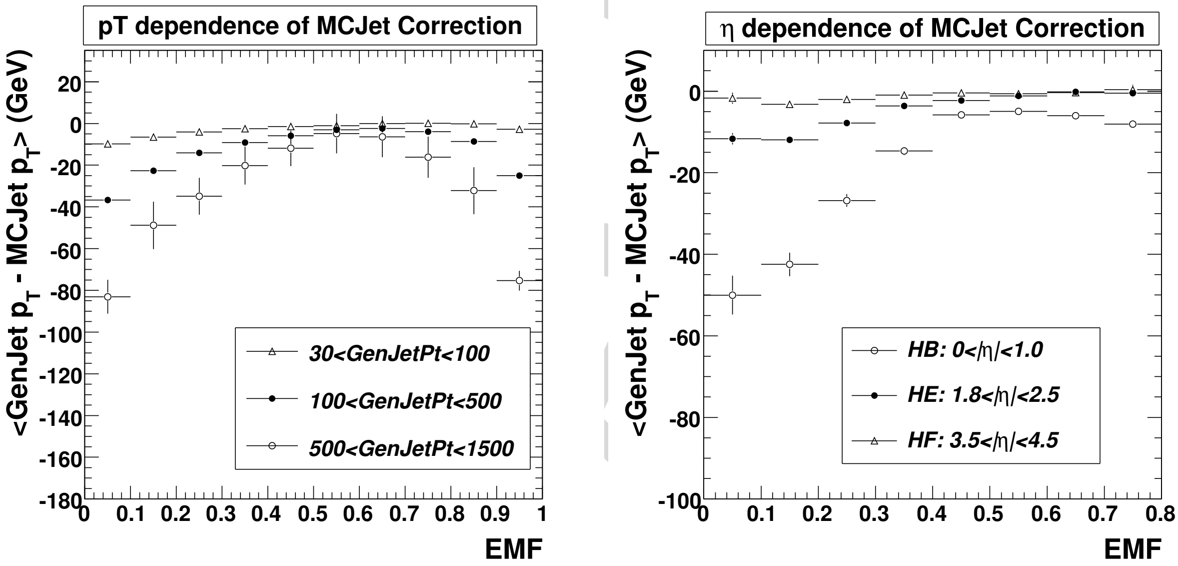

Figure 12: Left: p_T dependence of MCCorJet correction. Right: $|\eta|$ dependence of MCCorJet correction.

8 Level 5: Optional Flavor dependence

Contributors: J. Cammin, A. Nikitenko

In this section we discuss the flavor dependence of the jet energy corrections with a focus on b-jet specific corrections.

Jets that originate from heavy quarks (b and c quarks) differ from light quark jets in several aspects. The average number of charged hadrons is higher for b jets than for light jets, but the average charged hadron momentum is smaller. B hadrons also have a large branching fraction into semileptonic decays. These decays can occur directly ($b \to \ell\nu + X, \sim 11\%$) or via cascade decays ($b \to c \to \ell\nu + X, \sim 11\%$, here also the charmed hadron can decay semileptonically), and the energy of the neutrino(s) cannot be measured in the detector. The combined effects lead to a charged hadron energy fraction that is considerable smaller than that of light quarks jets. Hence the average calorimeter energy response of b jets is lower than that of light jets.

The size of the effect can be seen in Fig. 13, where the ratio of reconstructed to particle-level transverse jet momentum is shown for different jet flavors in dijet and $\mathrm{t\bar{t}}$ events as a function of GenJet p_T. The response for

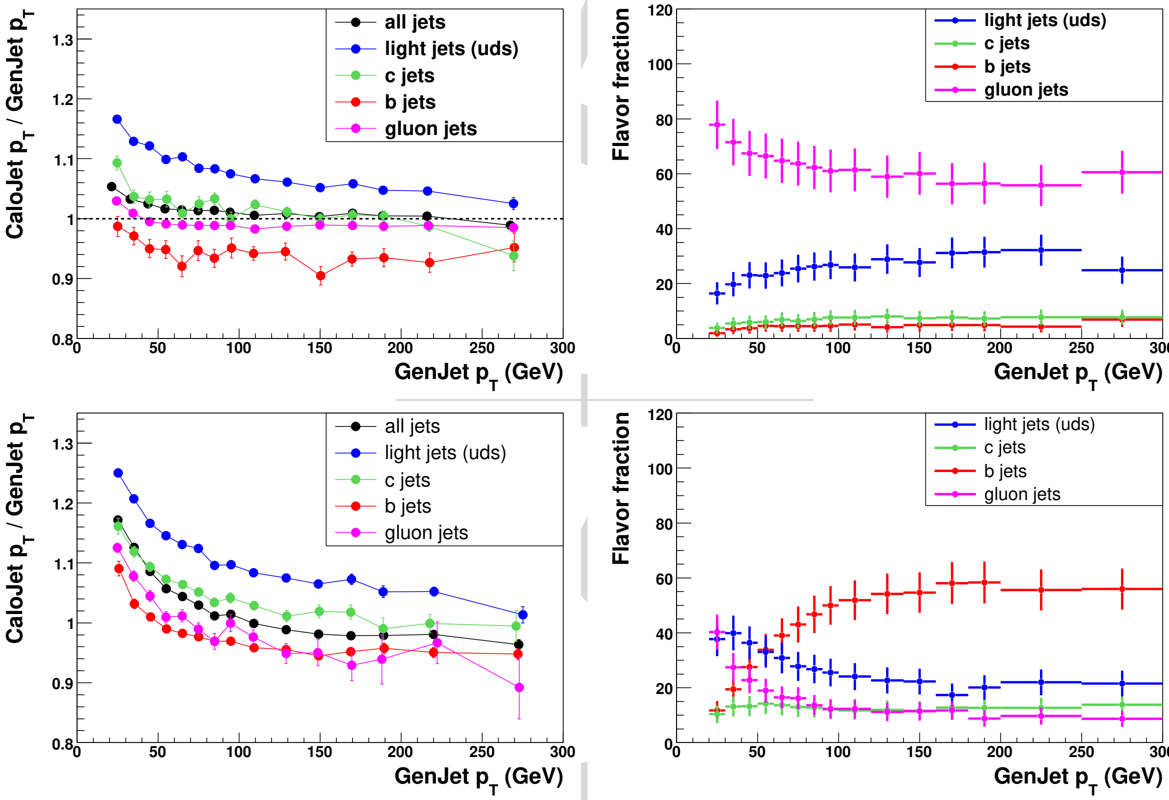

Figure 13: Left: Ratios of reconstructed to particle-level transverse jet momentum for different jet flavors after MCJet corrections (top: dijets events, bottom: inclusive $t\bar{t}$ events). Right: Fractional contribution of jet flavors to the average jet in dijet events (top) and $t\bar{t}$ (bottom). (**FIXME: need to update plots with higher statistics.**)

b-jets is about 10 % smaller compared to uds jets.[4)]

Inclusive b decays consist of both hadronic and semileptonic decays. The energy response for these two classes of decays is quite different since there is at least one neutrino in semileptonic decays whose energy cannot be measured, resulting in a smaller response than that of purely hadronic b decays. As special algorithms exist to identify muons and electrons in semileptonic b decays ("soft lepton taggers") it makes sense to derive corrections separately for inclusive b jets and for semileptonic b jets. Inclusive corrections would then be applied to jets which are considered b jets by a lifetime b-tagging algorithm and semileptonic corrections to jets with a soft lepton tag. The separation of these two rather distinct cases can improve the overall energy resolution of b jets, especially at low transverse momentum.

Several procedures should be explored to determine the residual b jet corrections, after the inclusive (light) jet energy scale corrections have been applied. In all cases it is necessary to cross check the corrections for different b-tagging algorithms, since different taggers select different subsets of b jets, which can have different energy responses.

8.1 b jet corrections from Monte Carlo

Before data taking, or when there is only a small data sample of b jets available for calibration, b jet corrections can be derived from Monte Carlo as the ratio of $\langle \mathrm{CorJet}p_T/\mathrm{GenJet}p_T \rangle$ in bins of jet p_T and η for b-tagged jets. In the case of semileptonic corrections the corrections might also be parameterized in quantities of the soft lepton, for example as a function of p_T^{rel}, the transverse momentum of the lepton with respect to the jet axis.

8.2 b jet corrections from $t\bar{t}$ events

A large number of top-quark pairs will be produced at the LHC and can be used for data-driven b jet corrections. After reconstruction of the lepton+jets final states the invariant three jet mass m_{jjb} can be constrained to the world average of the top quark mass measured at Fermilab's Tevatron experiments. If this is done after applying the light jet corrections to all jets then the only parameter that needs to be adjusted is the residual scale factor between light jets and b jets. The light jet energy scale can be cross-checked by calculating the dijet invariant mass of the two jets assigned to the W decay, which should be close to m_W.

8.3 b jet corrections from $\gamma + b,\ b\bar{b}Z,\ Z \to b\bar{b}$ events

With increasing data statistics it will also be possible to determine the b jet energy scale from momentum balance in $\gamma + b$ or $b\bar{b}Z$ using either the MPF method (see section 5) or simple p_T balancing. The position of the mass peak in $Z \to b\bar{b}$ can serve as a cross check for the b jet energy scale, but because of huge backgrounds to the $Z \to b\bar{b}$ signal it might be too difficult to actually extract corrections from this process.

With γ+jets events the b response as a function of p_T can be derived with the same methods as the inclusive absolute response described in section 6, but applied to b-tagged events. Some analyses will need the pure b response and so the b response needs to be disentangled from the "tagged" response. This requires knowledge of the fractions of light, b, and c jets in b-tagged jets.

The process $pp \to b\bar{b}Z/\gamma^*, Z/\gamma^* \to \ell\ell$ has been shown to be useful for providing energy corrections for b jets [18]. The p_T balance between the Z boson and pair of b jets is exploited. About 1000 events, useful for calibration, are expected with 10 fb^{-1}, of these $\simeq$ 25% are due to background contamination from Drell-Yan and $t\bar{t}$. No effect of background contamination on the energy corrections was found, however a more dedicated study requiring larger background Monte Carlo statistics is required. It has been shown that with 10 fb^{-1} it will be possible to evaluate the energy corrections for b jets as a function of raw energy in the interval between 20 and 150 GeV and $|\eta| <$2.5 with an uncertainty between $\simeq$ 10 % (at 20 GeV) and $\simeq$ 4 % (at 150 GeV). However, the binning in η in addition to E_T requires more data. The b-jet energy corrections restore the energy of the original b quark. They can be directly applied to the b jets in the associated production of the Higgs bosons in MSSM $pp \to b\bar{b}\phi,\ \phi \to \tau\tau\ (\mu\mu)$ or for the Higgs boson mass reconstruction in the $pp \to t\bar{t}H,\ H \to b\bar{b}$ process. For the latter process, however the impact of the difference in the color reconnections in the $pp \to b\bar{b}Z,\ Z \to \ell\ell$ and $pp \to t\bar{t}H,\ H \to b\bar{b}$ processes has to be studied and taken into account.

[4)] The ratio of $\mathrm{CaloJet}p_T/\mathrm{GenJet}p_T$ for dijet events has been derived in a slightly different way than the default MCJet corrections, and hence Fig. 13 should not be understood as a closure plot. What matters here is the relative difference between jet flavors.

8.4 Biases from flavor dependent corrections and event topologies

Figure 13 demonstrates that the jet energy response has a strong dependence on the jet flavor. The average flavor composition of jets also depends on the process and so it cannot be expected that corrections derived from one sample provide the correct scale for jets in another sample. As an example, it can be seen from the right-hand plots in Fig. 13 that the fraction of light, gluon, and b jets is very different for dijet events and $\mathrm{t\bar{t}}$ events, which leads to differences in the average energy response for dijet and $\mathrm{t\bar{t}}$ events. Additional process dependent corrections will therefore be necessary.

For b jets an additional bias can occur from the choice of a particular b-tagging algorithm and its operating point, since they represent different efficiencies and mistag rates. A specific b-tagging algorithm selects a biased subsample of b jets and the energy response of those subsets can vary significantly between b-tagging algorithms. It will therefore be necessary to cross-check the corrections on different b-taggers or to derive b-tagging algorithm-specific corrections.

9 Level 6: Optional UE correction

Contributor : O. Kodolova

The underlying event depends on the details of the hard interaction. Thus, it is not possible to introduce a unique correction for the underlying event energy falling into the jet area. In addition, many Monte Carlo generators have included the underlying event based on different models. The recommendation is to make this correction optional and to provide a tool to estimate the underlying event contribution. We propose to use the schema (1) described in subsection 3.3 in addition to the recommended schema (3) for the runs without zero-suppression. This schema will only have HCAL readouts without zero-suppression, but selective readout for ECAL cells.

Events with one vertex have to be selected. The estimation of the energy has to occur in the area with the jet square (**FIXME: GD: ??**) in the direction perpendicular to the direction of the jet [13]. In order to estimate the effect of zero-suppression, the UE energy estimation in events with jet triggers in schema (1) have to be compared with the UE energy estimation in one-vertex events taken with whatever trigger is chosen for the runs with schema (3).

10 Level 7: Optional Parton correction

Contributor : J. Cammin

The corrections described in the previous sections correct back either to the parton level or the particle (hadron) level. Whatever correction will eventually be adopted, there will be a need to switch from the particle level to the parton level or vice versa. Take the initial scenario when jet energy scale corrections are available from the comparison of CaloJets to GenJets, hence delivering corrections back to the particle level. Measurements such as the top quark mass need to apply additional corrections for the parton level.

Those corrections can be large, as has been shown at the Tevatron. As an example, the CDF measurement of the inclusive jet cross section ([**?**]), which is compared to pQCD predictions, shows that the parton-level corrections are 9 % at $p_T \approx 50$ GeV (though with rather large uncertainties).

The parton level corrections account for effects from hadronization and the underlying event and can conveniently be derived from Monte Carlo simulations. The general procedure is to compare jets in Monte Carlo events at the hadron level to the initiating partons in bins of the jet transverse momentum after all other jet energy scale corrections have been applied to the CaloJets. Different hadronization and underlying event models result in different parton-level corrections, and so this type of corrections are specific for a given event generator. The size of the systematic uncertainty arising from the chosen model can be estimated by comparing parton-level corrections with different hadronization and underlying event parameters or even with different models.

11 Closure Tests

Contributors: A. Bhatti, D. Elvira, J. D'Hondt, P. Van Mulders

Closure tests are an essential tool to validate the procedures to derive the jet energy calibration as well as the actual correction values.

11.1 Validation of Methods

This type of closure tests have been used by both the DØ and CDF experiments at the Tevatron [4, 2] to compare corrected measured jets in the Monte Carlo with their associated particle level jets. If the corrections are derived following the same procedure as in data, the tests are a validation of the methods. DØ uses both γ-jets and dijets samples and defines a "direct" closure quantity:

$$D(p'_T, |\eta^{\mathrm{det}}_{\mathrm{jet}}|) = \frac{\langle E^{\mathrm{Corr}}_{\mathrm{jet}} \rangle}{\langle E^{\mathrm{ptcl}}_{\mathrm{jet}} \rangle} , \tag{24}$$

where $\langle E^{\mathrm{Corr}}_{\mathrm{jet}} \rangle$ is the mean corrected energy of a jet in a given p'_T and $|\eta^{\mathrm{det}}_{\mathrm{jet}}|$ bin, with p'_T the projection of the probe jet transverse momentum onto the direction of the tag object and $\eta^{\mathrm{det}}_{\mathrm{jet}}$ the probe jet pseudorapidity measured from the geographical center of the detector.$\langle E^{\mathrm{ptcl}}_{\mathrm{jet}} \rangle$ is the mean energy of the parent particle level jets. The only source of systematic uncertainty associated with this closure test is the matching criteria to relate a "measured" jet to its parent particle level jet, $\Delta R < R_{\mathrm{cone}}/2$ for a jet reconstructed with a fixed cone algorithm.

Deviations from unity outside of the errors from the JES correction and those of the closure test would point to flaws in the derivation methods. This test may also be used to validate the use of a given JES correction in samples different from those used for its derivation. In other words, it may help to answer the question on the bias associated with the use of a γ-jets derived JES on dijet, W+jets, top events, etc.

11.2 Validation of Correction Values

This type of closure tests are based on real collider data and use physics knowledge such as the known mass of unstable particles or transverse momentum balance in real collider events. We will concentrate on two methods associated to two sets of samples: γ/Z-jet and dijet events, $t\bar{t}$ (leptons plus jets) events.

11.2.1 Tests on γ/Z-jets Samples

This analysis involves stringent cuts on the purity of the photon or the Z samples, as well as on the $\Delta\varphi$ separation between them. Results will therefore be affected by systematic errors comming from the residual QCD background, and the asymmetric radiation patterns in the gamma and jet hemispheres. In addition, this is a test of the correction to the "parton level" since it involves a gamma or a Z and a deviation from closure is therefore expected in case we are testing particle level corrections.

One way to reduce systematic uncertainties is by defining a closure variable:

$$R_{\mathrm{closure}}(p'_T, |\eta^{\mathrm{det}}_{\mathrm{jet}}|) = \frac{\langle E^{\mathrm{Corr,Data}}_{\mathrm{jet}} \rangle}{\langle E^{\mathrm{Corr,MC}}_{\mathrm{jet}} \rangle} \quad . \tag{25}$$

Deviations from unity beyond the errors propagated from the JES correction and the closure test would imply biases in the JES correction being tested either because of problems in the sample selection and methods, or because it is being tested inconsistently on a sample different to the one used for its derivation. Although the uncertainties in R_{closure} are mitigated by the use of a MC to model the QCD background and the event topology, there are still errors coming now from the accuracy with which the Monte Carlo models the data. Sources of uncertainty will be the measurement of the gamma energy scale, the difference in topology between a pure photon sample in the MC and a contaminated one in the data, the modeling of hadron response in the Monte Carlo. The use of a realistically mixed MC sample (signal+background) and a "data driven" MC with a pion response tuned to data could help to reduce the error on the closure method.

A similar approach to the one described above is the so called "Hemisphere Method" which uses transverse momentum balance in the whole event rather than in a back-to-back di-object system. We define the observable H as:

$$H = \frac{\Sigma_{\mathrm{Probe}} |\vec{p}_T \cdot \hat{n}_{\mathrm{tag}}|}{\Sigma_{\mathrm{Tag}} |\vec{p}_T \cdot \hat{n}_{\mathrm{tag}}|} , \tag{26}$$

where the denominator is a sum of the p_T's of the objects in the hemisphere defined by the tag object and the numerator is the sum over the objects on the other side. The sources of error come from the energy scale of the tag object, the event selection (not the same sample used to derive the JES), energy resolution biases affecting low p_T jets, physics out-of-cone and unclustered energy effects. Again, these uncertainties can be mitigated by defining a $\Delta H = H^{\mathrm{data}} - H^{\mathrm{MC}}$. However, it is difficult to bring the resolution biases under control except with a Monte Carlo that models jet multiplicity, offset, and resolution very accurately. This method, pioneered by DØ , was discarded for this and for the reason that it does not test the JES in the sample it was derived from.

11.2.2 Corrections from mass constraints in top quark events

Processes in which top quarks appear have an important cross section in the proton collisions at the LHC. The production of top quark pairs, $pp \rightarrow t\bar{t}$, has a Next-to-Leading Order cross section of about 830 pb. With only a modest integrated luminosity of 100 pb^{-1}, a relevant sample of top quarks can be collected. With respect to the Tevatron $p\bar{p}$ collisions, the influence of the background processes like the production of W bosons with jets is neglegible due to a smaller cross section with respect to the top quark processes. The top quark decays with a branching ratio of about 100% to a W boson and a bottom quark. Therefore in the decay $t \rightarrow Wb \rightarrow q\bar{q}b$ two strong mass constraints are present. The first one comes from the precise measurement of the mass of the W boson, M_W, from the LEP and Tevatron experiments, while the second constraint is given by the Tevatron measurement of the top quark mass, M_t. This information together with the jet flavour tagging capabilities of the CMS experiment, we can identify a constraint system of 3 reconstructed jets in the final state of the top quark processes,

$$M_t^2 = \sum_{i=q,\bar{q},b} E_i^2 - \sum_{i=q,\bar{q},b} p_{x,i}^2 - \sum_{i=q,\bar{q},b} p_{y,i}^2 - \sum_{i=q,\bar{q},b} p_{z,i}^2 \quad (27)$$

$$M_W^2 = \sum_{i=q,\bar{q}} E_i^2 - \sum_{i=q,\bar{q}} p_{x,i}^2 - \sum_{i=q,\bar{q}} p_{y,i}^2 - \sum_{i=q,\bar{q}} p_{z,i}^2 \quad (28)$$

where the energies E_i and momenta $\vec{p}_i$ are from the jets in the hadronical decaying top quark. In general the 4-momenta of the reconstructed jet should fulfill these constraints if the jet energy scale is properly calibrated. This can be checked on an event-by-event basis with the selected and reconstructed top quark events. It was shown that an excellent statistical precision can be obtained with 1 fb^{-1} using only the W boson mass constraint [19].

The golden $t\bar{t}$ decay channel for this study is the semi-leptonic channel, where the W boson from one top quark decays hadronically and the W boson from the other top quark decays leptonically, reflecting a $t\bar{t} \rightarrow WbWb \rightarrow q\bar{q}bl\nu\bar{b}$ topology. When applying only loose selection cuts on the data, a rather pure sample of semi-leptonic decaying top quark pair events can be obtained [20]. With b-tagging tools a differentiation can be obtained between the light quark jet from the W boson decay and the remaining bottom quark jets from the top quark decay [21]. The remaining ambiguity between the bottom quark jets can be solved by the information provided by several topological observables. Therefore a relatively pure sample of three-jet systems reflecting the $t \rightarrow Wb \rightarrow q\bar{q}b$ topology can be selected from the data.

For each three-jet system the above constrained mass equations can be verified. For a succesfull closure test the equations should hold on average. Corrections to the jet energy scale, ΔE_i, can be introduced if the constraints are not fulfilled, yielding corrected jet energy scales, E_i^{corr}. In this procedure the mass of the jet can remain constant by rescaling the magnitude of the jets momentum accordingly. Because the mass constraints are true at the parton level, the residual corrections ΔE_i will correct the jet energy scale back to the primary parton level.

As an illustration of the residuals to be estimated an inclusive shift is being determined on the light and heavy quark jets from these top quark decays. The jets in top quark pair events reconstructed with the iterative cone algorithm of opening angle 0.5 are calibrated with Monte-Carlo based techniques using QCD events. From the knowledge of the generated parton momenta, an angular matching has been made between the reconstructed jet and the primary parton. When the jets do match the primary parton to better than $\Delta R(\eta, \phi)$ ¡ 0.3, the relative difference can be calculated between the energy scale of the jet and the parton. In Figure 14 the relative shifts are shown for respectively light and heavy quark jets depending on the flavour of the parton. In Figure 15 the mean of this distribution is differentiated as a function of the transverse momentum of the reconstructed jet. In these Figures, produced with TopReX events in CMSSW 131, the jets are within of pseudorapidity of $|\eta|$ ¡ 2.5.

With the application of a kinematic fit [22] the hypothesis of both mass constraints can be fitted together, obtaining optimal corrections on both the light and heavy jet energy scale. Forcing the three-jet system to fulfill the mass constraints will provide a χ^2 value. For each event this χ^2 value can be determined as a function of the corrections to be estimated, hence $\chi^2 = \chi^2(\Delta E_b, \Delta E_q, \Delta E_{\bar{q}})$. The minimum of the total χ^2 over all events can be projected into one dimension of either the light or heavy quark jets, after which a differentiation can be made as a function of the properties of the individual jets, for example the transverse momentum or the pseudorapidity. This method

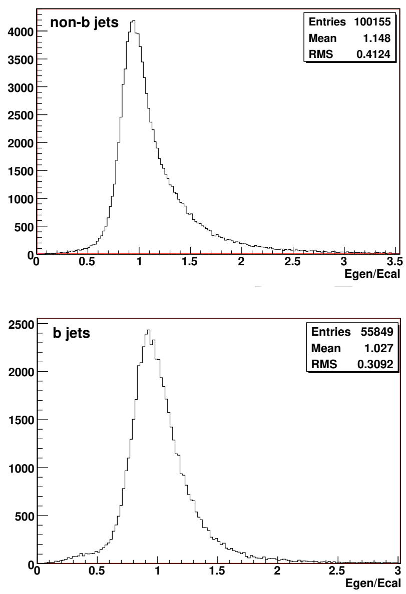

Figure 14: Distribution of E_{gen}/E_{cal} for light quark jets (left) and heavy bottom quark jets (right) for jet with p_T^{rec} ¿ 10 GeV and $|\eta|$ ¡ 2.5.

has the potential to perform the closure test to parton level jet energy scales for both light and heavy quark jets to a percent precision with only 1 fb^{-1} of data. The method can also provide estimates for the residual corrections after a calibration procedure applied to a certain level. It could check therefore the consistency of each of the individual steps in the jet energy scale calibration procedure. Unique for these processes and this method is the extraction of residual corrections of both light and heavy quark jets from the same event.

11.3 Summary

An accurate "data driven" Monte Carlo simulation is a key tool for validating data based methods. There are many ways to use the physics of collider events to validate the JES corrections. These methods may also be modified, as discussed in previous chapters, to derive the JES corrections.

12 The Role of the Monte Carlo Simulation

Contributors: D. Elvira.

A realistic and accurate, data driven, Monte Carlo simulation is an essential tool for the Jet Energy Scale project. As mentioned throughout this note, the simulation will be used to understand and demonstrate data based analysis procedures (closure tests), as well as for the direct derivation of some of the subcorrections. For example, a Monte Carlo simulation tuned to cosmic ray and test beam data may be the only tool we have in day one of data taking to anchor the absolute jet energy scale in the central pseudorapitity "control" region. That is because it will take some time to understand the calibration data taken during the first few months of the collider run. These data will eventually be used to perform an independent in-situ measurement of the jet energy corrections, as well as to re-tune the Monte Carlo to reproduce data distributions more precisely and therefore reduce the sytematic uncertainties of a Monte Carlo based jet energy calibration. In other words, at different stages of the experiment, we may use different tools to perform the jet calibration as they become available. The strategy should follow closely the physics goals and associated timescales. The first round of search analyses will probably not need a JES uncertainty better than 5-7%. Many measurements actually depend on the difference between MC and data corrections rather on its absolute error. Longer term precision measurements will probably care about ¡1% uncertainties, a physics based JES factorization, and good knowledge of error correlations. For example:

- The nominal "day one" JES corrections may be derived from a Monte Carlo simulation which should incor-

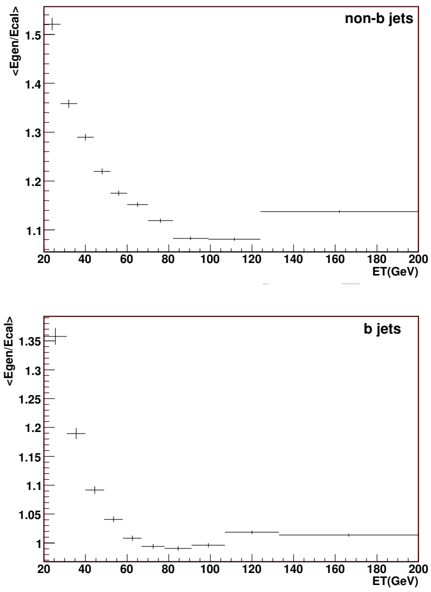

Figure 15: Dependency of the mean of E_{gen}/E_{cal} for light quark jets (left) and heavy bottom quark jets (right) as a function of the reconstructed E_T of the jet.

porate the lessons from the Cosmic Ray and Test Beam experiments. The first pass to data based corrections could be used as a measure of the systematic uncertainties which should be contained within the ¡5-7$\%$ range.

- The intermediate JES with a target uncertainty of 3-4a better understanding of the data based methods and closure tests, as well as from the progress in Monte Carlo tuning using dedicated calibration triggers.

- The long term JES with a target uncertainty of $\approx 1\%$ will probably be based on a very accurate but biased data based correction and a highly tuned data driven MC which could be used to extend the range of validity of the corrections to different samples, energy ranges, and algorithms.

The current simulation effort within the offline and DPG groups is targeting a simulation tool suited to the needs of day one physics. Through the Spring and Summer of 2007, work has been performed to improve and validate the detector geometry descriptions, including the detector parts, and material budget. The digitization code was also significantly improved, with calibration constant tunning, and updated descriptions of signal and noise based on recent test beam and cosmic ray experiments. One important step toward achieving a realistic Monte Carlo is related to the development of a Monte Carlo/data mixing tool to be able to overlay, for example, a real Min-Bias or Zero-Bias collider event to a Monte Carlo signal event. Another big step will be given with the availability of the GFlash infrastructure to parameterize both EM (already available) and Hadronic showers. The elements and tools described above are critical to achieve accurate descriptions of the offset energy, the hadronic energy response linearity and resolution, as well as transverse and longituninal shower shapes. These characteristics are essential for the JES corrections to be similar in the data and the Monte Carlo. The MC/data mixing tool will allow to produce MC samples with a realistic modeling of pileup and underlying event and reproduce accurately the effects of pedestal subtruction, zero suppression, and algorithm parameter thresholds which contribute the jet energy scale. Geant4 physics is based on theoretical calculations and data based parameterizations to describe the many different processes associated with the interaction of particles with matter. Different sets of models are grouped in a small set of "physics lists" which we can select from. It is not possible, however, to tune the Geant4 simulation to the CMS collider data. The tuning may be done at different levels by, for example, re-weighting the EM and HAD fractions of the simulated hit energy to adjust response linearity, including calibration constants and sampling weights following the same model as in the data to optimize resolution and detector uniformity, replacing the Geant4 shower by parameterizations after the first inelastic interaction to adjust linearity, resolution, and shower shapes. The tuning of the Monte Carlo will grow into a very challenging and man-power demanding collaboration effort between the simulation, DPG, and POG groups. The program would involve the definition of calibration triggers to collect Min-Bias and Zero-Bias events, and measure isolated charged tracks and shower properties.

13 Conclusions

References

[1] R. Harris, **CMS AN-2005/034**, *"Jet calibration from dijet balancing"*.

[2] A. Bhatti et al. (CDF Collaboration), **hep-ex/0510047** (2005), *"Determination of the Jet Energy Scale at the Collider Detector at Fermilab"*.

[3] L. Babukhadia, D. Elvira (D0 Collaboration), **D0 Note Number 003554** (1998), *"Method to Test the Jet Energy Scale Using Transverse Momentum Balance"*; B. Abbott et al. (D0 Collaboration), **Nucl. Instrum and Methods A424**, 352-394 (1999)

[4] Jet Energy Scale Determination At DØ Run II, private communications with the DØ Jet Energy Scale group. NIM article in preparation.

[5] CMS Physics Technical Design Report (Vol I).

[6] B.Hirosky, Jet energy scale at CDF and D0, talk at JetMET meeting, http://indico.cern.ch/conferenceDisplay.py?confId=15199

[7] J.L.Agram et al, Jet energy scale at D0, Run II, D0 Note 4720.

[8] P.Janot and C.Bernet, see for example, "Clustering in ECAL & Eflow framework", CMS Energy Flow Meeting, C ERN, 20 January, 2006.

[9] P.Janot, see for example, "Cluster and calibration for hadrons", CMS HCAL Detector Performance Group Meeting, CERN , 5 February, 2007.

[10] P.Janot, Private Comunication.

[11] O.Kodolova,I.Vardanyan,A.Nikitenko,A.Oulianov,CMS NOTE-2006/050, 2006, http://dx.doi.org/10.1140/epjc/s10052-007-0223-9, 2007.

[12] V.Gavrilov, O.Kodolova, A.Oulianov, I.Vardanyan, CMS-RN-2003-004.

[13] D.Acosta et al, CMS NOTE 2006-067, 2006.

[14] V. Konoplinikov, A Ulyanov and O. Kodolova, *Jet Calibration using γ-jet events in the CMS Detector* CMS Note 2006/042

[15] D. Acosta $\widehat{e}$t al. The CDF Collaboration *Study of Jet Shapes in Inclusive Jet Production in $p\bar{p}$ Collisions at $\sqrt{s}$ Phys. Rev.* **D***71, 112002 (2005).*

[16] S. Esen, G. Landsberg, **CMS Internal Note in Preparation**, *"3D MCJet Corrections in CMSSW_1_2_0"*.

[17] M. Vazquez Acosta et al., *"Jet and MET Performance in CMSSW_1_2_0"* CMS Internal Note, CMS IN-2007/053.

[18] A. K. Nayak, A. Nikitenko, T. Aziz, **CMS Note 2007/014** (2007), *"Evaluation of the b-jet energy corrections using $b\bar{b}Z$, $Z \to \ell\ell$ process."*.

[19] J.D'Hondt *et al.*, *' Light quark jet energy scale calibration using the W mass constraint in single-leptonic'*, CMS Note 2005-027

[20] J.D'Hondt, J.Heyninck, S.Lowette, *'Measurement of the cross section of single leptonic ttbar events'*, CMS Note 2006-064

[21] J.D'Hondt, J.Heyninck, S.Lowette, *'Top quark mass measurement in single leptonic ttbar events'*, CMS Note 2006-066

[22] J.D'Hondt *et al.*, *'Fitting of Event Topologies with External Kinematic Constraints in CMS'*, CMS Note 2006-023