Measurement of jet energy scale corrections using top quark events

Petra Van Mulders *(on behalf of the CMS Collaboration)*
Vrije Universiteit Brussel, Inter-university Institute for High Energies, Pleinlaan 2, B-1050 Brussels, Belgium

Top Quark conference 2008, La Biodola (Elba), Italy, 19-23 May 2008
email: pvmulder@vub.ac.be
Reference Physics Analysis Summary page CMS PAS TOP-07-004.

A least-square kinematic fit using Lagrange multipliers is applied to enforce the mass constraints in the reconstructed $pp \to t\bar{t} \to q\bar{q}b\mu\nu_\mu\bar{b}$ events. Both the W boson mass and the top quark mass in the hadronic top quark decay, $t \to Wb \to q\bar{q}b$, are constrained to agree with their measured values. Residual corrections are estimated on the energy scale of the jets arising from both the light quarks in the W boson decay and the heavier bottom quark in the top quark decay. Utilizing the first integrated luminosity of 100pb^{-1} of proton collisions at 14 TeV detected by the CMS detector, an uncertainty smaller than 1% can be obtained on the jet energy scale for both light and heavy jets. This analysis is based on detailed simulations of both the physics processes and the detector responds.

1 Introduction

Top quark events are abundant in the 14 TeV proton collisions at the LHC. Within CMS we will trigger and reconstruct these events in the muonic decay channel, being $t\bar{t} \to bWbW \to bq\bar{q}b\mu\nu_\mu$. It is a challenge however to calibrate the energy scale of the reconstructed jet to the parton level quark. Several data driven methods will be applied on these events to correct the jet energy scale. Applying on an event-by-event basis a kinematic fit forcing the world average W boson and top quark mass constraints to the reconstructed event, the residual jet energy scale corrections can be estimated. The potential of this new method with 100pb^{-1} of integrated luminosity is studied with detailed simulated events. The AlpGen generator has been utilized to simulate the physics processes of interest in this paper, being $t\bar{t} + jets$ (NLO cross section) and $Z/W + jets$ (LO cross section) events. The QCD multi-jet background has been simulated with PYTHIA. A state-of-the-art GEANT based simulation program is applied to simulate the interaction with the CMS detector. Alignment and calibration correction factors are used as they would have been estimated from a data sample with an integrated luminosity of 10pb^{-1}.

The detector final state can be characterized by four hadronic jets of which two originate from a heavy quark, an isolated muon and missing transverse energy. The events are reconstructed using the general CMS Software framework and the single-muon High Level Trigger criterium is applied. The jets are reconstructed from the combined electromagnetic and hadronic calorimeter deposits and clustered with the Iterative Cone algorithm using an opening angle of $\Delta R = 0.5$ in an (η, ϕ)-metric. The energy scale of the jets is calibrated with corrections based on simulated events applying a procedure which compensates the responds on E_T in QCD events as a function of the transverse energy of the reconstructed jet. The four jets with the highest transverse energy in the pseudo-rapidity range $|\eta| \leq 2.4$ are considered to be the hadronic decay products of the semi-leptonic $t\bar{t} \to bq\bar{q}b\mu\nu_\mu$ process. A muon candidate is formed when a muon track is reconstructed in the muon chambers and a matching track is found in the main tracker. These reflect the so-called global reconstructed muons in the CMS detector. In the analysis the muon with the largest transverse momentum in the pseudo-rapidity range $|\eta| \leq 2.1$ is used to reconstruct the topology of the top quark pair. To differentiate between light and heavy quark jets, the nominal track counting b-tagger is used when identifying the correct jet combination.

2 Event selection and jet combination

The four jets with the highest transverse energy in $|\eta| \leq 2.4$ are required to have a transverse energy above 40 GeV. The muon with the highest transverse momentum in $|\eta| \leq 2.1$ is requested to have a transverse momentum above 30 GeV/c. The muon tracker isolation variable $p_T^{iso} = \left(\sum_{\Delta R(track,muon)<0.3} p_{T,tracks}\right) - p_{T,muon}$ and a calorimeter isolation variable $E_T^{iso} = \sum_{\Delta R(calo,muon)<0.3} E_{T,calo}$ are required to be smaller than respectively 3 GeV/c and 5 GeV/c. The direction of the muon in the above equations is taken at the vertex. To avoid difficulties in the interpretation of the result, the four jets are required not to overlap in the (η, ϕ) space in which they are constructed by the jet algorithm. It is observed that even after applying the muon isolation criterium described, the smallest angle between the muon and the leading jets has an important differentiation power between the QCD multi-jet events and the signal of interest. A selection cut is therefore applied requiring that this angle has to exceed 0.5.

For the selected events, four variables are being calculated to select the correct jet combination into a $t\bar{t} \to bq\bar{q}b\mu\nu_\mu$ topology. For each of the possible jet combinations the following variables are determined

- Transverse momentum of the hadronic top quark in the combination scaled to the average transverse momentum of the hadronic top quarks over all possible jet combinations.
- Sum of the transverse momenta of the two jets assumed b jets in the combination divided by the sum of the transverse momenta of the two other jets.
- The angle in (η, ϕ) space between the two assumed light jets in the combination.
- Sum of the track counting b-tag discriminants of the two jets assumed b jets.

The likelihood ratio to be a correct jet combination obtained from each observable is combined into a likelihood ratio variable per jet combination. The jet combination in the event with the highest logarithm of the combined likelihood ratio, P_C^{max}, is taken as the correct combination. Figure 2 shows the highest logarithm of the combined likelihood ratio value per event and the probability of having a correct combination as function of this variable. This variable helps in rejecting background processes, and an extra sequential cut is made demanding the event to have a value of P_C^{max} larger than 0. After this event selection the sample is dominated by signal muonic top quark pair decays. For an integrated luminosity of 100pb^{-1} we have 755 signal events compared to 404 background events. Only a small amount of QCD multi-jet events are expected at this point in the event selection.

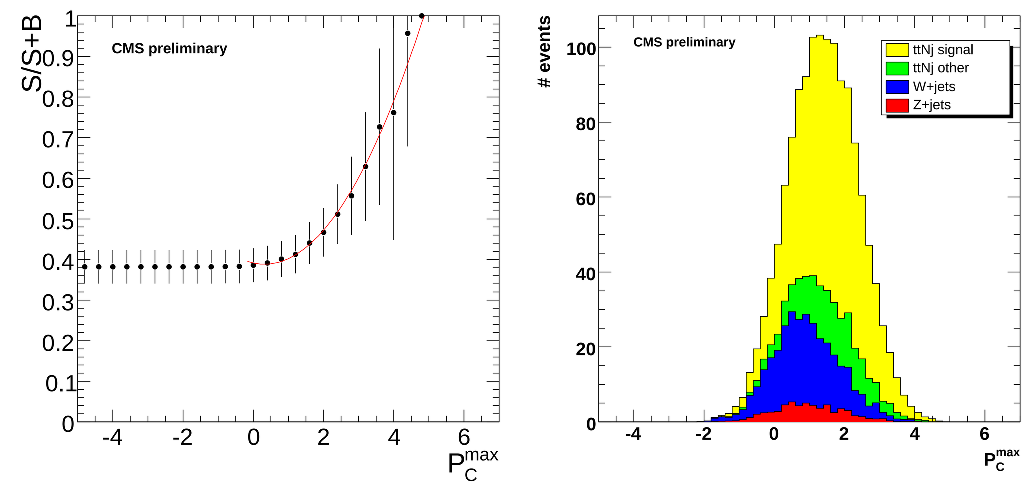

Figure 1: Probability of having the correct jet combination in the event as a function of the maximum of the logarithm of the combined likelihood ratio variable of the event (left) and distribution of the maximum of the logarithm of the combined likelihood ratio value in the event.

3 Method

At this point in the analysis we apply a general non-linear least square fit using Lagrange Multipliers constraining the reconstructed W boson and top quark masses in the hadronic top quark decay ($t \to bW \to bq\bar{q}$) of the selected jet combination to the world average values. The momenta of the jets are parametrized in the (E_T, θ, ϕ) dimensions and the resolutions of these parameters are obtained from simulation as a function of the pseudorapidity and transverse momentum of the object. The convergence rate of the applied kinematic fit for the selected signal events and the chosen jet combination is about 99%.

The kinematic fit provides a probability obtained from the χ^2 value of the fit. This probability reflects how likely the reconstructed event in the choosen jet combination agrees with the hypothesis of the world average empirical determinations of the top quark and W boson masses. The energy scale of the jets can be adapted to maximize this probability over all selected events. This was obtained by determining this χ^2 probability, P^{fit}, in the dimensions of the three residual jet energy corrections, resulting in $P^{fit} = P^{fit}(\Delta E_b, \Delta E_q, \Delta E_{\bar{q}})$. The energies and accordingly the momenta of the jets are shifted in steps of 2.5% between -50% and +50% for light quark jets from the W boson decay and for bottom quark jets, keeping the $E/|\vec{p}|$ ratio of the jet invariant. As we are estimating inclusive correction shifts on the jet energy scale the procedure was simplified by reducing the dimensions from three to two. For this the correction applied on both jets arising from the W boson decay in the event were taken to be equal, $\Delta E_q = \Delta E_{\bar{q}}$.

Some more event selection cuts are applied to make the analysis more robust, hence by enhancing events for which effectively a good reconstruction matching the parton level is found. This is obtained by requiring that the fit probability when applying no additional shifts on the jet energy scale exceeds 0.01 or $P^{fit}(\Delta E_b = 0, \Delta E_q = \Delta E_{\bar{q}} = 0) \geq 0.01$. Events are only taken into account in the method if they have in the scanned range of corrections a maximal fit probability exceeding 0.98. The distribution of the maximum fit probability for all events is shown in Figure 3 together with the fit probability distribution when no residual jet energy scale corrections are introduced, hence $P^{fit}(\Delta E_b = 0, \Delta E_q = \Delta E_{\bar{q}} = 0)$ for each event.

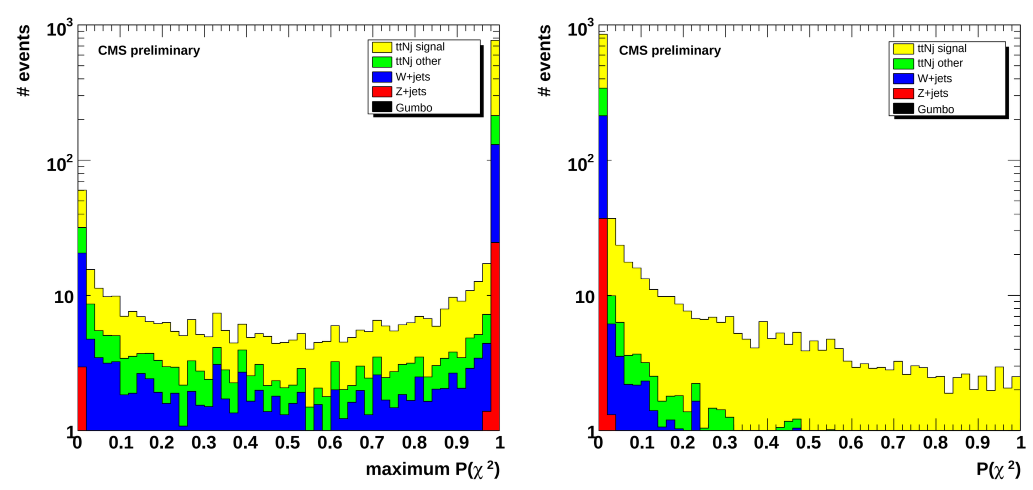

Figure 2: Maximal fit probability in the scanned range of jet energy corrections (left) and fit probability without any residual jet energy corrections (right)

The last event selection cut is applied on outliers with respect to the central value of the residual corrections to be estimated. To first order the residual corrections can be estimated from the inclusive W boson mass. The W boson mass distribution is constructed from all selected events and each event is weighted with the probability it has to have the correct selected jet combination. The relative difference between the fitted expectation value and the true world average W boson mass, is taken as an initial estimate of the residual shift on the light quark jet energy scale. The central value for the b quarks is estimated from the central value for the light quarks, obtained from the fit on the W boson mass spectrum, but with a shift calculated from the simulation being 7%.

The first estimates of this inclusive residual corrections obtained on the inclusive mass spectra of the W boson, are taken as central values for an outlier cut. All events for which the best values of the corrections deviate by more than 20% from these central values are rejected. At this point we have 130 signal events compared to 27 background events of which half of them are other $t\bar{t}$ decays in which also a hadronic top quark decay is present.

The best values of the residual corrections $(\Delta E_b, \Delta E_{\bar{q}} = \Delta E_q)$ are obtained when the combined $\Delta\chi^2(\Delta E_b, \Delta E_{\bar{q}} = \Delta E_q)$ distribution is minimized. Figure 3 shows the 5σ contour in dimensions of the two inclusive corrections. The estimate of the corrections is obtained by projecting this into each dimension separate.

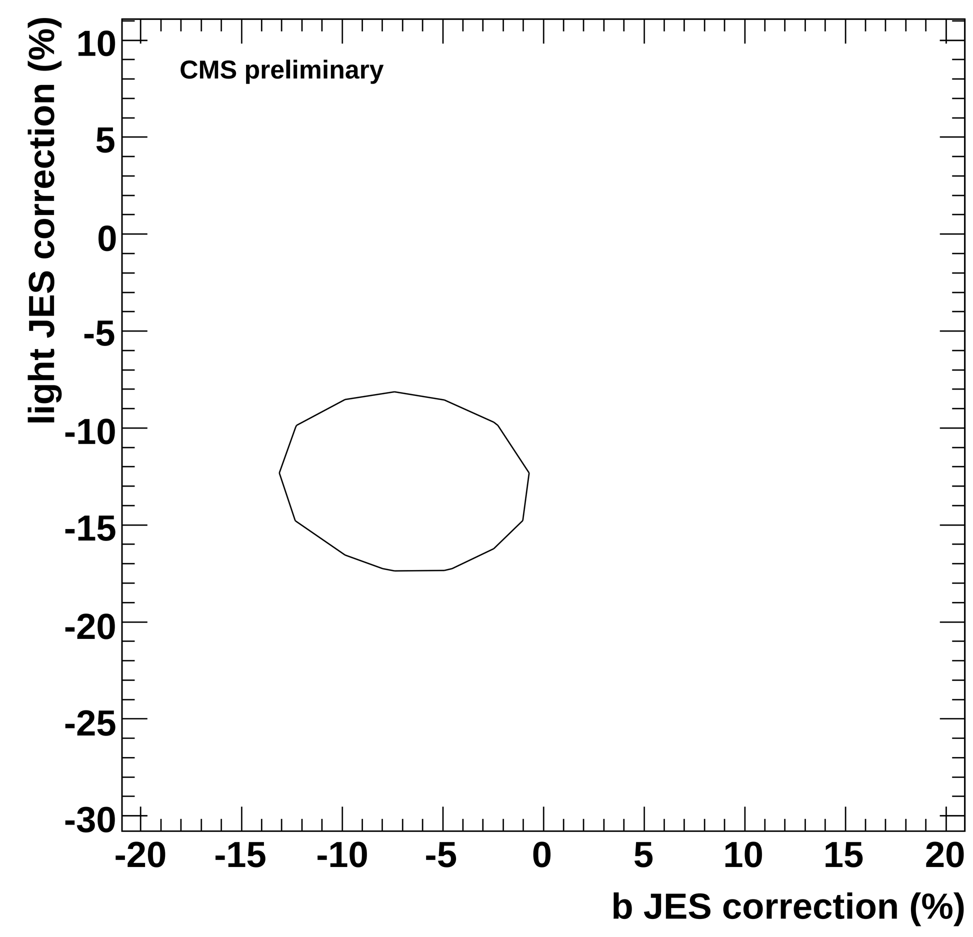

Figure 3: The distribution of $\Delta\chi^2(\Delta E_b, \Delta E_{q,\bar{q}})$ represented as a 5σ confidence interval.

4 Results and discussion

For the reconstructed jets initially calibrated with simulation based techniques, the residual corrections are found to be

$$\Delta E_b = -7.0 \pm 0.9\%$$

$$\Delta E_{q,\bar{q}} = -12.9 \pm 0.9\%$$

where the uncertainties reflect the expectation with a data set of 100pb^{-1} of integrated luminosity. For the light quark jet energy scale correction the information on both light quark jets is combined. The uncertainties are corrected to obtain a unity width of the pull distribution determined via resampling techniques.

A calibration curve is obtained by applying an inclusive relative shift on the energy scales of the reconstructed jets prior to the full analysis. The estimates of the residual corrections on the jet energy scale are compared to the expected corrections for the selected events. The slope of the calibration curve of the estimator for the light quark jet energy scale correction is found to be equal to 0.77 ± 0.02, while the one for the b quark jet energy scale correction is 0.87 ± 0.03. If we apply the presented method after a first jet energy scale calibration, the residual calibration should be close to zero. Hence the non-unity slope will not have an important impact in the measurement.

Several systematic effects can influence the method described. The main effect is expected to be the contribution of pile-up collisions although this could be considered as part of the residual jet energy scale correction to be estimated and differentiated versus the amount of reconstructed and identified primary vertices. Other sources of systematic uncertainties arise from the uncertainty in the level of several background contributions. The sensitivity to both the jet combinatorial and the process background is estimated and found to be negligible. An extra smearing on the jet energies is applied prior to the analysis increasing the Gaussian width of the residual distribution with a factor of 2, but without a significant effect on the estimated jet energy scale corrections. From data itself it will be possible to estimate the sensitivity of the results with respect to different kind of backgrounds by changing the selection cuts in the analysis. For example the combined likelihood variable and the maximum fit probability cut to estimate the sensitivity to respectively the combinatorial and process background.

5 Conclusions

Given the small sensitivity to systematic uncertainties, the jet energy scale can potentially be estimated from top quark events with a precision of about 1% using the first 100pb^{-1} of data. The method was applied on $pp \to t\bar{t} \to q\bar{q}b\mu\nu_\mu\bar{b}$ events, but can similarly be used on $pp \to t\bar{t} \to q\bar{q}be\nu_e\bar{b}$ events increasing the statistical information with about a factor of 2.