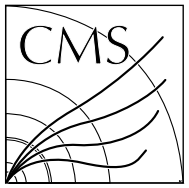
Available on CMS information server

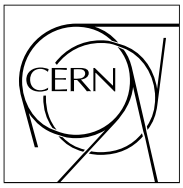
CMS AN-2008/001



The Compact Muon Solenoid Experiment

Analysis Note

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



14 March 2008

Performance of Jet Algorithms in CMS

P. Schieferdecker

CERN, Geneva, Switzerland

P. Kurt

Cukurova University, Adana, Turkey

G. Dissertori

ETH Zürich, Zürich, Switzerland

F. Chlebana, D. Elvira, K. Kousouris

Fermilab, Batavia, IL, USA

L. Apanasevich, C. Dragoiu, A. Smoron, N. Varelas

University of Illinois, Chicago, IL, USA

F. Ratnikov

University of Maryland, College Park, MD, USA.[1)]

Z. Qi, M. Zielinski

University of Rochester, Rochester, NY, USA

A. Bhatti

The Rockefeller University, New York, NY, USA

Abstract

The CMS Combined Software and Analysis Challenge 2007 (CSA07) is well underway and expected to produce a wealth of physics analyses to be applied to the first incoming detector data in 2008. The JetMET group of CMS supports four different jet clustering algorithms for the CSA07 Monte Carlo samples, with two different parameterizations each: Fast k_{T}, SISCone, MidPoint Cone, and Iterative Cone. We present several studies comparing the performance of these algorithms using QCD dijet and $t\bar{t}$ Monte Carlo samples. We specifically observe that the SISCone algorithm performs equal to or better than the Midpoint Cone algorithm in all presented studies and propose that SISCone be adopted

[1)] On leave of absence from Institute of Theoretical and Experimental Physics, Moscow, Russia.

as the preferred cone-based jet clustering algorithm in future CMS physics analyses, as it is preferred by theorists for its infrared- and collinear-safety to all orders of perturbative QCD. We furthermore encourage the use of the Fast k_{T} algorithm which is found to perform as good as any other algorithm under study, features dramatically reduced execution time w.r.t. previous implementations of the k_{T} algorithm, and is infrared- and collinear save as well.

1 Introduction

Almost every process of interest at the LHC contains quarks or gluons in the final state. The partons can not be observed directly, but fragment into stable hadrons, which can be detected in the tracking and calorimeter systems. This note describes the latest performance studies of several algorithms which cluster energy deposits in the CMS calorimeters into collimated objects of stable particles, "CaloJets". Calorimeter jets are expected to yield a good description of both the parton-level and the hadron showers emerging from the hard interaction. For Monte Carlo events, the hadron-level is defined by applying the same clustering algorithms, which are typically formulated to accept any set of four vectors as inputs, to all stable particles from the MC truth record ("GenJets"). Hadron-level is also referred to as "particle-level", and jet energy scale corrections based on MC are derived to correct back to this detector-independent level.

Calorimeter jets are reconstructed using energy deposits in calorimeter towers ("CaloTowers") as inputs: they are composed of one or more HCAL cells and corresponding ECAL crystals. The unweighted sum of energy deposits of one single HCAL cell and 5x5 ECAL crystals form a projective tower in the barrel ($|\eta| < 1.4$). A more complex association between HCAL cells and ECAL crystals is required in the forward region. The standard jet reconstruction applies the "Scheme B" thresholds [1] on calorimeter cells and an overall tower threshold $E_{\mathrm{T}} > 0.5\,\mathrm{GeV}$, summarized in Table 1 and relevant for all studies presented in this note.

Scheme	HB [GeV]	HO [GeV]	HE [GeV]	$\sum EB$ [GeV]	$\sum EE$ [GeV]
B	0.90	1.10	1.40	0.20	0.45

Table 1: Energy thresholds (in GeV) for calorimeter noise suppression "Scheme B". $\sum EB$ and $\sum EE$ refer to the sum of ECAL energy deposits associated with the same tower in the barrel and endcap respectively.

The studies presented in this note are based on QCD dijet and $t\bar{t}$ Monte Carlo samples without pileup, produced and reconstructed with CMSSW_1_5_2 and analyzed with CMSSW_1_6_X. It is often necessary to associate CaloJets with GenJets in these samples to probe how well the calorimeter-level reconstruction represents the hadron-level of the process. This association is based on spatial separation in η-ϕ-space between the two jet axes by requiring

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{1}$$

to be less than a certain value. Similarly, GenJets (and hence their associated CaloJets) are assigned the same parton flavor as the matched MC parton from the hard interaction.

Besides good correspondence to the parton-level and hadron-level, a successful jet algorithm should fulfill two important requirements. It should be *collinear-safe*, such that the outcome remains unchanged if e.g. the energy carried by a single particle is instead distributed among two collinear particles. Collinear safety is typically endangered if the jet finding is based on energetic seeds and a threshold is applied to these seeds. The algorithm should be *infrared-safe*, such that the result of the jet finding is stable against the addition of soft particles. Jet algorithms which don't comply with either or both of these requirements yield ambiguous results and lead to unnecessary uncertainties when applied to calculations in perturbation theory.

The performance of the following four jet clustering algorithms that are supported by the CMS JetMET group for CSA07 samples are discussed in this note:

- The Iterative Cone algorithm is a simple cone-based algorithm employed by CMS online in the High Level Trigger (HLT). It has a short and predictable execution time. Calorimeter towers and particles with $E_{\mathrm{T}} > 1\,\mathrm{GeV}$ are considered in descending order as starting points (seeds) for an iterative search for stable cones such that all inputs with $\sqrt{\Delta\eta^2 + \Delta\phi^2} \leq R$ from the cone axis are associated with the jet, R being the cone size parameter. A cone is considered stable if its geometric center agrees with the (η, ϕ) location of the sum of the constituent four vectors within a certain tolerance. Once a stable cone is found, it is declared a jet and its constituents are removed from the remaining inputs. The algorithm is neither collinear- nor infrared-safe.

- The Midpoint Cone [2] algorithm is based on an iterative procedure to find stable cones as well. Infrared-safety is addressed however by considering the midpoints between each pair of (proto-)jets which are closer than twice the cone radius R as additional seeds. Moreover, each input object (tower, particle, etc.) can initially be associated with several proto-jets, and a splitting and merging algorithm is applied afterwards to ensure each input object appears in one jet only. Despite these improvements to the cone-based clustering procedure, the algorithm has been shown not to be infrared-safe for pQCD orders beyond NLO. Note that the same seed requirements as for the Iterative Cone algorithm are imposed.

- SISCone [3] is the "Seedless Infrared-Safe Cone" jet algorithm. It is collinear- and infrared-safe to all orders of pQCD and demands only slightly higher execution time compared to the Midpoint Cone algorithm. The code is supported and available publicly with a detector-independent interface ensuring that different experiments can compare results with the exact same clustering code applied.
- Fast $\boldsymbol{k_T}$ [4] is a recent implementation of the k_T algorithm [5] which is also collinear- and infrared-safe. It has a dramatically reduced execution time w.r.t. previous implementations of the k_{T} algorithm. It is the only sequential recombination jet algorithm currently supported in CMS. The radius- or size-parameter D plays the corresponding role to the cone size parameter R for cone algorithms. By construction any pair of clustered k_{T} jets is guaranteed to be separated by $\sqrt{\Delta y^2 + \Delta\phi^2} > D$.

These four algorithms can be grouped into two general categories: seeded and seedless. The "E-Scheme" is used for all algorithms as the recombination scheme: the energy and momentum of a jet are defined as the sums of energies and momenta of its constituents.

Figure 1 shows the CPU time requirements for each algorithm to cluster all calorimeter towers in the event passing the E_{T} thresholds, using QCD dijet events. The execution time of the Fast k_{T} algorithm is comparable to the Iterative Cone algorithm without the discussed deficiencies of the latter. The SISCone algorithm requires more CPU resources compared to the Midpoint Cone algorithm. The time spent for the jet reconstruction ($\approx 0.02\,s$) of each event however is small compared to the total event reconstruction time ($\approx 10\,s$): the particular jet algorithm choice does therefore not impact the overall CPU requirements per event significantly. A version of SISCone with execution times reduced by about 10% was recently made available and will be adopted by CMS in the near future.

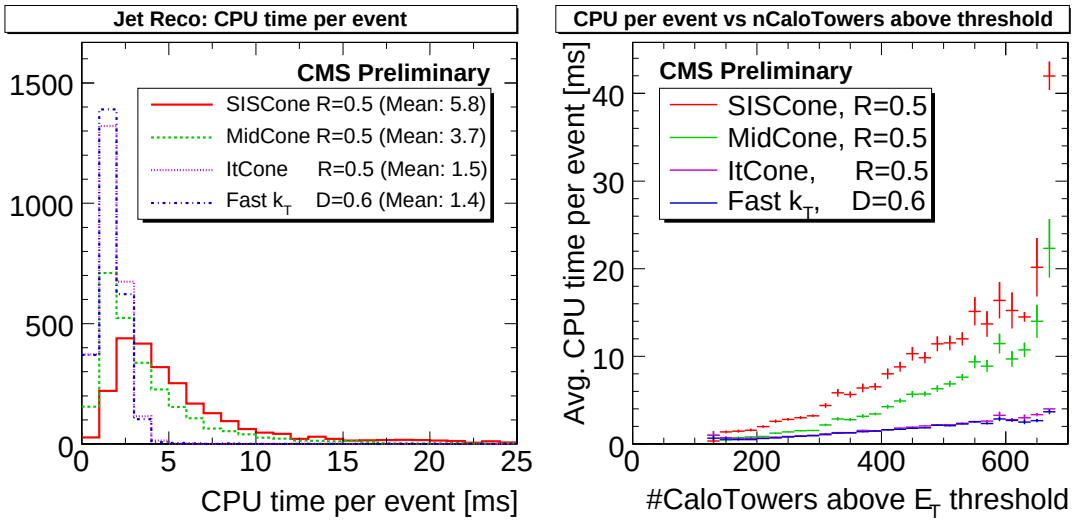


Figure 1: *Left:* CPU time required for each jet algorithm to cluster all CaloTowers above the E_{T} threshold of $0.5\,\mathrm{GeV}$ into jets. *Right:* Average CPU time as a function of the number of CaloTowers above E_{T} threshold.

Section 2 briefly discusses the procedure and results to derive the jet energy scale corrections; The jet response before and after application of these corrections in QCD dijet events is presented in Section 3, followed by the position resolutions in η and ϕ in Section 4; Section 5 is devoted to the quantification of reconstruction efficiencies and fake rates, based on the ability to match CaloJets to GenJets spatially within a certain ΔR; jet energy resolutions are derived in Section 6 based both on MC truth information and by means of data-driven methods, and the results are compared and discussed; Section 7 compares the performance of different jet algorithms to reconstruct $t\bar{t}$ events, considering the reconstruction efficiency of hadronically decaying top quarks, and their mass resolution; the energy distribution within a jet is discussed in Section 8; finally, the comparisons of the performance of the Midpoint Cone and SISCone algorithms are summarized in Section 9. We conclude with the recommendation to use SISCone as the default cone-based algorithm for future CMS analyses.

2 Jet Energy Scale Corrections

In this section we briefly describe the so-called "MCJet" corrections which are used to correct the calorimeter level jets (CaloJets) back to particle level jets (GenJets). These corrections are derived from the CMSSW_1_5_2 QCD datasets (data set names include the string *CMSSW_1_5_2-CSA07*). The samples were generated in 21 $\hat{p}_T$ bins in

order to have a sufficient number of events across the entire p_{T} range, and about 1M events from 17 of these $\hat{p}_T$ bins were used.

A detailed description of the MCJet corrections can be found in [6]. Note that the procedure to derive jet energy corrections at CMS both based on MC and data will undergo major changes in the near future [7].

2.1 MCJet Response Definition

The fundamental measurement in the process of deriving the jet corrections is that of the $Response$ which is defined as:

$$Response = \frac{CaloJetE_T}{GenJetE_T} \tag{2}$$

All GenJets in the event are considered and matched to the closest CaloJet, defined as the one which minimizes ΔR as defined in Equation (1). If $\Delta R > 0.25$, the GenJet is discarded. The $Response$ (eq.2) distributions for matched GenJets are recorded in bins of $GenJetE_T$ and $CaloJet|\eta|$, potential differences in the response for negative and positive η are not accounted for. A Gaussian fit in the $\pm 1\sigma$ range around the maximum determines the peak position of each distribution, as the mean itself is affected by asymmetric tails. The individual fit results of all $GenJetE_T$ bins are fitted with a smooth function for each $|\eta|$ bin to derive the $Response$ meaurement.

2.2 MCJet Correction Definition

The jet energy correction aims at determining the $GenJetE_T$ given a measured $CaloJetE_T$ at some η. Mathematically, it is equivalent to solving the nonlinear equation:

$$CaloJetE_T(\eta) = GenJetE_T \times Response(GenJetE_T, |\eta|) \tag{3}$$

The approach followed for this purpose is iterative, starting with:

$$k(CaloJetE_T, \eta) = \frac{1}{Response(CaloJetE_T, |\eta|)} \ . \tag{4}$$

The CaloJet Lorentz vector p is multiplied by the jet correction to obtain a *corrected* CaloJet Lorentz vector p':

$$p' = kp \tag{5}$$

However, the $Response$ is derived as a function of $GenJetE_T$ and the previous calculation needs to be repeated until the mean *corrected* $CaloJetE_T$ converges to $GenJetE_T$. If i is the iteration number, then k_i is the correction obtained in the i^{th} iteration and is equal to:

$$k_i(CaloJetE_T, \eta) = \frac{1}{Response(CaloJetE_T \times k_{i-1}, |\eta|)} \tag{6}$$

with $k_0 = 1$. The procedure is applied within a fixed $CaloJet|\eta|$ bin for all bins and the results are not interpolated, leading to a discontinuity at the η bin boundaries.

The MCJet corrections for different jet algorithms, shown in Figure 2, lead to a few general remarks:

- Jets reconstructed with larger jet size parameter ($D = 0.6$ and $R = 0.7$) tend to require larger corrections.
- The differences among jets with comparable jet size parameters are more pronounced at low E_T values.
- As $|\eta|$ increases, the differences among jets with different jet size parameters become larger.

3 Jet Energy Response

For the study described in this section, each CaloJet is matched to the closest GenJet within $\Delta R < 0.3$ or otherwise discarded. For successfully matched pairs, the jet response, $R_{\mathrm{jet}} = p_{\mathrm{T}}/p_{\mathrm{T}}^{\mathrm{gen}}$, is calculated. Note that this definition, based on the p_{T} variable, is different than the one used in the section on MCJet corrections.

Three generated pseudorapidity bins have been defined to highlight distinct regions of the detector. The first, representing the barrel region, includes jets with pseudorapidity between $0 < |\eta^{\mathrm{gen}}| < 1.4$. For the endcap region,

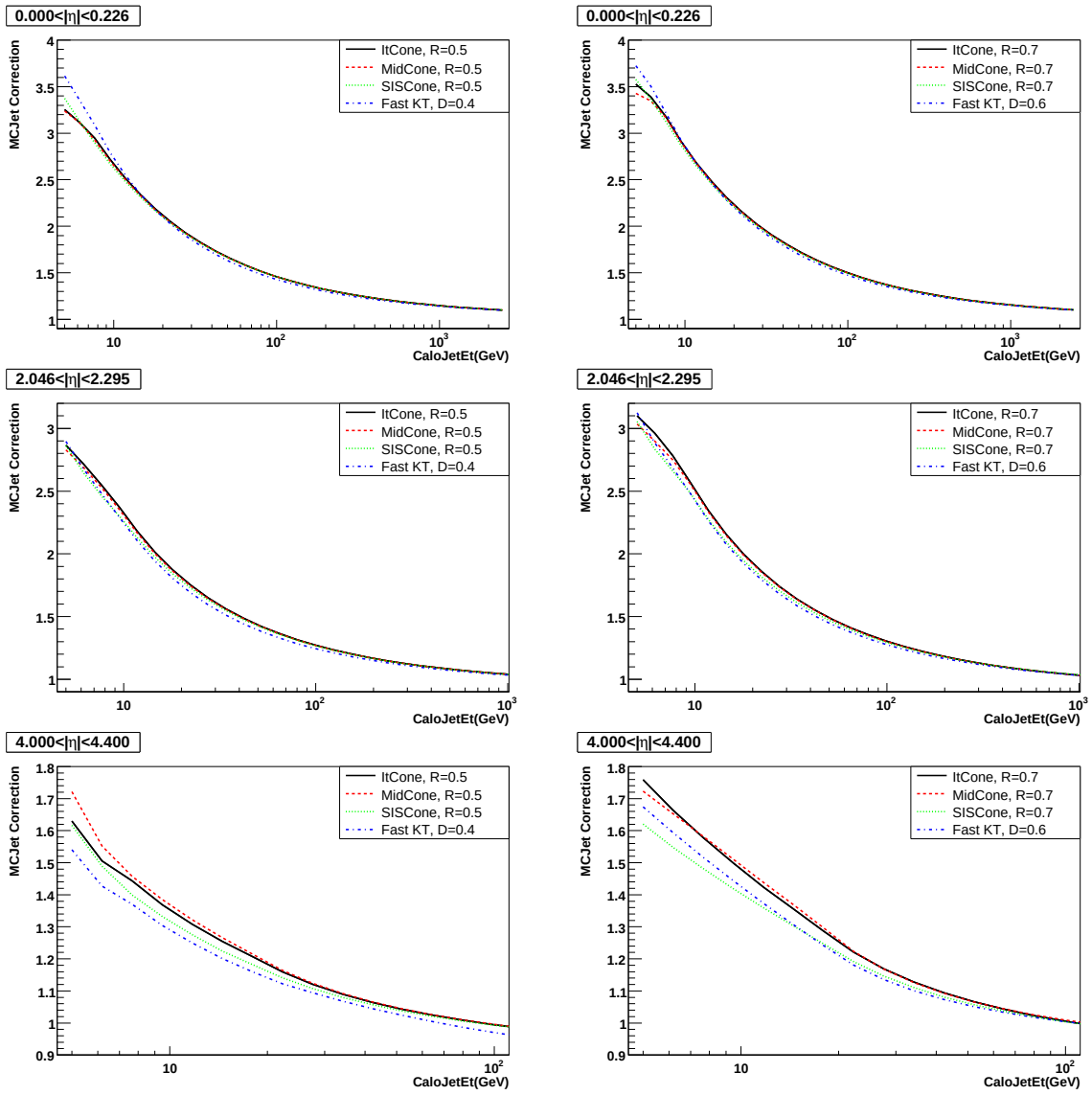


Figure 2: Jet energy correction factor as a function of $CaloJetE_T$ for jets reconstructed with smaller ($R = 0.5/D = 0.4$, left) and larger ($R = 0.7/D = 0.6$, right) radius parameter, for various $CaloJet|\eta|$ bins.

the pseudorapidity range of $1.4 < |\eta^{\rm gen}| < 3.0$ is used, and for the forward region the pseudorapidity range is $3.0 < |\eta^{\rm gen}| < 5.0$.

The histograms of the response have been fitted with a Gaussian function in the interval $\pm 1.5 \cdot$RMS centered around the mean. The mean parameters of the Gaussian fits have been extracted and used as data points to determine the average jet response as a function of $p_{\rm T}^{\rm gen}$. The results are shown in Figures 3, 4, and 5 for all eight algorithms considered. In general, the response curves for jets with similar size parameter (R or D) are comparable and not sensitive to the details of the clustering algorithm.

To test the performance of the MCJet corrections for the $p_{\rm T}^{\rm gen}$-based response, and the level of residual response effects between different algorithms, we present the analogous average response for the corrected jets in Figures 6, 7, and 8. The corrected response results are within a few percent of the desired value of 1.0 for jet $p_{\rm T}^{\rm gen} > 30$ GeV, but larger deviations are observed at lower $p_{\rm T}^{\rm gen}$.

4 Jet Position Resolutions

The resolutions of jet positions in ϕ and η have been calculated from distributions of the following variables for the matched jets:

$$\Delta\phi = \phi - \phi^{\rm gen}, \tag{7}$$

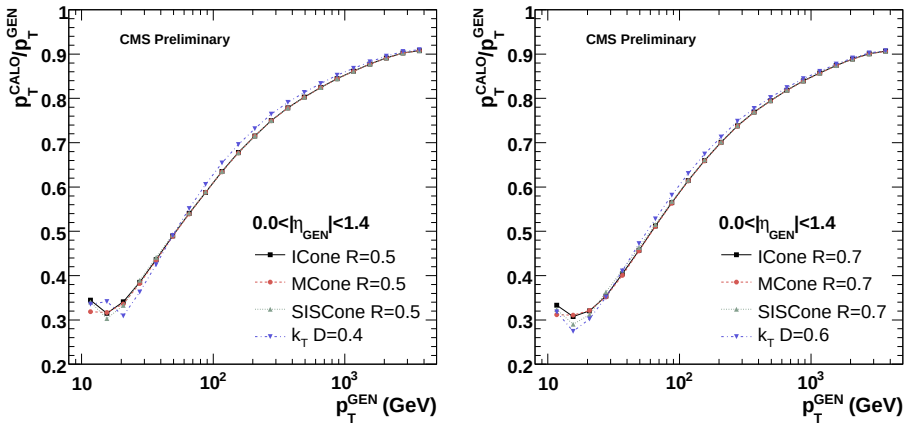


Figure 3: The jet response as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Barrel region, for jets clustered with smaller (left) and larger (right) size parameters.

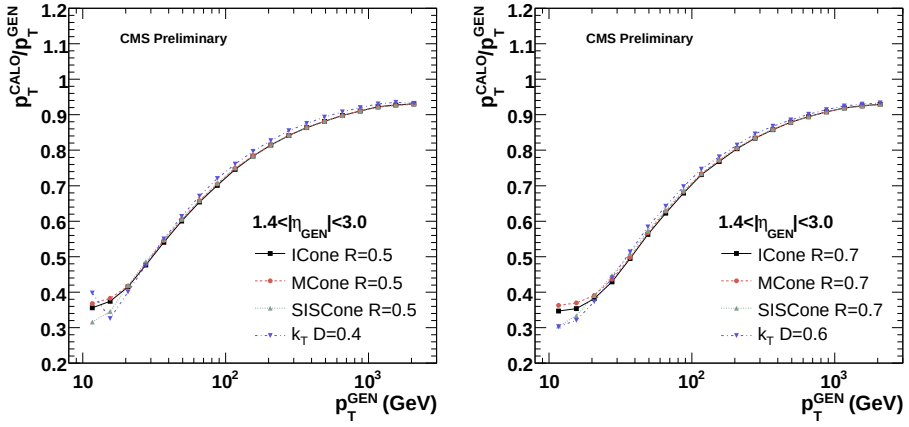


Figure 4: The jet response as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Endcap region, for jets clustered with smaller (left) and larger (right) size parameters.

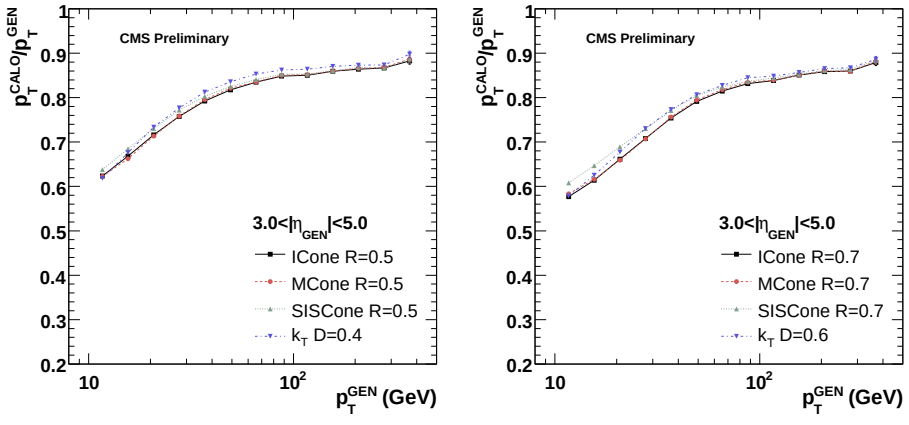


Figure 5: The jet response as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Forward region, for jets clustered with smaller (left) and larger (right) size parameters.

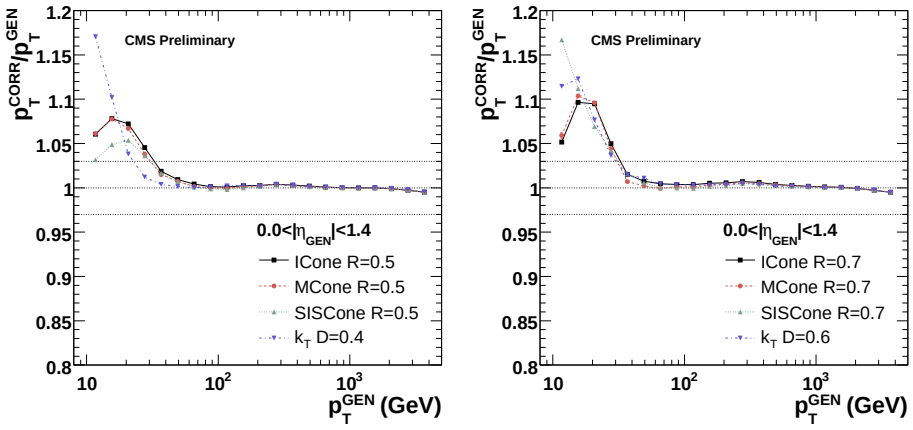


Figure 6: The MCJet-corrected jet response as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Barrel region, for jets clustered with smaller (left) and larger (right) size parameters. The dashed lines indicate the desired response of 1 as well as $\pm 3\%$ deviations.

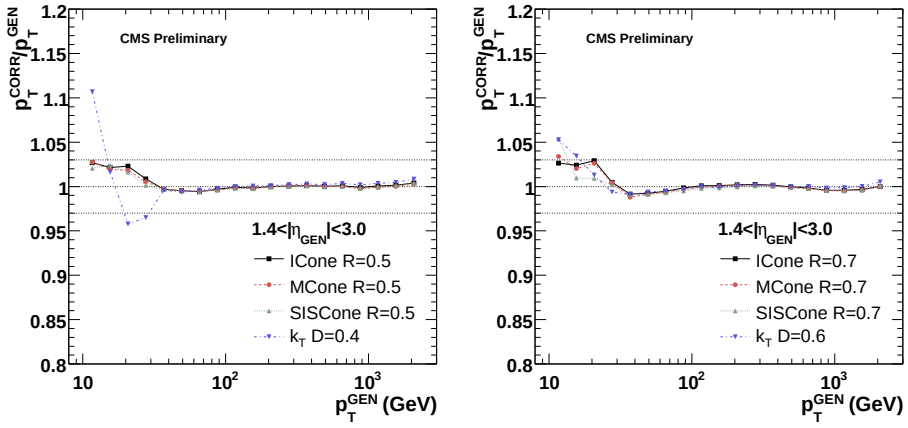


Figure 7: The MCJet-corrected jet response as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Endcap region, for jets clustered with smaller (left) and larger (right) size parameters. The dashed lines indicate the desired response of 1 as well as $\pm 3\%$ deviations.

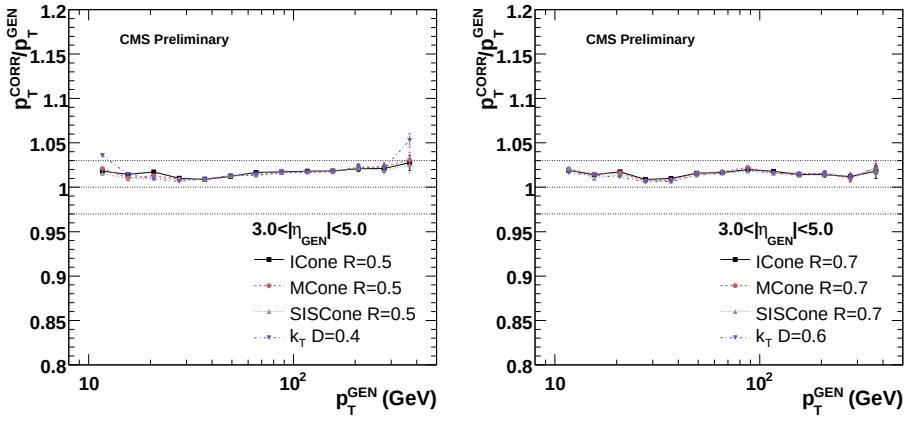


Figure 8: The MCJet-corrected jet response as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Forward region, for jets clustered with smaller (left) and larger (right) size parameters. The dashed lines indicate the desired response of 1 as well as $\pm 3\%$ deviations.

$$\Delta\eta = |\eta| - |\eta^{\rm gen}|, \tag{8}$$

and taking the $\pm\pi$ wrap-around in ϕ properly into account. We use the absolute values of generated and reconstructed jet η to eliminate the impact on the resolution from a reconstruction bias that pulls the calorimeter jet towards $\eta = 0$ relative to the generated jet. This effect visibly affects the results in HF.

Figures 9, 10, and 11 show the resolutions in ϕ for the three regions of the detector, as a function of $p_{\rm T}^{\rm gen}$. Similarly, Figures 12, 13, and 14 display the η resolutions.

The resolutions have been fitted with a function of the form

$$\sigma(p_{\rm T}^{\rm gen}) = \sqrt{\left(\frac{a}{p_{\rm T}^{\rm gen}}\right)^2 + \left(\frac{b}{\sqrt{p_{\rm T}^{\rm gen}}}\right)^2 + c^2}\,, \tag{9}$$

where a, b, and c are the fit parameters. This functional form describes poorly the data points in HF, which are rising as $p_{\rm T}^{\rm gen}$ increases. This behavior is interpreted as worsening of the resolution due to the large size of towers near the boundary between HE and HF. When $p_{\rm T}^{\rm gen}$ increases, the η distributions of jets become more central and the fraction of jets positioned in HF, but extending into HE, becomes greater, and so is the impact of these large-size towers on the positions of such jets.

Good agreement is found among all algorithms with comparable radius parameter, with marginal differences at low $p_{\rm T}^{\rm gen}$. Jets reconstructed with larger radius parameters yield slightly worse resolution both in η and ϕ. Note that the position of the primary vertex is assumed to be at $z = 0$, which dilutes the η resolution w.r.t. taking the correct position measured with the tracking detectors into account.

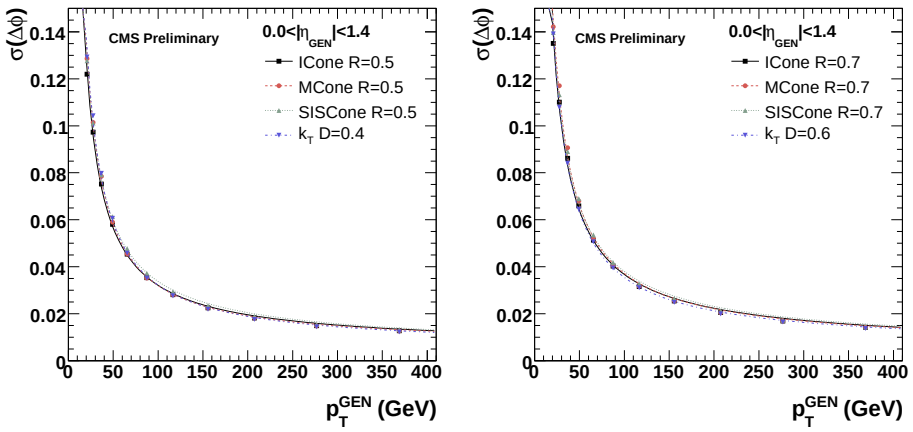


Figure 9: The jet ϕ resolutions as a function of $p_{\rm T}^{\rm gen}$, averaged over the Barrel region, for jets clustered with smaller (left) and larger (right) size parameters. The resolutions are derived using MC truth information.

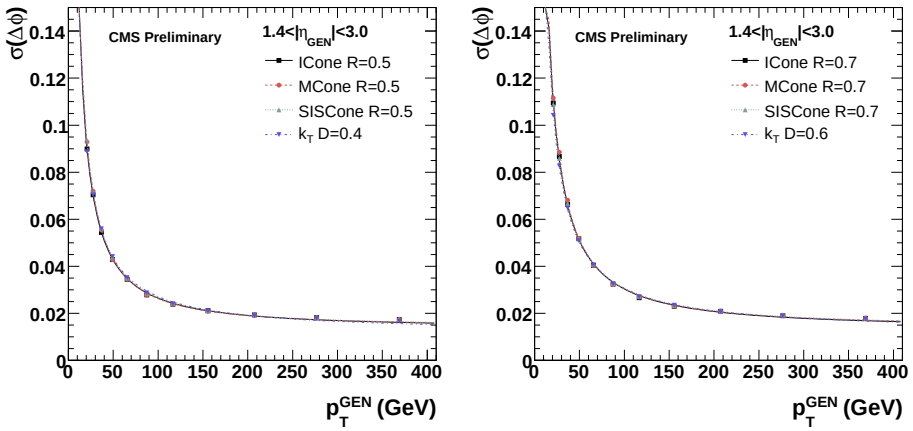


Figure 10: The jet ϕ resolutions as a function of $p_{\rm T}^{\rm gen}$, averaged over the Endcap region, for jets clustered with smaller (left) and larger (right) size parameters. The resolutions are derived using MC truth information.

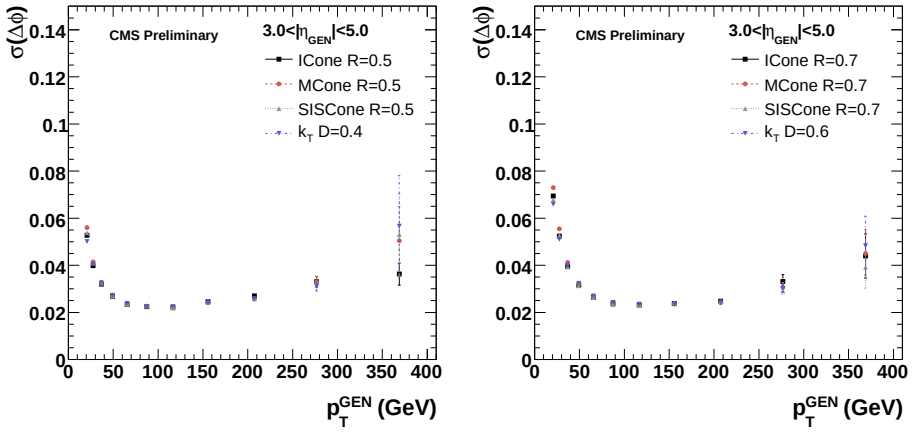


Figure 11: The jet ϕ resolutions as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Forward region, for jets clustered with smaller (left) and larger (right) size parameters. The resolutions are derived using MC truth information.

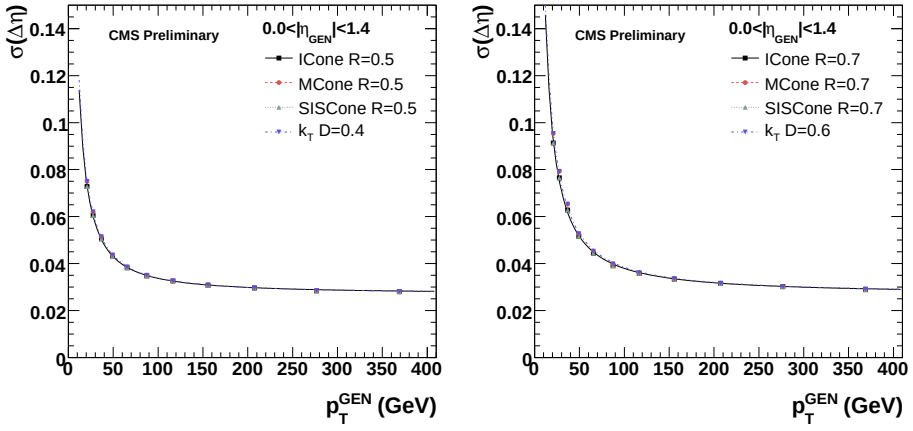


Figure 12: The jet η resolutions as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Barrel region, for jets clustered with smaller (left) and larger (right) size parameters. The resolutions are derived using MC truth information.

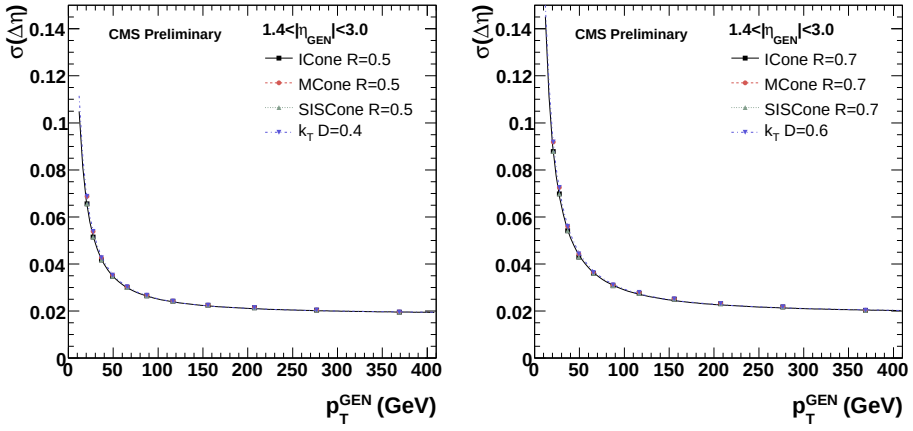


Figure 13: The jet η resolutions as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Endcap region, for jets clustered with smaller (left) and larger (right) size parameters. The resolutions are derived using MC truth information.

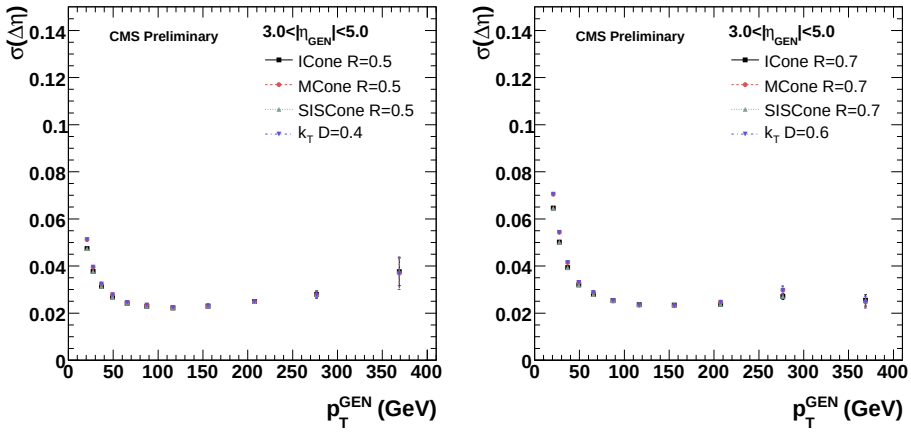


Figure 14: The jet η resolutions as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, averaged over the Forward region, for jets clustered with smaller (left) and larger (right) size parameters. The resolutions are derived using MC truth information.

5 Jet Matching Efficiencies and Mismatch Rates

The studies presented in this section are based on QCD dijet samples for $\hat{p}_T$ bins $15 - 300\,\mathrm{GeV}$ which are appropriately weighted according to the corresponding cross sections. For each particle jet the closest calorimeter jet is matched within $\Delta R < 0.5$. While no attempt was made to ensure that the same calorimeter jet is not assigned to multiple generator jets, it was shown in a separate study that this effect is insignificant for the presented results given the chosen ΔR requirement [8]. The matching efficiencies and mismatch rates presented here reflect primarily the consequences of jet position resolution on the matching procedure, and do not include effects due to leptons, high-multiplicity topologies, or instrumental effects not represented in the Monte Carlo. Figure 15 shows the resulting ΔR distributions for all algorithms. The distributions significantly differ for different regions of the detector in accordance to the position resolution presented in Section 4: The ΔR distribution tends to be narrower in the forward region than in the barrel and endcap. Figure 16 illustrates furthermore that particle jets with $p_{\mathrm{T}}^{\mathrm{gen}} > 60\,\mathrm{GeV}$ can be matched to a corresponding calorimeter jet within $\Delta R < 0.3$ with high efficiency, independent of the calorimeter region and algorithm used.

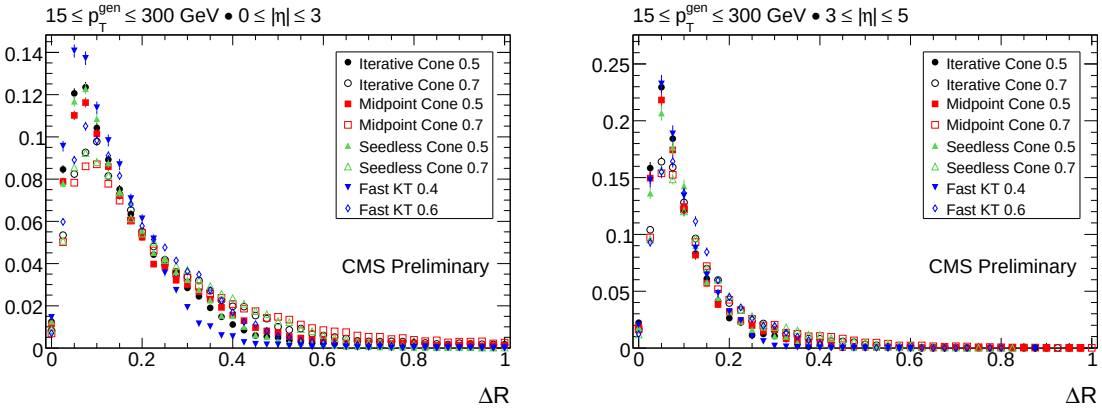


Figure 15: ΔR distributions for barrel and endcap region (left) and forward region (right).

The jet matching efficiency is defined as the ratio of the number of particle jets that match a calorimeter jet to the total number of particle jets. Figure 17 shows a comparison between algorithms with radius parameters $R = 0.5/D = 0.4$. The corresponding comparison for algorithms with $R = 0.7/D = 0.6$ is shown in Figure 18. The efficiency decreases for $p_{\mathrm{T}}^{\mathrm{gen}} < 40\,\mathrm{GeV}$, favoring smaller R/D parameters as well as SISCone and Fast k_{T} over Iterative Cone and Midpoint Cone. For particle jets with $p_{\mathrm{T}}^{\mathrm{gen}} = 10\,\mathrm{GeV}$ and $R = 0.5/D = 0.4$, SISCone and Fast k_{T} algorithms still yield efficiencies of around 90%, while Iterative Cone and Midpoint Cone drop to about 70%. The efficiencies as a function of η^{gen} are shown on the right side of Figure 17 and Figure 18 for $R = 0.5/D = 0.4$ and $R = 0.7/D = 0.6$ jets with $p_{\mathrm{T}}^{\mathrm{gen}} > 15\,\mathrm{GeV}$ respectively. Jet reconstruction is found to be fully efficient in the forward regions for all algorithms with $R = 0.5/D = 0.4$ and well above 95% across the entire calorimeter for SISCone and Fast k_{T}. Large variations are observed in the endcap ($1.4 < |\eta^{\mathrm{gen}}| < 3$) for Midpoint Cone and Iterative Cone jets, as well as for all algorithms with $R = 0.7/D = 0.6$. A separate study of

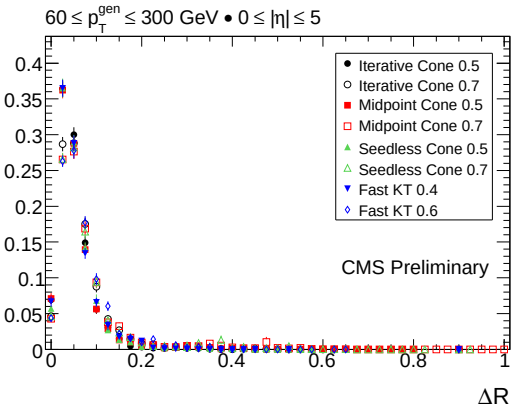


Figure 16: ΔR distribution for the entire calorimeter and $p_{\mathrm{T}}^{\mathrm{gen}} \geq 60\,\mathrm{GeV}$.

the seed thresholds, tower thresholds, and the magnetic field strength did not reveal the origin of this effect [9]. A comparison of SISCone and Fast k_{T} for small and large R/D parameters is shown in Figure 19.

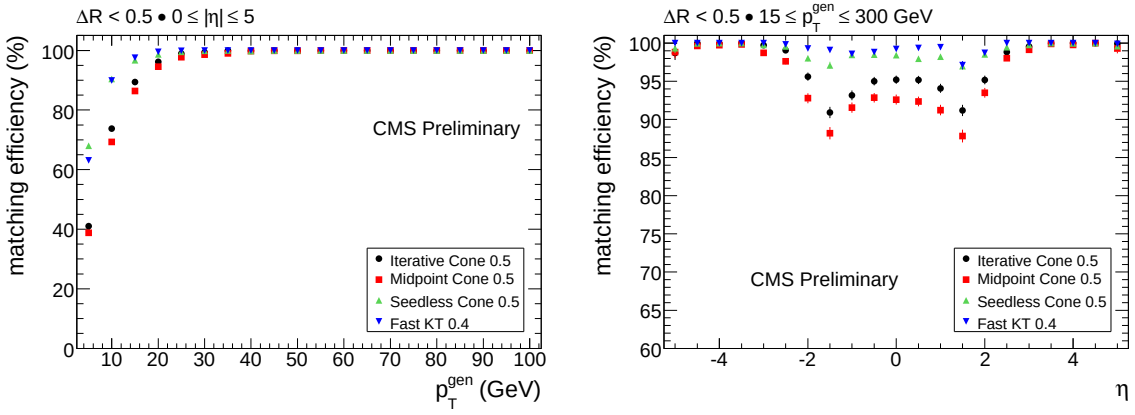


Figure 17: Matching efficiency vs $p_{\mathrm{T}}^{\mathrm{gen}}$ (left) and vs η^{gen} (right) for $R = 0.5/D = 0.4$ jets.

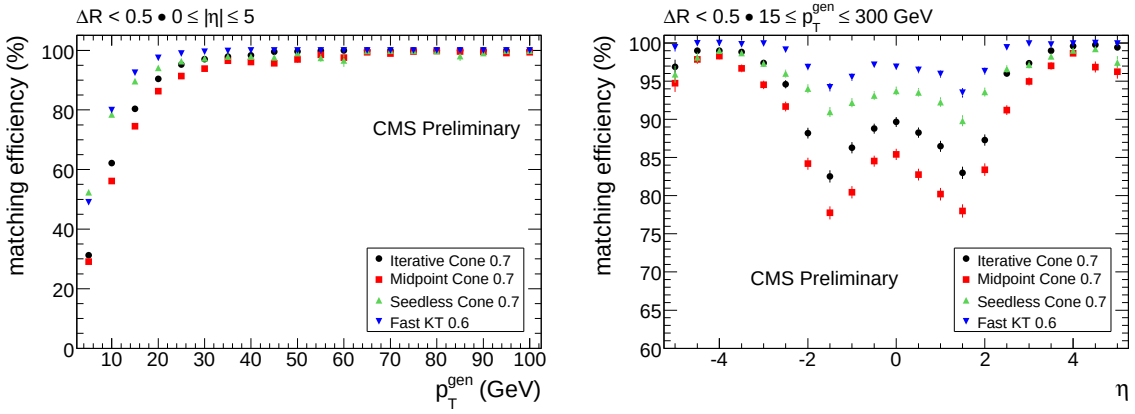


Figure 18: Matching efficiency vs $p_{\mathrm{T}}^{\mathrm{gen}}$ (left) and vs η^{gen} (right) for $R = 0.7/D = 0.6$ jets.

The jet mismatch rates are computed as the number of calorimeter jets which can not be matched within $\Delta R < 0.5$ to a particle jet divided by the total number of calorimeter jets. The MCJet energy corrections described in Section 2 are applied to the calorimeter jets, and the mismatch rates are plotted as a function of the corrected jet p_{T}. Mismatch rates for algorithms with $R = 0.5/D = 0.4$ and $R = 0.7/D = 0.6$ are compared separately in Figure 20 and Figure 21. For corrected calorimeter p_{T} below $50\,\mathrm{GeV}$ the mismatch rates start increasing such that at $10\,\mathrm{GeV}$ the mismatch rate varies between 5% and 25% depending on the algorithm. It is lower for $R = 0.5/D = 0.4$ than for $R = 0.7/D = 0.6$ algorithms. The mismatch rate as a function of η is shown in Figures 20 and 21 on the right for $R = 0.5/D = 0.4$ and $R = 0.7/D = 0.6$ respectively. It is negligible in the forward region and rises in the barrel

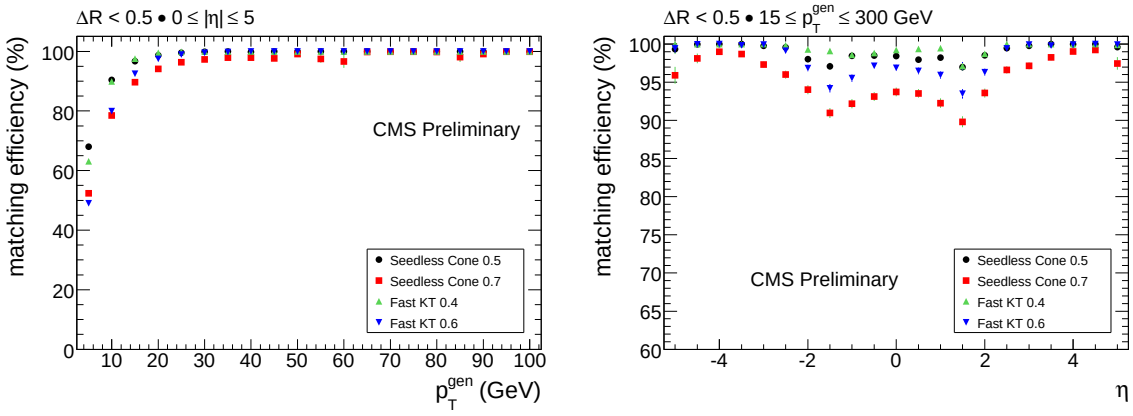


Figure 19: Matching efficiency vs $p_{\mathrm{T}}^{\mathrm{gen}}$ (left) and vs η^{gen} (right) for SISCone and Fast k_{T} jets.

and endcap region. A comparison between SISCone and Fast k_{T} mismatch rates is shown in Figure 22.

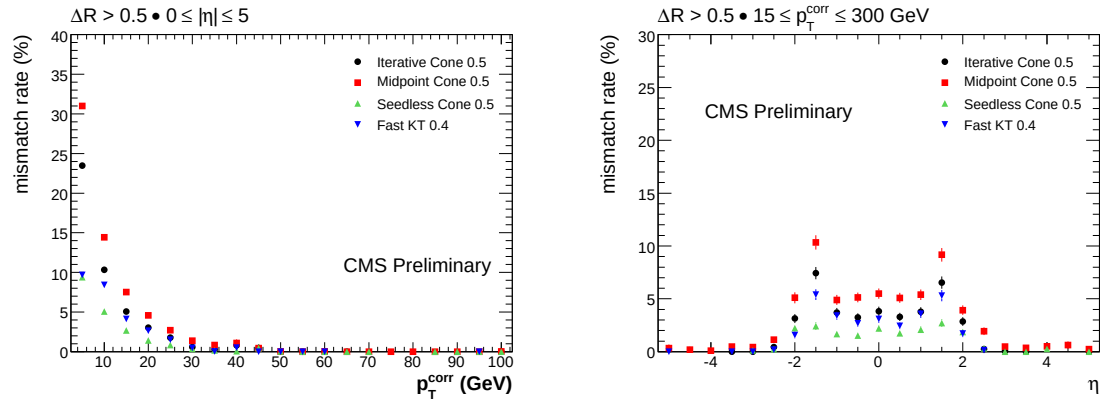


Figure 20: Mismatch rate vs corrected p_{T} (left) and vs η (right) for $R = 0.5/D = 0.4$ jets.

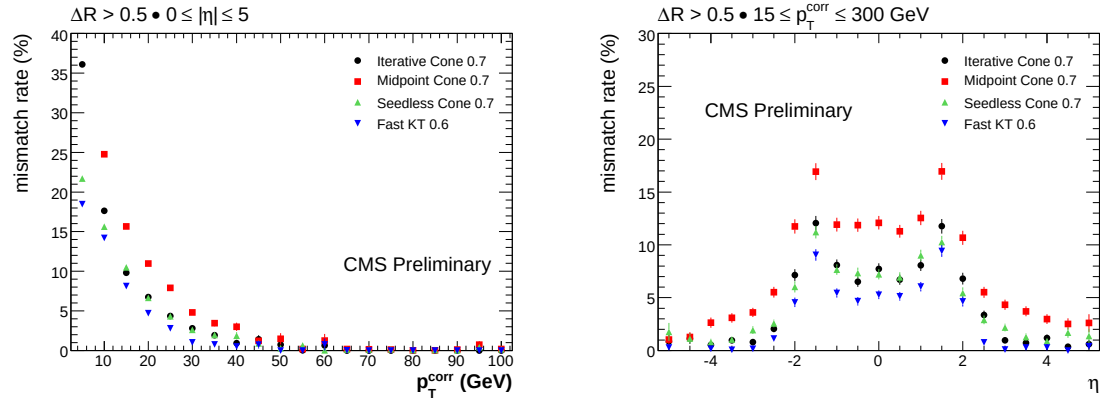


Figure 21: Mismatch rate vs corrected p_{T} (left) and vs η (right) for $R = 0.7/D = 0.6$ jets.

The p_{T} dependence of both matching efficiency and mismatch rate in the forward region is illustrated in Figure 23. Figure 24 shows the ΔR dependence of matching efficiency and mismatch rate. The results vary quite significantly with different ΔR cuts and the presented efficiencies and mismatch rates therefore need to be interpreted with care, taking into account the tight coupling of the definitions of both quantities to the p_{T} dependent position resolution.

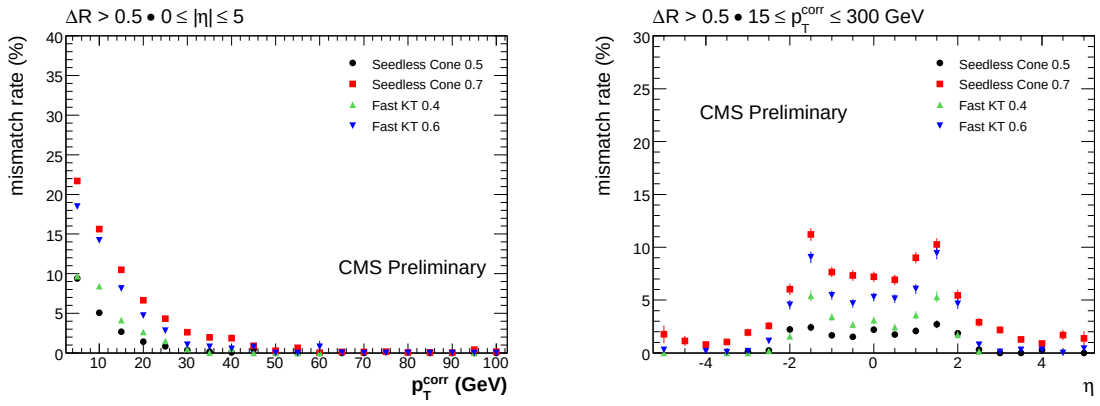


Figure 22: Mismatch rate vs corrected p_{T} (left) and vs η (right) for SISCone and Fast k_{T} jets.

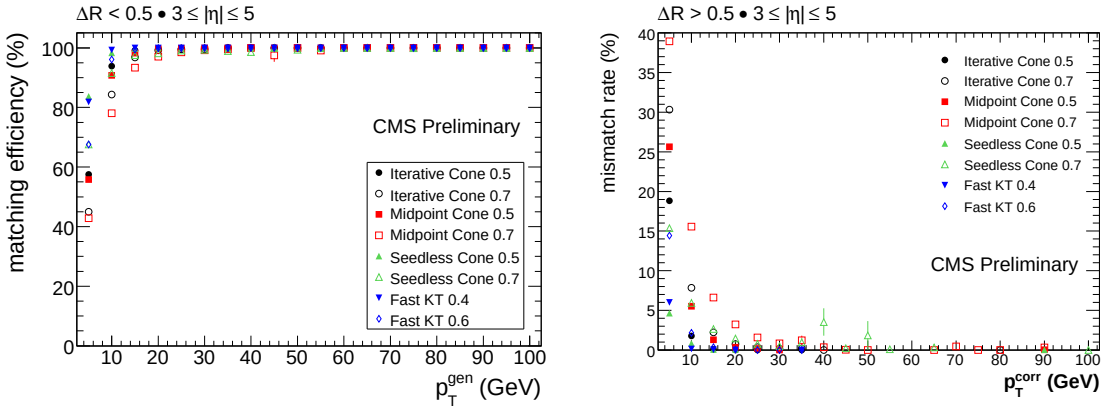


Figure 23: Matching efficiency (left) and mismatch rate (right) for the forward region.

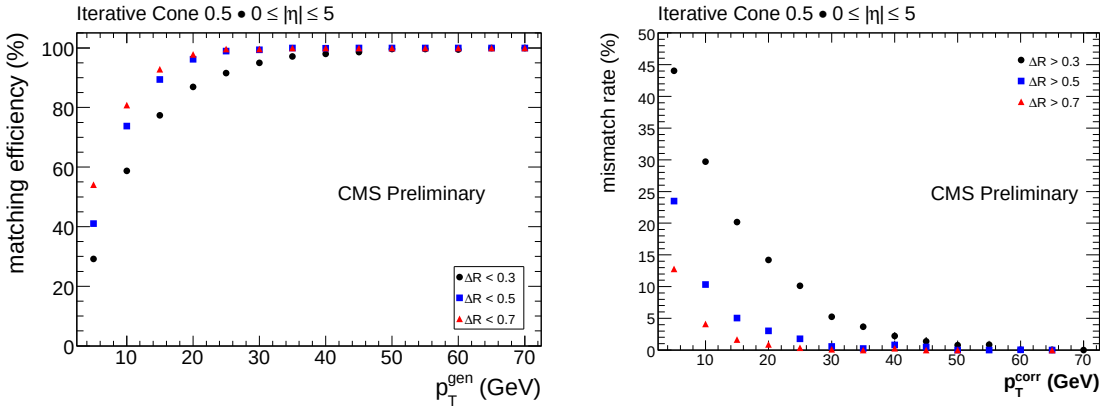


Figure 24: ΔR dependence of matching efficiency (left) and mismatch rate (right) for jets reconstructed with the Iterative Cone algorithm.

6 Jet Energy Resolution

6.1 Monte Carlo Truth Resolution

The Monte Carlo truth resolution is obtained from the $p_{\mathrm{T}}/p_{\mathrm{T}}^{\mathrm{gen}}$ distribution of corrected calorimeter jet p_{T} and particle-level $p_{\mathrm{T}}^{\mathrm{gen}}$. Only CaloJets matched to a GenJet within $\Delta R < 0.3$ are considered. The resolution is determined as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$, consistent with response and position resolution studies presented in Sections 3 and 4. The $p_{\mathrm{T}}/p_{\mathrm{T}}^{\mathrm{gen}}$ distributions are fitted with a Gaussian in the range of $2.5 \cdot \mathrm{RMS}$ around the mean value. The procedure is applied to several $p_{\mathrm{T}}^{\mathrm{gen}}$ bins and the fitted width is interpreted as the resolution for the mean $p_{\mathrm{T}}^{\mathrm{gen}}$ of each bin. The resolutions for Midpoint Cone, Iterative Cone, SISCone, and Fast k_{T} in the barrel region ($|\eta| < 1.4$) are presented in Figure 25 for radius parameters $R = 0.5/D = 0.4$ on the left and $R = 0.7/D = 0.6$ on the right. Similar comparisons are shown for the endcap ($1.4 < |\eta| < 3$) and forward regions ($3 < |\eta| < 5.2$) in Figures 26 and 27 respectively. The resolutions obtained for different jet algorithms are very similar for small and large radius parameters. In the high p_{T} region (above 100 GeV) they are almost identical, while small differences arise in the low p_{T} region. The resolution is worse in the forward region compared to both barrel and endcap for all algorithms.

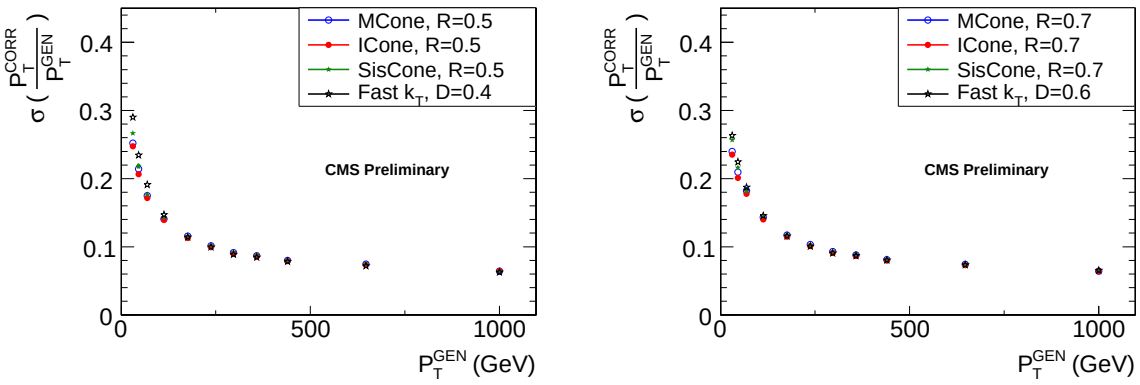


Figure 25: Jet energy resolution derived from MC truth for Midpoint Cone, Iterative Cone, SISCone, and Fast k_{T} with $R = 0.5/D = 0.4$ (left) and $R = 0.7/D = 0.6$ (right) in the barrel region ($|\eta| < 1.4$).

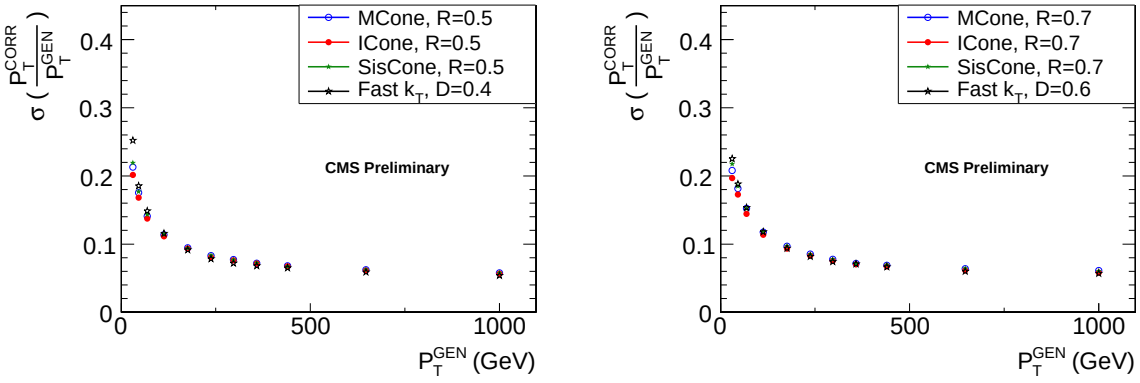


Figure 26: Jet energy resolution derived from MC truth for Midpoint Cone, Iterative Cone, SISCone, and Fast k_{T} with $R = 0.5/D = 0.4$ (left) and $R = 0.7/D = 0.6$ (right) in the endcap region ($1.4 < |\eta| < 3$).

6.2 Data-Driven Resolution Measurement

Data driven methods for the determination of the jet energy resolution are based on momentum conservation in the transverse plane. Events are selected for which the two leading jets are back-to-back ($\Delta\phi > 2.7$), to limit the influence of additional radiation in the event on the p_{T} balance of the leading jets. Both leading jets are required to be in the barrel ($|\eta| < 1.4$). The method can be applied to events with leading jets in rapidity regions other than the barrel, which is however not considered here.

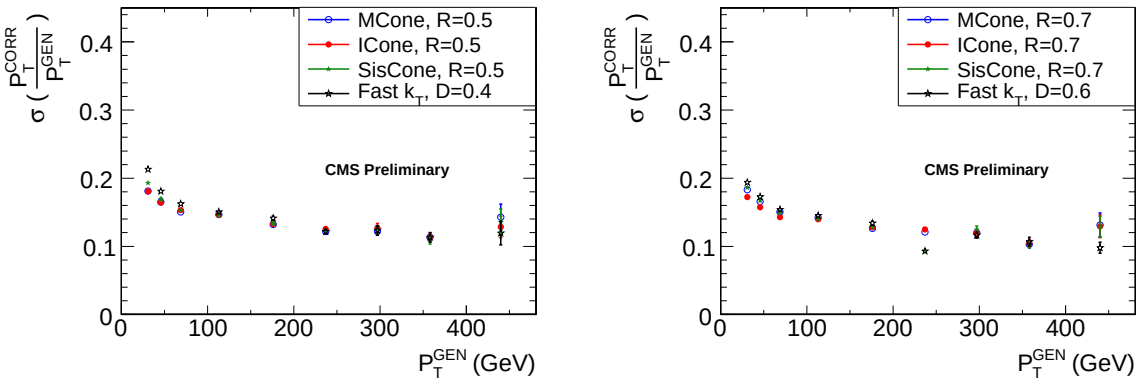


Figure 27: Jet energy resolution derived from MC truth for Midpoint Cone, Iterative Cone, SISCone, and Fast k_{T} with $R = 0.5/D = 0.4$ (left) and $R = 0.7/D = 0.6$ (right) in the forward region ($3 < |\eta| < 5.2$).

One such technique to obtain the jet energy resolution from data is the **Asymmetry Method** [10][11], which is based on the asymmetry variable A defined as

$$A = \frac{(p_{\mathrm{T}}^{jet1} - p_{\mathrm{T}}^{jet2})}{(p_{\mathrm{T}}^{jet1} + p_{\mathrm{T}}^{jet2})} \tag{10}$$

where p_{T}^{jet1} and p_{T}^{jet2} are the transverse momenta of the two leading jets in the event. The transverse momentum resolution can be expressed as a function of the variance of the asymmetry distribution:

$$\sigma_A^2 = \left|\frac{\partial A}{\partial p_{\mathrm{T}}^{jet1}}\right|^2 \sigma^2_{p_{\mathrm{T}}^{jet1}} + \left|\frac{\partial A}{\partial p_{\mathrm{T}}^{jet2}}\right|^2 \sigma^2_{p_{\mathrm{T}}^{jet2}} \tag{11}$$

With the assumption that $p_{\mathrm{T}} \equiv < p_{\mathrm{T}}^{jet1} > = < p_{\mathrm{T}}^{jet2} >$ and $\sigma_{p_{\mathrm{T}}} \equiv \sigma_{p_T}^{jet1} = \sigma_{p_{\mathrm{T}}}^{jet2}$ the relative transverse momentum resolution can be expressed as

$$\left(\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}}\right) = \sqrt{2}\sigma_A \, . \tag{12}$$

The resolution is obtained as a function of the average transverse momentum of the two leading jets from the width of the asymmetry distribution multiplied by $\sqrt{2}$. Most events in the QCD dijet sample however have additional reconstructed jets, primarily originating from soft radiation, which degrade the momentum balance in the transverse plane between the two leading jets. The resolution obtained from σ_A is therefore studied as a function of the maximum p_{T} of the third jet, and the final resolution is extracted by extrapolation of the p_{T} cut to zero. Examples of this procedure are shown in Figure 28 for Iterative Cone $R = 0.5$ (left) and Fast k_{T} $D = 0.6$ (right) jets respectively, for the average p_{T} bin $270 < p_{\mathrm{T}} < 330\,\mathrm{GeV}$. Note that the contribution from particle level imbalance is subtracted in quadrature, but it is so small that it can be neglected.

Figure 29 summarizes the results of the resolutions obtained from the Asymmetry Method and from MC truth for Iterative Cone $R = 0.5$ (left) and Fast k_{T} $D = 0.6$ (right) jets. The resolutions determined by the Asymmetry Method for Iterative Cone jets are in good agreement with the MC truth. However, for Fast k_{T} jets the Asymmetry Method underestimates the MC truth resolutions, especially at low p_{T}. Studies are underway to investigate the closure of the Asymmetry Method and examine sources of systematic uncertainty.

7 Jet Algorithm Performance in $t\bar{t}$ events

The Standard Model (SM) predicts the production of approximately one $t\bar{t}$ pair per second at the LHC, and each $t(\bar{t})$ subsequently to decay to $W^+b(W^-\bar{b})$. Events where either one ("lep+jets") or both ("alljets") W bosons decay into two quarks represent a good benchmark sample to study the performance of jet clustering algorithms in a high jet-multiplicity environment, yielding one two-jet (W) and one three-jet (t) mass resonance per hadronic W-decay. The efficiency to reconstruct all three quarks from the top decay as calorimeter jets and the resolution of the resulting m_W and m_t distributions are compared for several jet algorithms and levels of jet energy corrections.

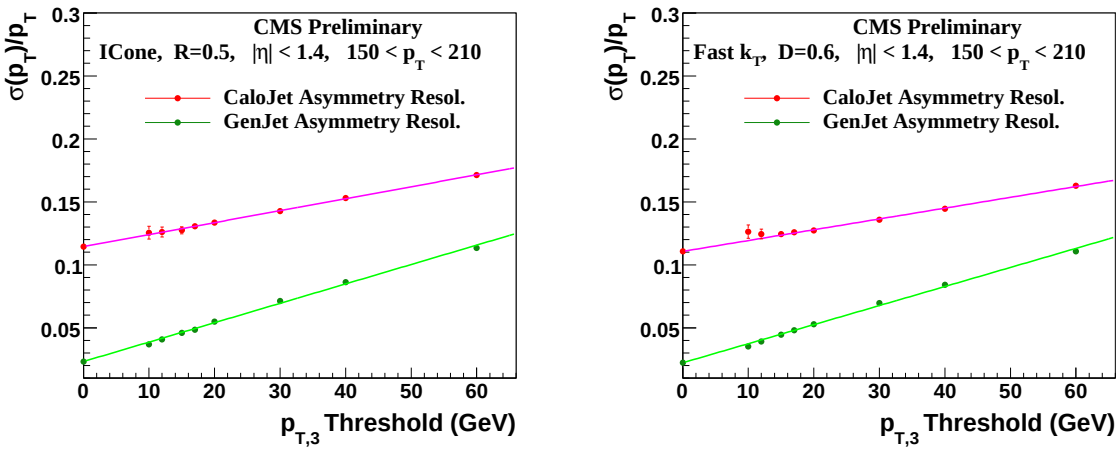


Figure 28: Resolution from dijet asymmetry as a function of the p_{T} threshold applied to the third jet in the event for Iterative Cone $R = 0.5$ (left) and Fast k_{T} $D = 0.6$, extrapolated to zero. The red and green lines correspond to detector and particle level respectively. The plots are taken from the average p_{T} bin with $150 < p_{\mathrm{T}} < 210\,\mathrm{GeV}$.

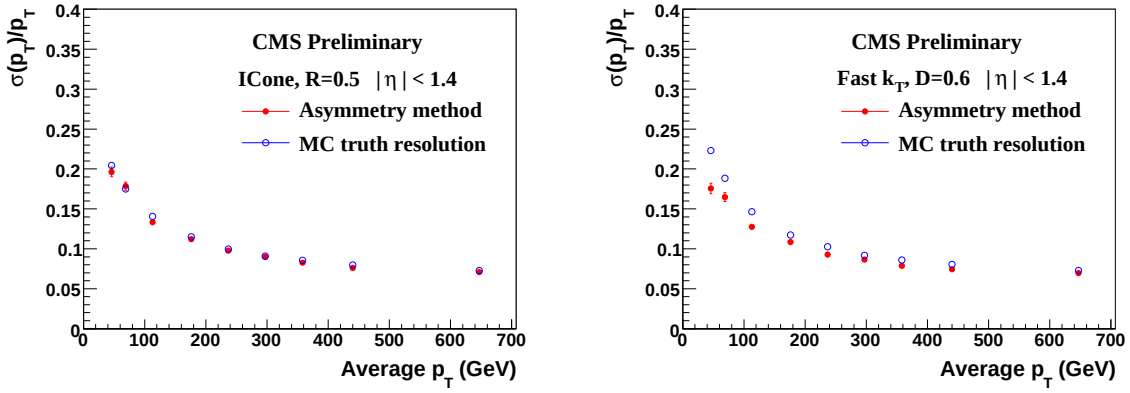


Figure 29: Resolution obtained with the Asymmetry Method and from Monte Carlo Truth for Iterative Cone $R = 0.5$ (left) and Fast k_{T} $D = 0.6$ (right) jets.

7.1 Event Selection

The study is based on ALPGEN $t\bar{t}$ MC events without additional jets ("$t\bar{t}+0jets$"), produced and reconstructed with CMSSW_1_5_2 without pileup. Based on MC truth information, events with either one or both W bosons decaying hadronically are selected, events with $\tau's$ in the final state are not considered. Four jet clustering algorithms are applied to these events, with three different parameter choices each:

- Fast k_{T}, $D = 0.3$, $D = 0.4$, and $D = 0.6$
- SISCone, $R = 0.4$, $R = 0.5$, and $R = 0.7$
- Midpoint Cone, $R = 0.4$, $R = 0.5$, and $R = 0.7$
- Iterative Cone, $R = 0.4$, $R = 0.5$, and $R = 0.7$

Calorimeter jets with uncorrected $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ are uniquely matched ($\Delta R < 0.5$) to generator jets passing the same p_{T} requirement. The flavor of each generator jet is determined by matching it within $\Delta R < 0.1$ to one of the partons (quarks) from the $t\bar{t}$ decay. The flavor of the calorimeter jet is considered to be the flavor of the matched generator jet. After matching procedures, only calorimeter jets with uncorrected $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| \leq 5$ are considered. Figure 30 shows the jet multiplicity, p_{T}, and η distributions for selected jets reconstructed with the Fast k_{T} algorithm ($D = 0.4$). The distributions look similar for other algorithms. Note that isolated electrons in lepton+jets events are typically reconstructed as jets as well, and no attempt is being made to remove them from

the list of jets. For p_{T} and η, the distributions for jets which are matched to one of the top decay partons are also shown, and are not significantly biased.

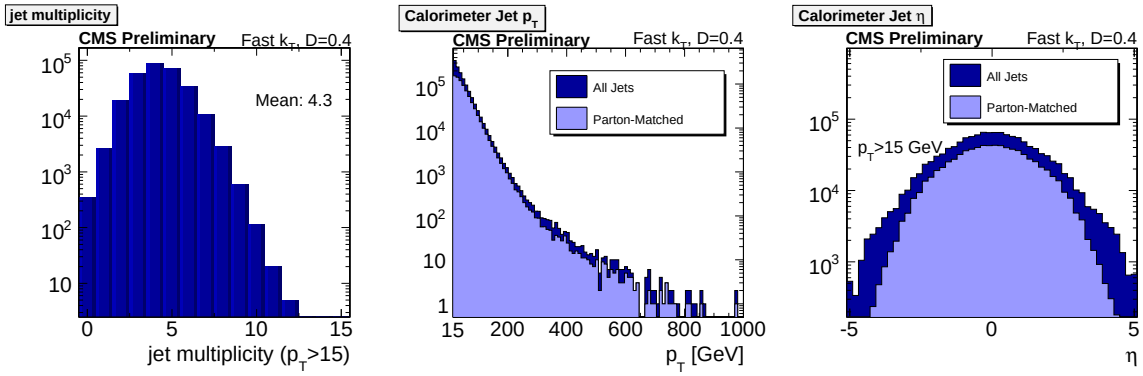


Figure 30: Jet multiplicity, p_{T}, and η for calorimeter jets reconstructed with the Fast k_{T} algorithm, $D = 0.4$. A preselection of uncorrected $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| \leq 5$ is applied to the jets. The light-blue distributions are for calorimeter jets which are matched to partons from the hadronic top decay in the MC truth record.

7.2 Top Selection Efficiency

The efficiencies to reconstruct all decay products of the hadronic top decay as calorimeter jets w.r.t. the number of events with at least four (six) reconstructed jets for lepton+jets (alljets) events are compared for all algorithms. For alljets events, the efficiency to reconstruct both top decays is considered as well. The absolute numbers of selected events and reconstructed decays are illustrated in Figure 31 for lepton+jets (top) and alljets (bottom) events. The relative efficiencies are summarized in Tables 2 and 3 for lepton+jets and alljets events respectively. For all cone-based algorithms, the efficiencies to reconstruct four (six) jets and to reconstruct hadronic top decays in lepton+jets (alljets) events decrease with increasing cone-size parameter R, as expected: The stable products from individual partons which are close in η-ϕ space are more likely to be clustered into one single jet with increasing R. The opposite behavior is observed for jets reconstructed with the Fast k_{T} algorithm however: Both efficiencies are increasing with the D parameter, indicating that the energy of the partons is distributed over too many individual jets with too little energy to fulfill the $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ requirement, if the parameter is chosen too small. Specifically, there seems to be no gain in selection efficiency for $D < 0.4$. The jet-selection and top-reconstruction efficiency for events clustered with Fast k_{T} and $D \geq 0.4$ is slightly better than for all cone-based algorithms.

Algorithm	Fast k_{T}			SISCone			Midpoint Cone			Iterative Cone		
D/R	0.3	0.4	0.6	0.4	0.5	0.7	0.4	0.5	0.7	0.4	0.5	0.7
$N(jet) \geq 4$ [%]	53.7	59.5	64.3	58.7	59.9	57.2	59.6	60.8	58.1	59.2	61.6	62.3
$N(t \to bq\bar{q}')$ [%]	17.1	18.4	17.6	17.5	15.8	9.3	17.7	16.0	9.4	18.9	18.5	14.3

Table 2: Efficiency to reconstruct at least four jets with uncorrected $p_{\mathrm{T}} > 15\,\mathrm{GeV}$, and efficiency to reconstruct all three partons from the $t \to bq\bar{q}'$ decay as calorimeter jets in **lepton+jets** events. The second efficiency is calculated w.r.t. the number of events with four or more jets.

Algorithm	Fast k_{T}			SISCone			Midpoint Cone			Iterative Cone		
D/R	0.3	0.4	0.6	0.4	0.5	0.7	0.4	0.5	0.7	0.4	0.5	0.7
$N(jet) \geq 6$ [%]	21.6	25.6	28.9	23.6	23.8	19.9	24.3	24.4	20.4	24.2	25.6	24.5
$N(t \to bq\bar{q}')$ [%]	42.1	42.0	38.8	43.6	40.2	28.3	43.6	40.4	28.3	45.5	43.9	36.3
$N(t\bar{t} \to b\bar{b}q\bar{q}'q\bar{q}')$ [%]	4.3	4.9	4.8	4.9	4.4	2.4	5.0	4.5	2.5	5.4	5.4	4.1

Table 3: Efficiency to reconstruct at least six jets with uncorrected $p_{\mathrm{T}} > 15\,\mathrm{GeV}$, efficiency to reconstruct all three partons from any $t \to bq\bar{q}'$ decay as calorimeter jets, and efficiency to fully reconstruct both top decays in **alljets** events. The second and third efficiencies are given w.r.t. the number of events with six or more jets.

7.3 W Boson and Top Quark Mass Resolution

To evaluate the m_W and m_t resolution for different algorithms, the two- and three jet masses are formed from the calorimeter jets matched to the partons from $t \to bq\bar{q}'$ decays. The resulting mass spectra peak at values

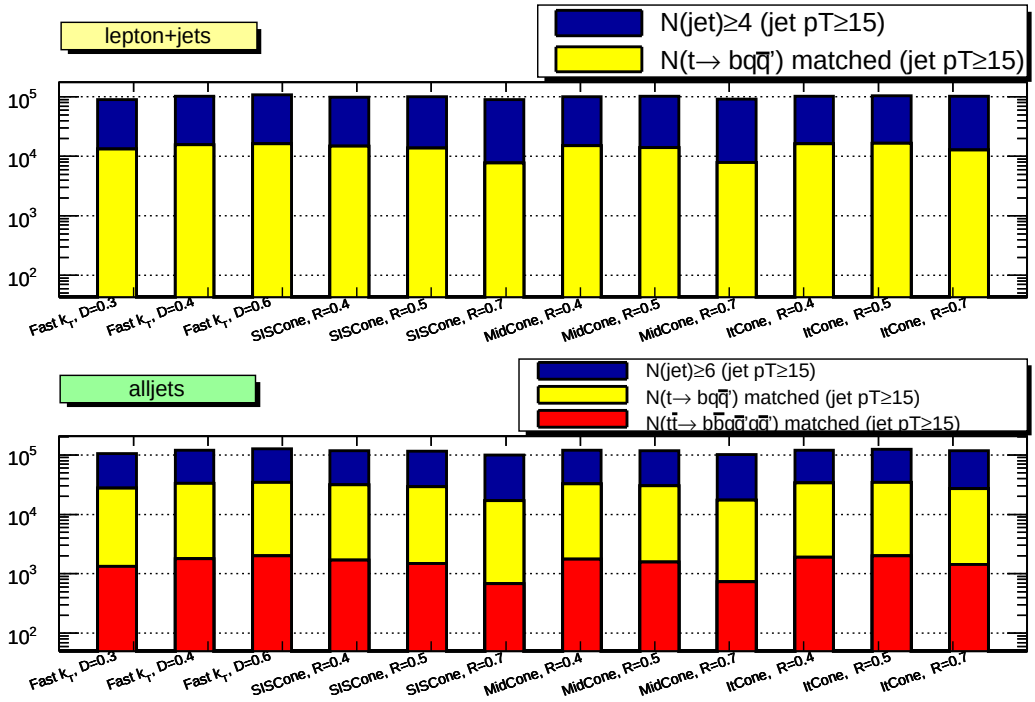


Figure 31: Number of hadronic top decays for which all three jets are reconstructed, per jet algorithm. The number of events with at least four (lepton+jets) and six (alljets) reconstructed jets with $p_{\mathrm{T}} \geq 15\,\mathrm{GeV}$ (uncorrected) is shown as well. For alljets events, the number of events with both top decays being fully reconstructed is included additionally.

considerably lower than the generated W boson ($80.42\,\mathrm{GeV}$) and Top Quark ($175\,\mathrm{GeV}$) masses if raw calorimeter jets are considered ("CALO"). After application of the official MCJet corrections to all calorimeter jets ("CORR"), the mass spectra are shifted slightly higher than the generated mass values: as the MCJet corrections are obtained from a QCD sample dominated by jets originating from gluons, additional flavor-dependent corrections must be applied to jets originating from u/d/s, c, and b quarks separately. In reference to the proposed factorized multi-level correction scheme for CMS jets [7], these flavor-dependent corrections are referred to as "Level 5" corrections ("L5"). Since no official MCJet corrections exist for $D = 0.3$ and $R = 0.4$, the respective corrections for $D = 0.4$ and $R = 0.5$ are applied to these jets. The effect of jet corrections for Fast k_{T}, $D = 0.4$ jets is illustrated in Figure 32 for the m_W (left) and m_t (right) mass spectra. The spectra using generator jets (GEN) are also shown.

The mean of the distributions for all jet algorithms under study is shown in Figure 33 for m_W (top) and m_t (bottom): after the application of both MCJet and flavor-dependent Level 5 corrections, the true W boson and Top Quark masses are reasonably well reproduced. Note that the dependence of the W and top masses as a function of the radius parameter on generator level (green stars in Figure 33) is consistent with the claim [13] of quadratic dependence of $p_{\mathrm{T}}^{\mathrm{gen}}$ due to the underlying event. The relative width ($\mathrm{RMS}(m)/m$) of these distributions yields a good measure of the mass resolution performance of each algorithms, and Figure 34 depicts the comparisons for m_W (top) and m_t (bottom). The resolutions are in good agreement for all algorithms and radius parameters, with the exception that the mass resolution improves slightly for jets reconstructed with Fast k_{T} when choosing larger radius parameters D.

8 Energy Distribution within a Jet

In this section, we compare the internal properties of the calorimeter and particle jets reconstructed with different algorithms. This study is performed using CMSSW_1_5_2 QCD dijet events. Different $\hat{p}_T$ samples are combined using appropriate weights. The jet collections for Iterative Cone ($R = 0.5, 0.7$), Midpoint Cone ($R = 0.5, 0.7$), and Fast k_{T} ($D = 0.6$) clustering algorithms are taken from official production. The SISCone jets are reconstructed using CMSSW_1_6_0.

Only events where the two leading jets are within $|y| < 1.0$ are used, where y is the rapidity of the jet. In addition, we use only those calorimeter jets which are within $\Delta R < 0.3$ of a particle jet (GenJet), to ensure correspondence between calorimeter and particle level distributions. For all studies in this section, we use only the particles (towers)

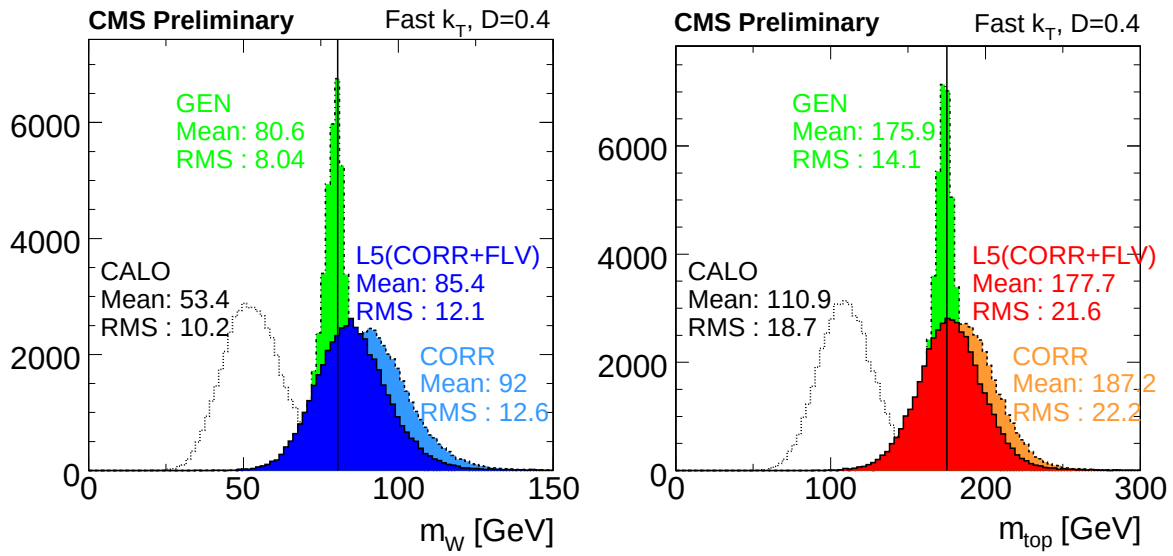


Figure 32: m_W and m_t distributions for hadronic top decays reconstructed with the Fast k_{T} algorithm, $D = 0.4$. Distributions are shown for particle-level jets (GEN), calorimeter jets (CALO), calorimeter jets corrected with MCJet corrections (CORR), and corrected calorimeter jets with an additional flavor correction ("Level-5 correction") applied (L5). Only jets with uncorrected $p_{\mathrm{T}} \geq 15\,\mathrm{GeV}$ and $|\eta| \leq 5$ are considered. The generated W boson ($80.42\,\mathrm{GeV}$) and Top Quark ($175\,\mathrm{GeV}$) masses are indicated by the black vertical lines.

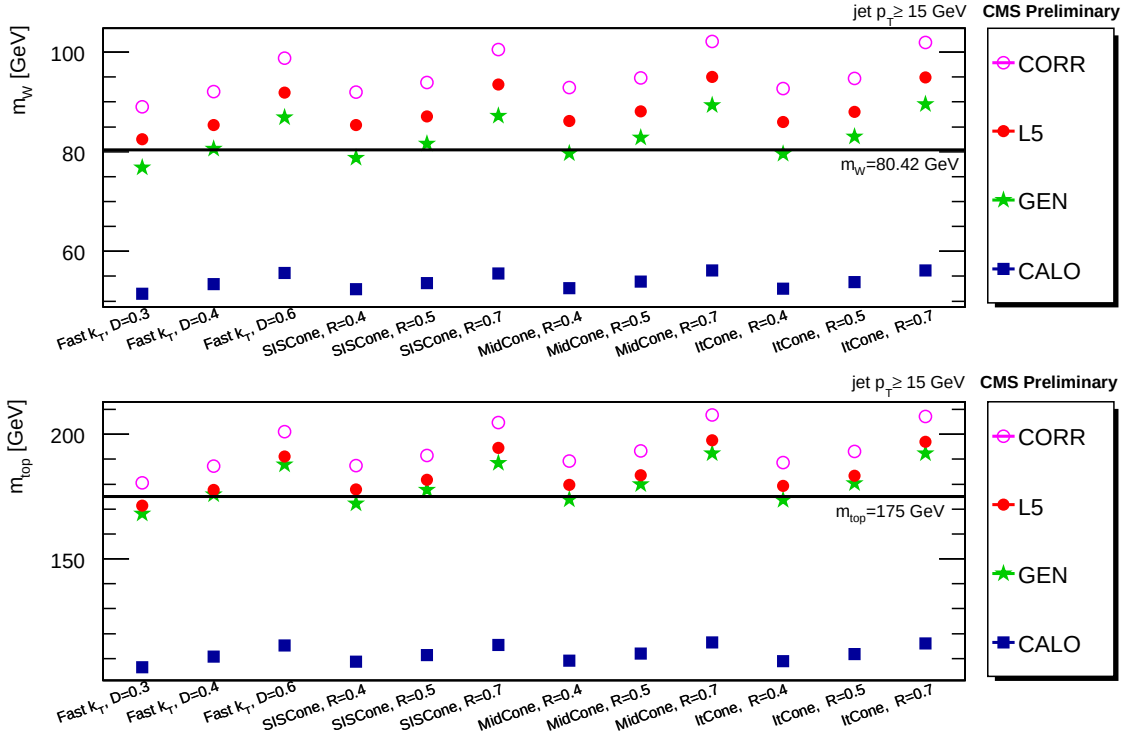


Figure 33: Mean of m_W (top) and m_t (bottom) distributions for all studied jet algorithms for particle-level jets (GEN), calorimeter jets (CALO), corrected calorimeter jets (CORR), and corrected calorimeter jets with an additional flavor correction applied (L5). Only jets with uncorrected $p_{\mathrm{T}} \geq 15\,\mathrm{GeV}$ and $|\eta| \leq 5$ are considered.

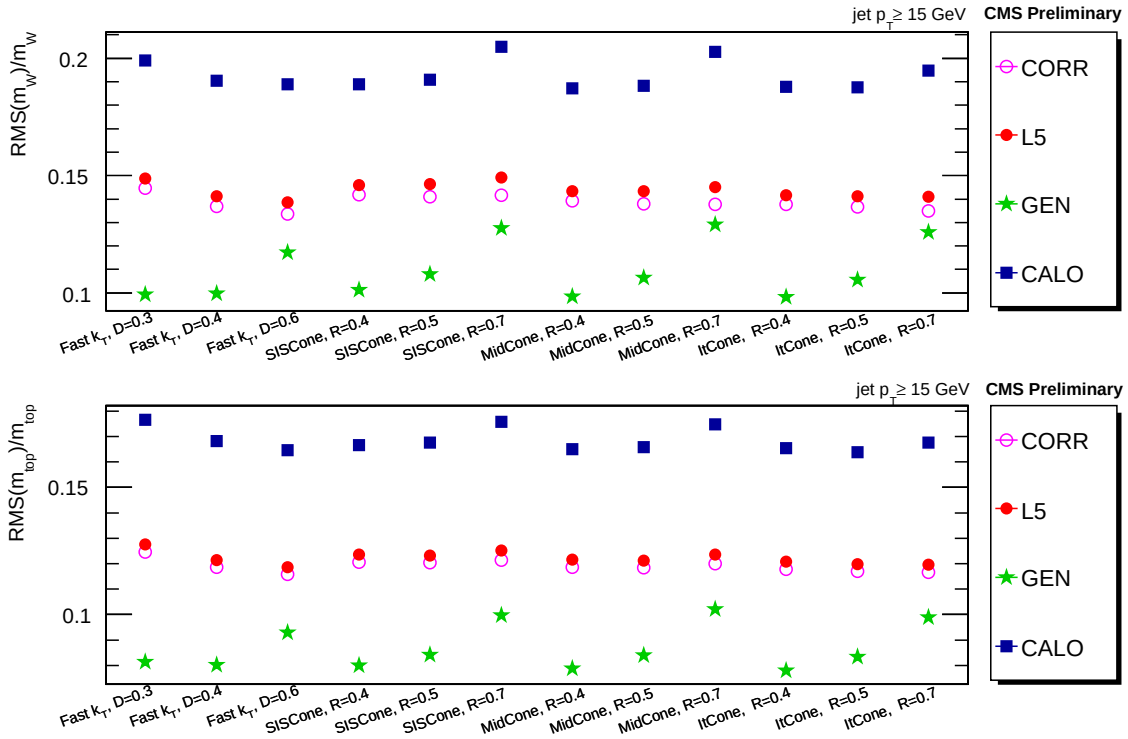


Figure 34: Relative width ($\mathrm{RMS}(m)/m$) of m_W (top) and m_t (bottom) distributions for all studied jet algorithms for particle-level jets (GEN), calorimeter jets (CALO), corrected calorimeter jets (CORR), and corrected calorimeter jets with an additional flavor correction applied (L5). Only jets with uncorrected $p_{\mathrm{T}} \geq 15\,\mathrm{GeV}$ and $|\eta| \leq 5$ are considered.

which are associated with the jet by the clustering algorithm.

Multiplicity and p_{T} distributions for particles (towers) in particle (calorimeter) jets are shown in Figures 35 and 36 (37 and 38). As expected, both the particle and tower mean multiplicities in a jet increase logarithmically with jet p_{T}. The p_{T} distribution of particles and towers becomes harder with increasing jet p_{T}.

The internal structure of the jet can also be described by the energy distribution within a jet which is characterized by the differential jet shape, $\rho(r)$, defined as

$$\rho(r) = \frac{\sum p_{\mathrm{T}}(r - \Delta r/2, r + \Delta r/2)}{\Delta r \sum p_{\mathrm{T}}^{Jet}}$$

where the sum is over all the jet constituents in the range $(r - \Delta r/2, r + \Delta r/2)$ in the numerator and $r = \sqrt{(y_{jet} - y_c)^2 + (\phi_{jet} - \phi_c)^2}$ with (y_{jet}, ϕ_{jet}) and (y_c, ϕ_c) being the position of the jet and the constituents. The denominator $\sum p_{\mathrm{T}}^{Jet}$ is the scalar sum of the transverse momenta of all the jet constituents. The jetshapes for different particle and calorimeter jets are shown in Figures 39 and 40 respectively. Jets become narrower with increasing jet p_{T}. Note that the Iterative Cone algorithm is based on $\Delta R(\eta)$, while the differential jet energy density is defined using $\Delta r(y)$, explaining contributions to the density for $r > R$.

The cone clustering algorithms using the same radius have similar properties. The Iterative Cone and Midpoint Cone algorithms mainly differ due to the splitting-merging stage. This difference has a small effect on the two leading jets in the sample used which is dominated by back-to-back jets. The SISCone algorithm finds all stable cones. The two leading jets produced by splitting and merging of stable cone are expected to be similar to the jets found by the Midpoint Cone algorithm.

The k_{T} clustering algorithm with $D = 0.6$ has similar characteristics as the cone jets with $R = 0.7$ except for jets in the $30 < p_{\mathrm{T}} < 50\,\mathrm{GeV}$ range where k_{T} jets are slightly narrower.

9 Cone Algorithm Considerations: SISCone and Midpoint Cone

At present, two modern cone clustering algorithms are available at CMS: SISCone and Midpoint Cone. In contrast to the Iterative Cone algorithm, they both implement the splitting-merging step to address the reconstruction of

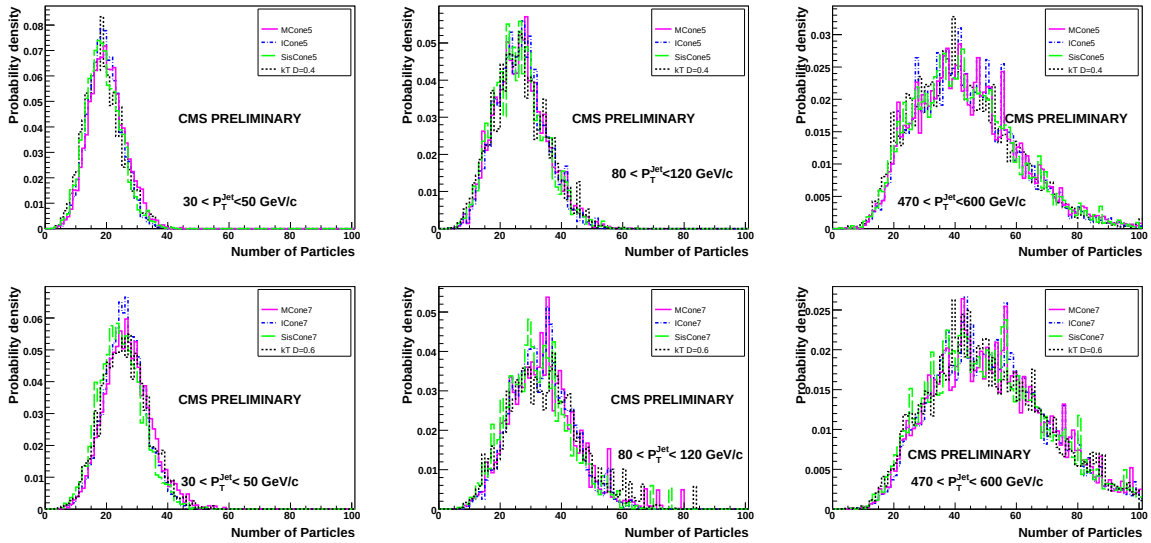


Figure 35: The multiplicity of particles in a jet for different algorithms. Different cone algorithms with the same clustering radius have almost the same multiplicity. The multiplicity for Fast k_{T} $D = 0.6$ jets is close to the observation for cone jets with $R = 0.7$.

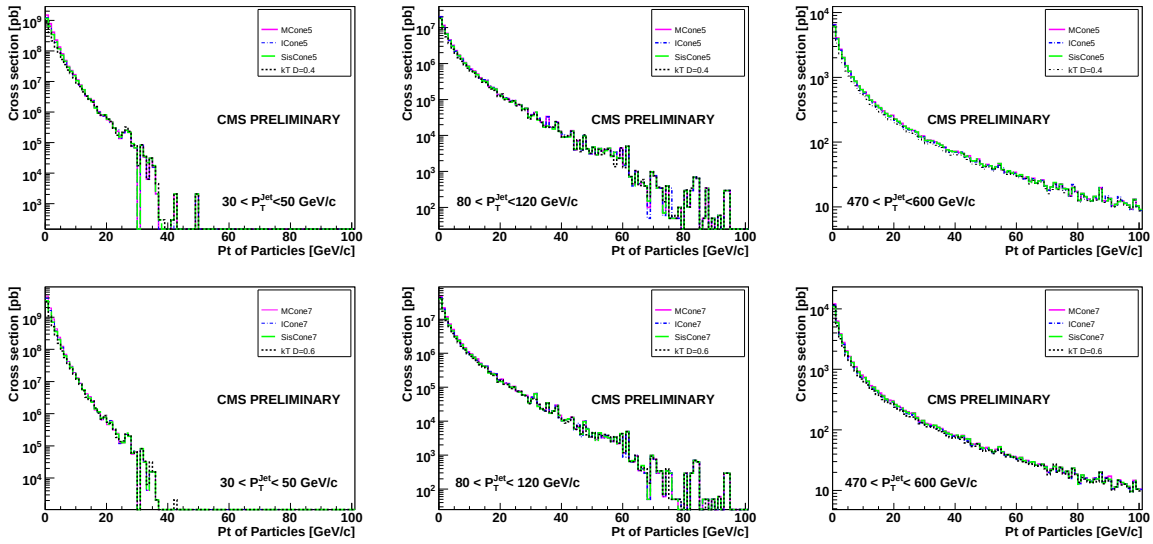


Figure 36: The transverse momentum of particles clustered in to a jet for different jet clustering algorithms. The p_{T} distribution for Fast k_{T} $D = 0.6$ jets is close to the one observed for cone jets with $R = 0.7$.

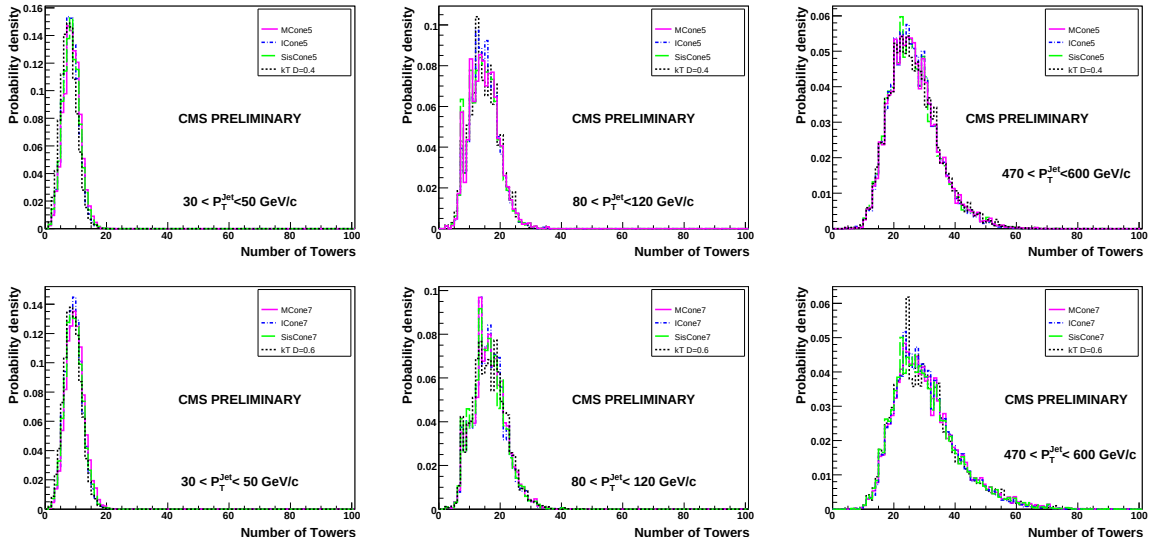


Figure 37: The multiplicity of towers in jets reconstructed using different algorithms. Different cone algorithms with the same clustering radius have almost the same multiplicity. The multiplicity of Fast k_{T} $D = 0.6$ jets is close to the one observed for cone jets with $R = 0.7$. Plots are labeled with corresponding particle jet p_{T} bins.

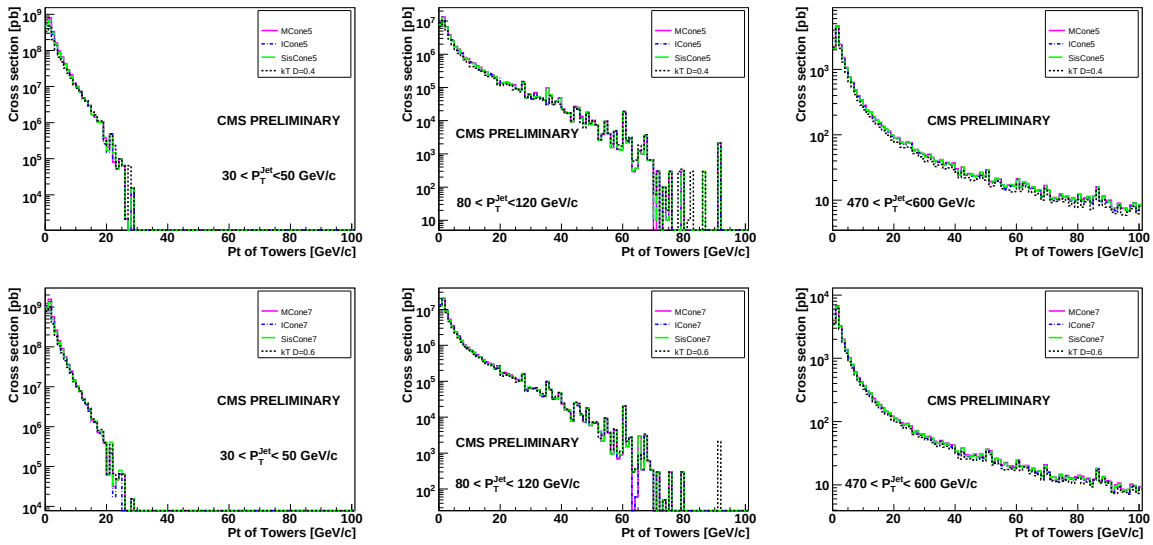


Figure 38: The p_{T} of towers in jets reconstructed using different algorithms. Different cone algorithms with the same clustering radius have almost the same p_{T} distributions. The p_{T} of towers clustered into Fast k_{T} $D = 0.6$ jets is close to the observation for cone jets with $R = 0.7$. Plots are labeled with corresponding particle jet p_{T} bins.

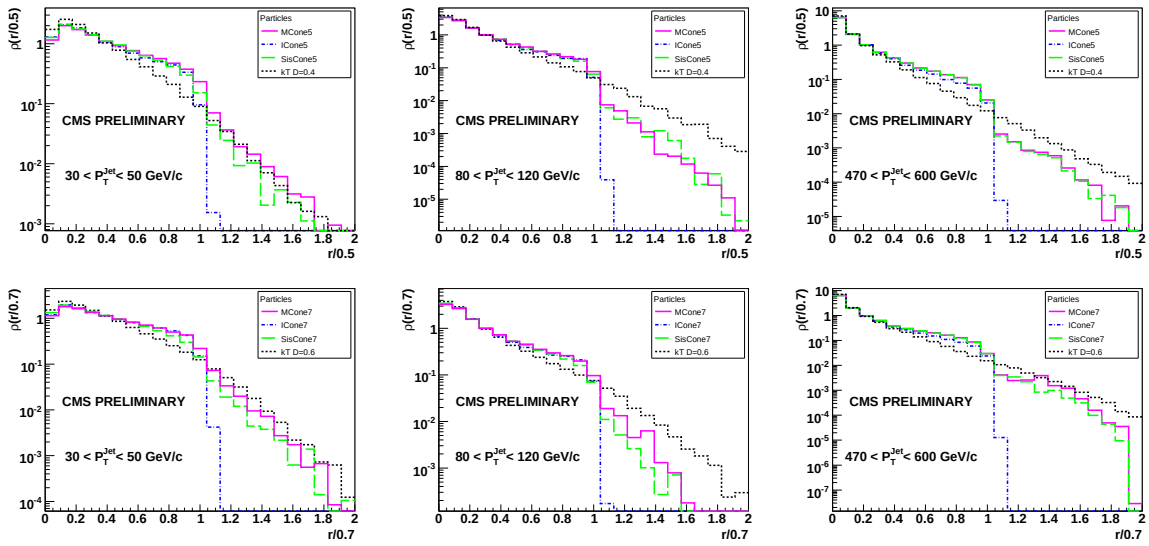


Figure 39: Normalized transverse energy density distributions ($\rho(r)$) in particle jets reconstructed using different algorithms.

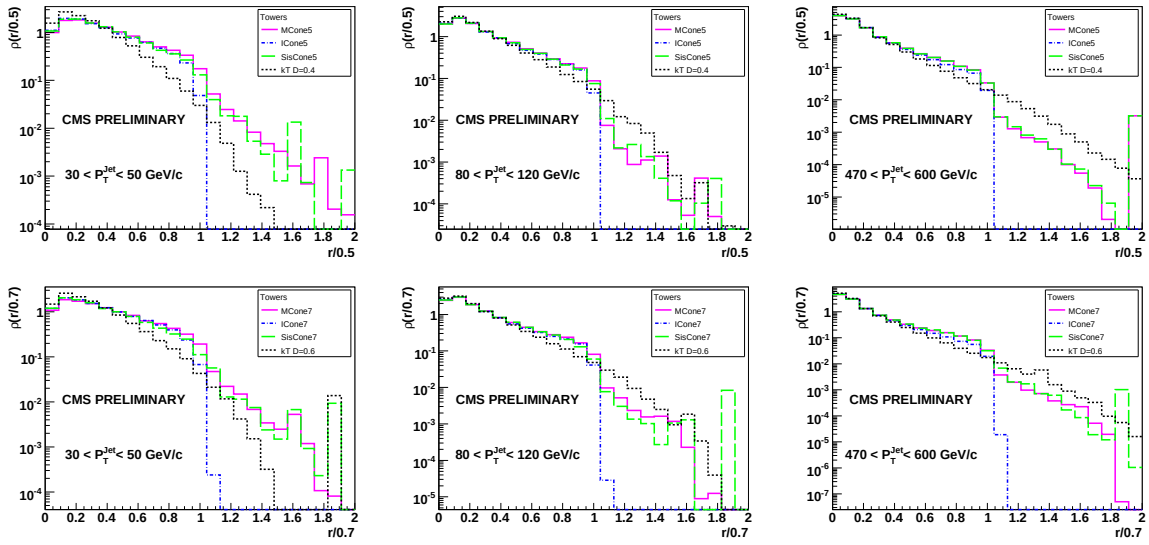


Figure 40: Normalized transverse energy density distributions ($\rho(r)$) in calorimeter jets reconstructed using the different clustering algorithms. Plots are labeled with corresponding particle jet p_{T} bins.

overlapping jets and to facilitate a treatment of QCD radiation effects.

Both SISCone and Midpoint Cone are currently supported by the CMS JetMET group, and the respective MCJet corrections are provided, to allow studies of their behavior in physics analyses and an evaluation of more technical performance benchmarks. However, due to constraints imposed by maintenance and computing considerations, such as the effort required to support and validate many algorithms, overall reconstruction time and the size of full-event and AOD outputs, choices will have to be made in the near future to limit the number of algorithms included in the regular production.

In this section we briefly summarize the information concerning the relative performance of the SISCone and Midpoint Cone algorithms; the related distributions have been presented throughout the note. Results of additional detailed studies, including the impact of pileup, are presented in a separate note [14].

The Midpoint Cone algorithm has been developed for and extensively used during the Tevatron Run II. While it is preferred to the earlier Snowmass Accord algorithm used in Run I, the accumulated experience has uncovered some deficiencies (see eg. Ref. [15]). The primary theoretical issue for Midpoint Cone is the lack of full InfraRed and Collinear (IRC) safety. While adding the midpoint seeds makes this algorithm IRC safe at the next-to-leading order, it is not IRC safe at higher orders. In addition, the algorithm is known to occasionally leave energetic detector towers unclustered, referred to as "dark towers" (leading to unreconstructed "dark jets"). A solution to this problem based on identifying stable cone positions using a "search cone" of smaller size than the one used subsequently in the final jet reconstruction, was investigated during Run II. It is theoretically disfavored and has been abandoned.

The SISCone algorithm has been developed specifically to address the IRC problem by removing the use of seeds; as a consequence, the algorithm is IRC safe to all orders. Special techniques have been employed to achieve reconstruction speed comparable to Midpoint Cone. Therefore, the SISCone algorithm is clearly theoretically preferred for QCD measurements and for other applications where the accuracy of higher-order QCD calculations is desired.

The SISCone algorithm also avoids the "dark jet" problem by implementing the option of sequential multiple passes which cluster the towers/particles left unclustered by the previous pass. This approach is much less controversial than the "search cone" fix for Midpoint Cone.

These features make SISCone a superior algorithm for theory predictions and for reconstruction at particle level. For the calorimeter-level reconstruction, the performance of SISCone is similar to Midpoint Cone, and the increase in the jet reconstruction time is modest, in fact negligible, within the full-event time budget.

It seems worthwhile to mention that the use of the SISCone algorithm, implemented as an external package, by both CMS and ATLAS will go a long way towards the goal of direct comparisons of jet based measurements without the unnecessary uncertainties induced by the experiment-specific implementations; it is possible that this will be true for comparisons to the Tevatron results as well.

Therefore we recommend the SISCone algorithm to be adopted as the default CMS cone-based jet clustering algorithm for offline reconstruction, and possibly to discontinue the use of Midpoint Cone if not required by other production considerations. We also recommend that all algorithms discussed in this note continue to be supported within the software framework, so that individual users can investigate them by running the desired jet reconstruction on the analysis level (using either full-event or AOD samples). To this end, it seems crucial that the energy corrections are officially derived and provided for all the algorithms, regardless of their inclusion in the AOD.

10 Conclusions

We presented detailed comparisons of performance between four jet reconstruction algorithms currently available in CMSSW with two radius parameter choices each: Iterative Cone, SISCone, and Midpoint Cone with $R = 0.5$ and 0.7, and Fast k_{T} with $D = 0.4$ and 0.6. The performance comparisons presented in this note include jet energy response, position resolutions, energy resolutions, efficiencies and fake rates in QCD dijet samples, reconstruction of the more complex $t\bar{t}$ signal, and jet composition and shape distributions. We have developed two data-based techniques to derive the jet energy resolution, which agree well with results based on MC truth for $p_{\mathrm{T}} > 300\,\mathrm{GeV}$ and are within 10% for lower momenta. All results are based on QCD dijet and $t\bar{t}$ samples generated with CMSSW_1_5_2, thus providing a new baseline for the jet reconstruction.

We find similar performance on the calorimeter level between algorithms with similar size parameter. The impact

of the detector effects appears to be more pronounced than the algorithmic differences studied in this note. We also find that the SISCone algorithm performs as well or better than the Midpoint Cone, while known to be preferred theoretically. Therefore we recommend to adopt SISCone as the default cone-based jet algorithm and consequently to include it in the reconstruction in future standard event processing at CMS.

The k_{T} algorithm is infrared- and collinear safe to all orders of pQCD as well and complementary to the cone-based algorithms. The execution time of Fast k_{T} is dramatically reduced w.r.t. earlier k_{T} implementations and it is therefore well suited for the high multiplicity environment of LHC pp collisions, in fact executing faster than all cone-based algorithms but Iterative Cone. We find that it performs as good or better than any other algorithm in this note and strongly encourage its use as an alternative to SISCone. Further studies will be conducted regarding the performance of all algorithms in events with high pileup and more realistic calorimeter noise.

References

[1] R. Demina *et al.*, *"Calorimeter Energy Thresholds for Jet Reconstruction in CMS"*, CMS NOTE 2006-020 (2006).

[2] G. C. Blazey *et al.*, *"Run II jet physics: Proceedings of the Run II QCD and Weak Boson Physics Workshop"*, hep-ex 0005012 (2000).

[3] G. P. Salam and G. Soyez, *"A practical seedless infrared-safe cone jet algorithm"*, JHEP05(2007)086 (2007).

[4] M. Cacciari and G. P. Salam, *"Dispelling the N^3 myth for the Kt jet-finder"*, Phys.Lett. B**641** 57-61 (2006).

[5] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, *"Longitudinally invariant Kt clustering algorithms for hadron hadron collisions"*, Nucl.Phys.B**406**:187-224 (1993).

[6] M. Vazquez Acosta *et al.*,*"Jet and MET Performance in CMSSW_1_2_0"*, CMS IN-2007/053 (2007).

[7] R. Cavanaugh *et al.*, *"Plans for Jet Energy Corrections at CMS"*, CMS AN-2007/055 (2007).

[8] C. Dragoiu, *"Jet Matching"*, Presentation in CMS Jet Algorithms Meeting on Nov. 13^{th} 2007: `http://indico.cern.ch/materialDisplay.py?contribId=3&materialId=slides&confId=23822`.

[9] C. Dragoiu, *"Update for Jet Finding Efficiency & Jet Fake Rate"*, Presentation in CMS Jet Algorithms Meeting on Oct. 2^{nd} 2007: `http://indico.cern.ch/getFile.py/access?contribId=6&resId=1&materialId=slides&confId=22001`.

[10] I. A. Bertram *et al.*, *"Single Jet Resolutions at DØ for Run I"*, DØ Note 3414 (1998).

[11] B. Abbott *et al.* [DØ Collaboration], *"High-p_T Jets in $\bar{p}p$ Collisions at $\sqrt{s} = 630$ and 1800 GeV"*, Phys.Rev.D**64**, 032003 (2001).

[12] O. N. Francisco, *"Inclusive Jet Production Studies at the Tevatron using the CDF Detector"*, Ph.D. Dissertation, Universitat Autònoma de Barcelona, Barcelona, Spain (2007) (unpublished).

[13] M. Dasgupta, L. Magnea, G. P. Salam, *"Non-perturbative QCD effects in jets at hadron colliders"*, DFTT-27-2007, MAN-HEP-2007-41 (2007).

[14] A. Bhatti *et al.*, *"Performance of the SISCone Jet Clustering Algorithm"*, CMS IN in preparation.

[15] M. G. Albrow *et al.*, *"Tevatron-for-LHC report of the QCD working group,"*, arXiv:hep-ph/0610012 (2006).