Available on CMS information server

CMS AN 2007/002

The Compact Muon Solenoid Experiment

Analysis Note

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

March 31, 2008

Performance of the SISCone Jet Clustering Algorithm

A. Bhatti

The Rockefeller University, New York, NY, USA.

F. Chlebana, R. Harris, K. Kousouris

Fermilab, Batavia, IL, USA.

Z. Qi, M. Zielinski

University of Rochester, Rochester, NY, USA.

F. Ratnikov

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

N. Varelas, C. Dragoiu

University of Illinios at Chicago, Chicago, IL, USA.

M. Jha

University of Dehli, India and INFN, Bologna, Italy.

P. Kurt, H. Topakli

Cukurova University, Adana, Turkey.

G. Dissertori

ETH Zurich, Zurich, Switzerland.

P. Schieferdecker

CERN, Geneva, Switzerland.

Abstract

We compare the performance of the Seedless Infrared Safe Cone (SISCone) jet clustering algorithm with the Midpoint algorithm for jet reconstruction in CMS calorimeters. It is shown that reconstructed

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

quantities are similar for the two algorithms and they have similar performance for multijet processes such as top production. Unlike the Midpoint algorithm, SISCone is both infrared and collinear safe, does not leave unclustered energy, and is preferred by theorists over traditional cone based clustering algorithms. SISCone has been fully integrated into the CMS software framework. We propose that SISCone be adopted as the default cone based jet clustering algorithm for CMS.

1 Introduction

Standard model processes in proton-proton collisions involving large momentum transfers are described by the scattering of partons. While partons are not directly observable they manifest themselves through hadronization as stable particles which can then be detected in tracking chambers and calorimeters. Perturbative theory and the hadronization model describe the interaction between constituent partons of the protons and the subsequent showering into stable particles. In addition to the hard interaction effects such as the underlying event and multiple pp interactions, which will change the observable energy flow, also have to be modeled. The evolution of a jet from the hard interaction to observable energy deposits is shown schematically in Figure 1.

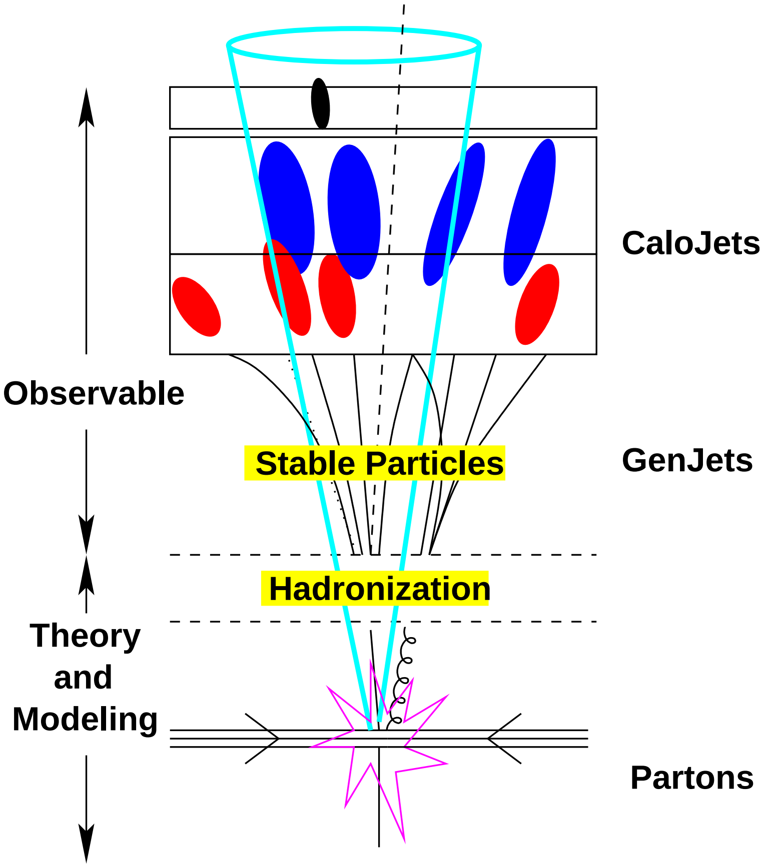

Figure 1: Evolution of a jet

Jet algorithms cluster energy deposits in the calorimeter or four-vectors of particles. A successful jet algorithm will provide a good correspondence between the parton level and the particle level, where particle level refers to the stable particles remaining after the hadronization stage. Traditionally at hadron colliders, jets have been defined using cone based clustering algorithms which search for stable cones around the direction of significant energy flow. The steps in a typical cone based algorithm are shown in Figure 2. Initially, a cone is defined using the highest E_T particle (or four-vector) and the summed four-vector is calculated for all particles within the cone resulting in a proto-jet. The procedure is repeated until a stable proto-jet is found such that the proto-jets's four-vector coincides with the sum of the four-vectors of all the particles within the cone. Once all stable proto-jets are found, a splitting/merging procedure is applied to ensure that all particles will end up in only one jet.

Iterative cone algorithms that consider every tower as the starting direction for the initial cone take a prohibitively long time to execute. In order to reduce the computation time, a minimum p_T requirement is applied to the four-vectors resulting in a subset referred to as "seeds" for the initial trial cone direction. If a p_T cut is applied to the particles used as seeds, then the procedure becomes collinearly unsafe at pQCD parton level and different sets of stable jet configurations can be found depending on the p_T cut. Cone algorithms which use seeds also have the problem of being infrared unsafe at pQCD parton level. The addition of a soft parton can lead to a new stable cone configuration. These two effects are illustrated in Figure 3 and Figure 4.

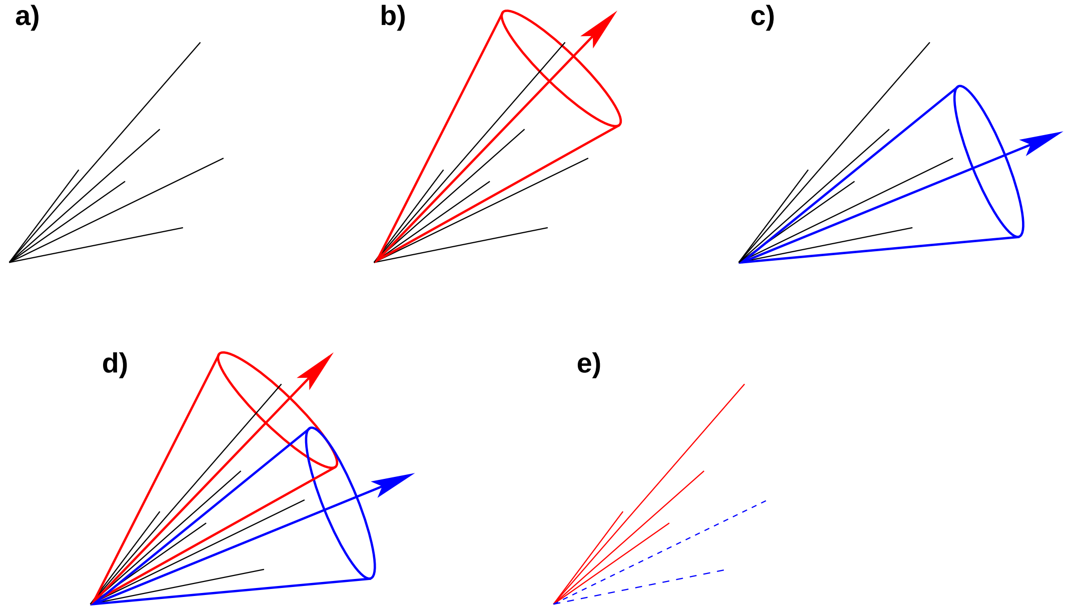

Figure 2: Steps in a jet clustering algorithm. a) Starting from an ordered list of four-vectors. b) and c) Stable cones are found. d) A splitting/merging algorithm is applied so that any four-vectors within the overlap region of cones is assigned to a single jet. e) We end up with a final list of jets.

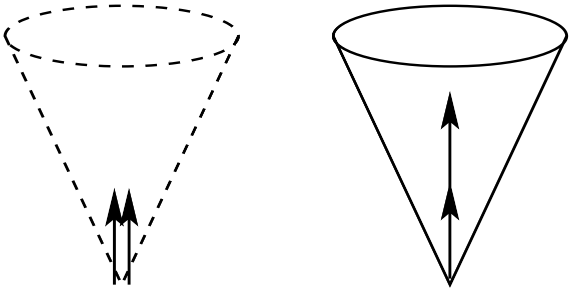

Figure 3: Collinear unsafe. Changing the p_T cut used for the seeds can lead to different stable cone configurations.

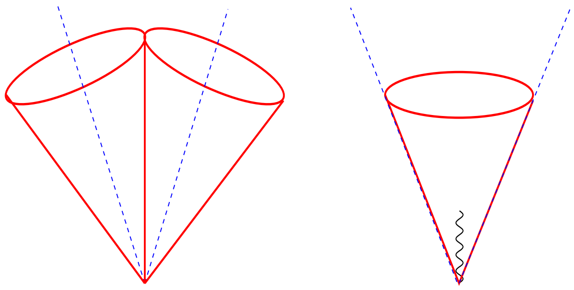

Figure 4: Infrared unstable. The addition of a soft particle can lead to new stable jet configurations.

Additional seeds were added to the midpoint between stable proto-jets in order to make the clustering algorithm less sensitive to the infrared safe problem. This procedure, coded in the Midpoint algorithm [1], is known to only delay the problem of infrared safety to a higher order of the perturbative calculation [2]. Table 1, extracted from Reference [2], lists some processes together with the order at which some jets become unstable. As higher order calculations become available it will be necessary to use jet algorithms that are not sensitive to these problems.

The problems discussed above are not present in clustering algorithms based on sequential recombination such as k_T, Jade, and Cambridge/Aachen [3]. The k_T algorithm merges pairs of four-vectors in order of increasing relative transverse momentum. The procedure is repeated until some stopping requirement is achieved, typically the distance between adjacent "jets" is greater than some value. These algorithms are infrared and collinear safe, have no artificial parameters, do not leave unclustered energy, and can be applied equally well to both data and theory. One feature of these algorithms is that the jet area is not well defined making the subtraction of the underlying event more difficult. Initial implementations of these algorithms were also very CPU intensive making them impractical to use. A faster implementation of the k_T algorithm is now available [5].

Recently the Seedless Infrared Safe Cone (SISCone) [2] algorithm has been developed which significantly improves the computation time allowing for all towers to be used as seeds. SISCone is both infrared and collinear safe and avoids some of the problems seen with previously used cone based algorithms. Both SISCone and the

Table 1: Summary of the order at which stable cones are missed in various processes when using the Midpoint algorithm, taken from Reference [2]

Observable	1st miss cone at	Last meaningful order
Inclusive jet cross section	NNLO	NLO
W/Z/H + 1 jet cross section	NNLO	NLO
3 jet cross section	NLO	LO
W/Z/H + 2 jet cross section	NLO	LO
jet masses in 3 jets, W/Z/H + 2 jets	LO	none

fast k_T algorithms have been fully integrated into the CMS software framework. In general cone based algorithms and sequential recombination algorithms will be sensitive to different effects and having both types of algorithms available allows for important cross checks.

In this note we compare the performance of SISCone with Midpoint. As will be shown, the performance of the two algorithms are similar and we propose that SISCone replace Midpoint as the default cone-based jet algorithm used by CMS.

1.1 Midpoint

The Midpoint algorithm is a modified cone jet clustering algorithm [1] [6] which iteratively clusters particles using their four-momenta. For each particle with $p_T > 1\mathrm{GeV}$ (seed), the $P^{Jet} = \Sigma P^k$ is calculated, where P^k is the four-momentum of a particle and the sum is over all particles within distance R from the seed particle. Using $\eta - \phi$ from P^{Jet} as the center of a new cone, P^{Jet} is re-calculated and its contents are compared with the contents of the previous cone. The process is iterated until the contents of the cone are the same as those from the previous iteration *i.e.* the cone is stable. The algorithm is also terminated if the number of iterations exceeds 100 even if no stable configuration can be determined. In this procedure, a particle can belong to many proto-jets. In order to reduce the sensitivity to soft radiation, additional seeds are added at the midpoint of pairs of protojets which are less than $2R$ apart, where $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, and stable proto-jets are searched for starting from these seeds.

Once all stable proto-jets are found, a splitting/merging procedure is applied to resolve the assignment of particles shared by different proto-jets. The splitting/merging procedure starts by ordering the proto-jets by their p_T. For the highest p_T proto-jet, i, the algorithm searches for the highest p_T overlapping proto-jet. If none exists, proto-jet i is moved to the list of jets. In case there exists an overlapping proto-jet and its shared p_T fraction is $f >= 0.75$, it is merged with proto-jet i. The original proto-jets are replaced by a single merged proto-jet. If the overlap fraction is $f < 0.75$, then shared particles are assigned to the one whose axis is the closest and the proto-jet momenta are re-calculated. The list of proto-jets is reordered after each splitting/merging step. The process is repeated until no proto-jets are left. After the splitting/merging procedure, particles are uniquely assigned to jets.

1.2 SISCone

The Seedless Infrared-Safe Cone (SISCone) jet algorithm [2] is a cone clustering algorithm which is reasonably fast, infrared safe to all orders in the perturbative expansion, and thus is theoretically sound. In contrast to iterative cone clustering algorithms which look for stable cones by starting only at the particles above a threshold (seeds), the SISCone algorithm searches for all possible stable cones. A brute force technique of finding all stable cones is to test all possible subsets of N particles for stability, N being the total number of particles. Although this technique is used in some parton level calculations [7],[8] ($N < 4$), it is not practical for large N as the number of distinct subsets grow as 2^N while the execution time grows as $\mathcal{O}(N2^N)$. The SISCone algorithm exploits the fact that a circle enclosing a set of particles can be moved around such that two of the particles lie on its circumference. Conversely, all possible stable circles of radius R can be determined by testing the circles defined by a pair of particles and radius R. The radius, R, used in the SISCone algorithm uses rapidity, y, rather than pseudo-rapidity, η, as for Midpoint. The algorithm first finds all the stable cones. Then, these stable cones are split/merged using the same procedure as the Midpoint Algorithm [6] except that it uses the scalar sum of P_T of particles in the jet as the ordering parameter. The stable cones, circles in $y\phi$ space, are determined as specified below.

For a given particle i, loop over all particles j with $\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} < 2R$, find two circles determined by the pair i, j and test for their stability. For a given circle, its four-momentum $P^{Jet} = \Sigma p^k$ where all the enclosed particles are included in the sum. A circle is stable if the same set of particles is enclosed by the circle of radius R centered at its four-momentum, P^{jet}, as enclosed by the original circle. A stable circle is added to the list of the proto-jets. An unstable circle is marked as such and added to a list. For each of two circles, defined by i, j being on the circumference, the algorithms checks four subsets of particles for stability. These four subsets are formed by including/excluding the particles i, j from the list of particles enclosed.

Sometimes a cluster of particles is not stable due to presence of a near-by jet and thus these particles, initially, are not clustered into a jet. These particles can be clustered by running the algorithm a second time after removing particles associated with stable jets found in the first pass. In the SISCone algorithm, the number of passes is controlled by an externally set parameter.

The infrared stability of the algorithm was tested by adding additional soft particles and re-running the algorithm. It was found that for the SISCone algorithm the hard jets are affected by the addition of soft particles in a fraction less than 10^{-9} of the events[2].

The source code is maintained in HepForge [4] which has a detector independent interface ensuring that the same clustering algorithm is applied by different experiments which facilitates comparisons.

2 Comparisons

In this section we compare the performance of SISCone with Midpoint. Unless otherwise noted, the comparisons were done using the CMSSW 1.5.2 based sample produced in Summer07 consisting of one million QCD dijet Monte Carlo events. The sample was generated in 21 $\hat{p_T}$ bins in order to provide sufficient statistics at high p_T.

2.1 Timing

Compared with existing seedless cone algorithms, the computation time for the SISCone algorithm has been reduced from $\mathcal{O}(N2^N)$ to $\mathcal{O}(N^2 \ln N)$, where N is the number of four-vectors being clustered. The external package FastJet has been interfaced to the CMS software framework and different clustering algorithms use the same standard interface. Tests were done comparing the execution times of several clustering algorithms available at CMS including fast k_T, Iterative Cone, Midpoint, and SISCone. QCD MC samples with $\hat{p_T}$ in the ranges 30-50, 50-80, 80-120, and 3000-3500 GeV were used. The tests were done using FastJet v.2.1.0, SISCone v.1.1.1, and CMSSW 1.7.1 and run on a desktop computer with a 3GHz Xeon processor.

Results are summarized in Figure 5 which shows the time distribution for all events. The execution time as a function of the number of towers is shown in Figure 6. Although this version of SISCone is slower than Midpoint, the execution time is still reasonable. The authors of SISCone have released a version of SISCone which is 10-15% faster. This has not yet been interfaced to CMSSW. The total CPU time spend for jet reconstruction is small compared with the total reconstruction time. For a QCD dijet sample with $\hat{p_T}$ ranging from 80 to 120 GeV, the CPU time spent on jet reconstruction is $\sim 0.02s$ compared with the total reconstruction time of $\sim 10s$.

2.2 Jet Energy Corrections

Currently available jet corrections are based on the procedure outlined in Reference [9]. The jet corrections have been determined using a QCD MC sample of about one million events generated using CMSSW 1.5.2. The jet response,

$$\text{Response} = \frac{E_T^{\text{Calo}}}{E_T^{\text{Gen}}}, \tag{1}$$

is defined as the ratio of calorimeter (CaloJet) jet E_T, to that of the particle level (GenJet) jet E_T. Calorimeter jets are uniquely matched to particle jets by finding the closest pair in $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ with the requirement that $\Delta R < 0.25$. The response is determined using all jets found in the event and binned in E_T and η in the region $|\eta| < 5$. In the current version of the corrections, it is assumed that the detector is symmetric in η and no attempt was made to smooth the response when going across the η bin boundaries.

The jet response is binned in Gen E_T and fit to a Gaussian. The means of the Gaussian are parameterized as a function of E_T for 16 η bins. The response for both SISCone and Midpoint as a function of the jet E_T is shown in Figure 7. The top row of plots are for $R = 0.5$ while the bottom row is for $R = 0.7$. The response is shown separately for one bin in the Barrel, Endcap, and Forward region. The jet correction is then provided as a function of the jet (detector) η and E_T. Figure 7 compares the correction factor for SISCone and Midpoint in different η

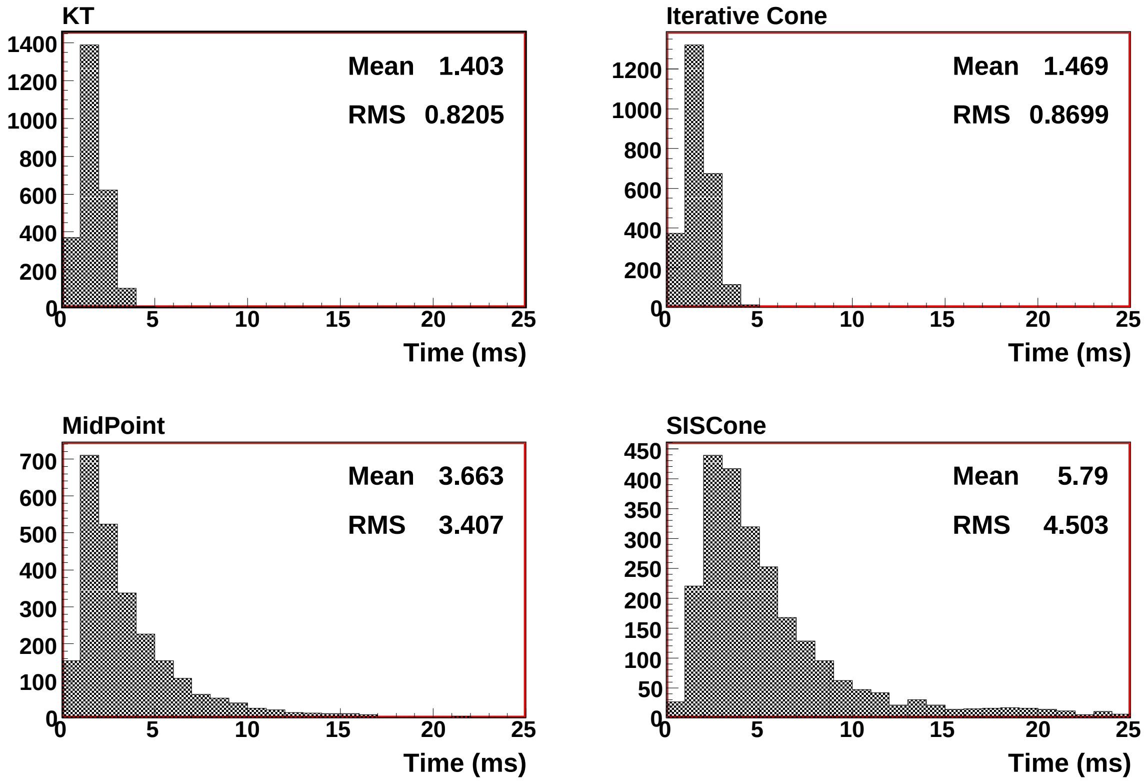

Figure 5: Execution time of several clustering algorithms available in the CMS software framework including fast k_T, Iterative Cone, Midpoint, and SISCone.

regions. The correction for SISCone and Midpoint is similar over most of the E_T and η region. For lower E_T values and for the higher η region the correction factors start to diverge.

Although the response of different jet clustering algorithms may not be the same, one should be able to correct the CaloJets back to GenJets with similar precision. This is tested by applying the correction to the jets in the sample and plotting the corrected jet response in Figure 9. The corrections for both Midpoint and SISCone are good to within about 1% for jets with $E_T > 30\text{GeV}$. More details on the correction procedure and how they can be applied can be found in the "Jet Corrections using MCJet" section of the CMS Workbook [10].

2.3 Jet Position Resolution

While the MCJet corrections are defined in terms of the jet E_T, for the analysis and results below we use p_T. The position resolution is determined by finding the CaloJet which is closest to a GenJet in $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ with the restriction $\Delta R < 0.3$. The distribution of $\Delta\phi = \phi_{Calo} - \phi_{Gen}$ and $\Delta\eta = |\eta_{Calo}| - |\eta_{Gen}|$ is then plotted in p_T bins and fit to a Gaussian over the range, mean - 1.5×RMS, mean + 1.5×RMS. The resultant widths obtained from the Gaussian fit is then fit to the function given in Equation 2.

$$\sigma(\phi, \eta) = \frac{a}{p_T} \oplus \frac{b}{\sqrt{p_T}} \oplus c \tag{2}$$

The η resolutions for SISCone and Midpoint are shown in Figure 10. The top row shows the results for $R = 0.5$ and the bottom row shows the results for $R = 0.7$. Different η bins are shown in the three columns. The ϕ resolution is plotted in Figure 11. Both SISCone and Midpoint give comparable position resolutions.

2.4 p_T Resolution of Corrected Jets

The jet p_T resolution is defined as

$$\frac{\sigma(p_T)}{p_T} = \frac{\sigma(p_T^{Corr}/p_T^{Gen})}{< p_T^{Corr}/p_T^{Gen} >}. \tag{3}$$

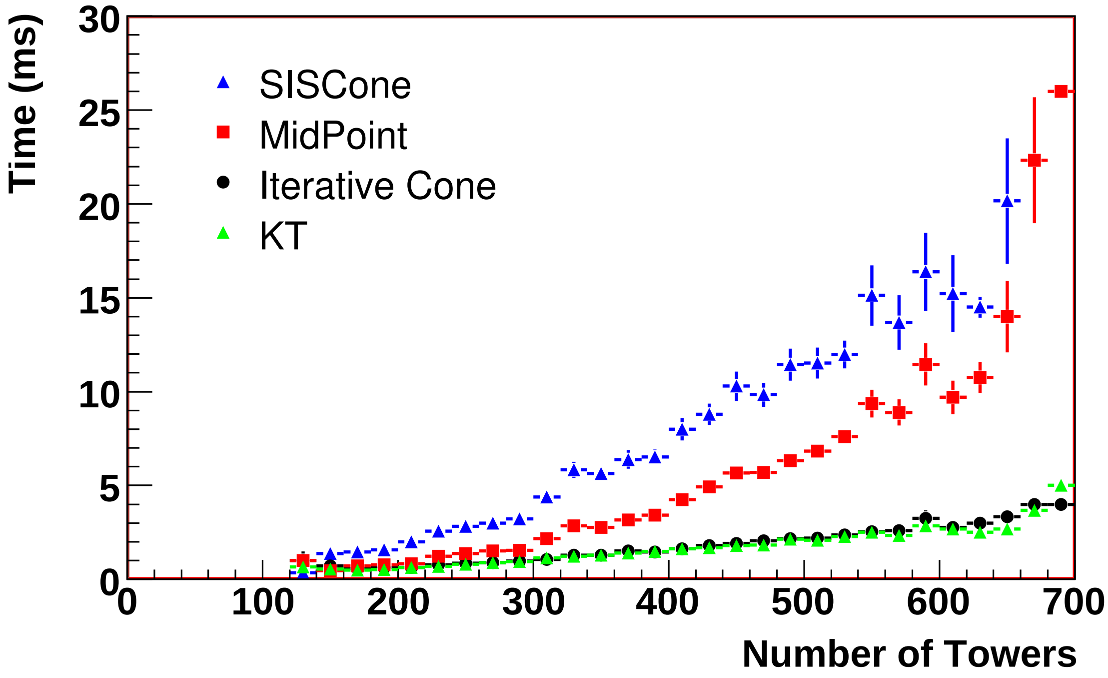

Figure 6: Execution time per event of several clustering algorithms available in the CMS software framework including Midpoint, SISCone, IterativeCone, and fast k_T clustering.

Equation 3 is a measure of how well the jet algorithm is able to measure the particle energy deposition. Additional effects which will smear the resolution arise from the hadronization process of going from the parton level to stable particles and are not included in Equation 3. The jet p_T resolution for SISCone and Midpoint is shown for the two leading jets in different η bins in Figure 12. The top row shows the results for $R = 0.5$, while the bottom row shows the results for $R = 0.7$. The resolution for the third leading jet is shown in Figure 13. The p_T resolution for the two algorithms is similar.

2.5 Matching Efficiencies

The jet matching efficiency is defined as the fraction of the GenJets matched to CaloJets using the matching requirement ΔR. The matching procedure does allow for multiple matches. Figure 14 shows ΔR for jets with $15 < p_T < 25$ GeV and ΔR for jets with $40 < p_T < 50$ GeV in the Barrel region. The matching requirement of $\Delta R < 0.3$ works well for jets with $p_T > 40$ GeV. The position resolution is not very good for jets with lower p_T and a larger ΔR matching cut should be used.

Figure 15 compares the matching efficiency of SISCone and Midpoint as a function of the GenJet p_T for $\Delta R < 0.3$ (left) and $\Delta R < 0.5$ (right) in the Barrel region. As expected the matching efficiency improves for low p_T as the ΔR requirement is increased. The matching efficiency for SISCone is comparable to Midpoint if not slightly better.

Figure 16 shows the matching efficiency as a function of the jet η for the low p_T jets with different matching requirements. The left plot shows the matching efficiency with a matching requirement of $\Delta R < 0.3$ while the right plot shows the results for $\Delta R < 0.5$. While the choice of the clustering distance (R determined using y or η) affects the values of the efficiencies, the performance for SISCone seems better than for Midpoint.

2.6 Unclustered Energy

In some cases the Midpoint algorithm will leave towers unassigned to a jet. These unclustered towers can have a significant E_T as is illustrated in Figure 17. The left plot shows the total E_T of all clustered towers in the jets. The

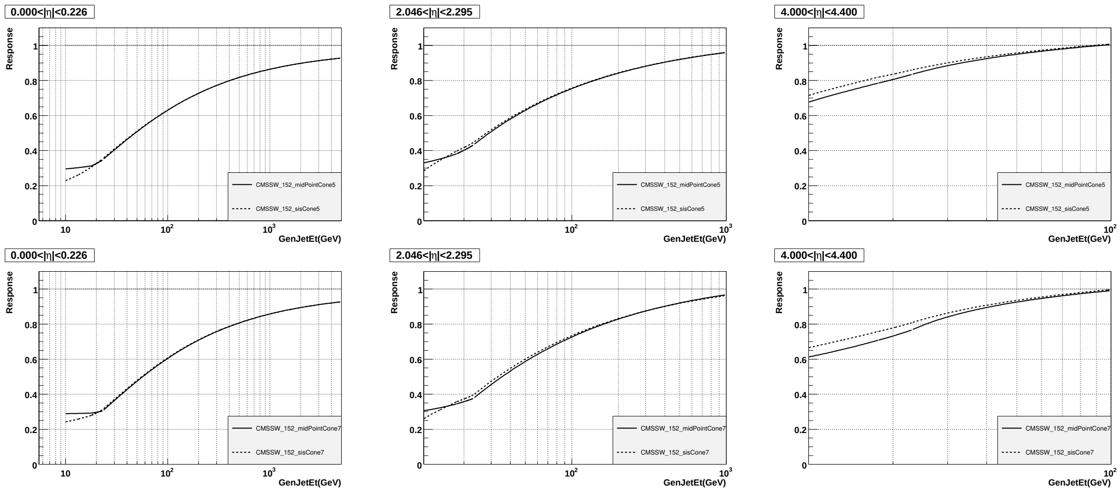

Figure 7: Jet response as a function of the GenJet E_T for Midpoint (solid) and SISCone (dashed). The top row shows the response for $R = 0.5$ and the bottom row is for $R = 0.7$. The response is shown for one of the 16 η bins in the Barrel (left column), Endcap (middle column), and the Forward (right column) regions.

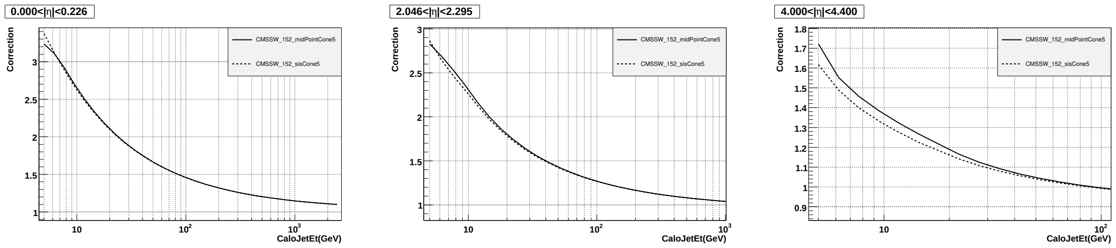

Figure 8: Comparison of the correction factor for SISCone and Midpoint as a function of the CaloJet E_T. Three separate η bins are shown in the Barrel (left column), Endcap (middle column), and the Forward (right column) regions.

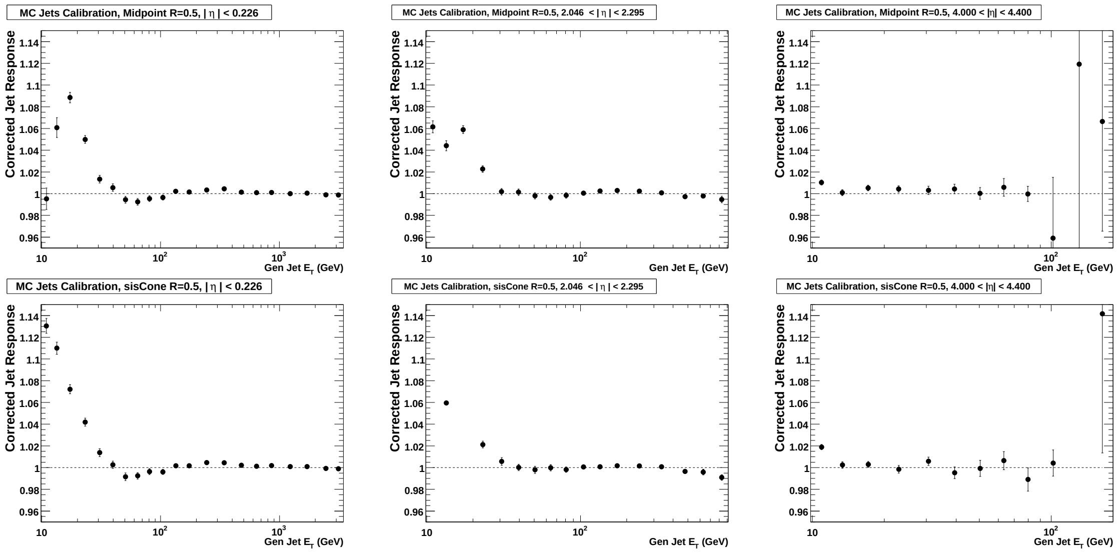

Figure 9: The corrected jet response shown as a function of GenJet E_T. The top row shows the results for Midpoint and the bottom row shows the results for SISCone. Three separate η bins are shown in the Barrel (left column), Endcap (middle column), and the Forward (right column) regions.

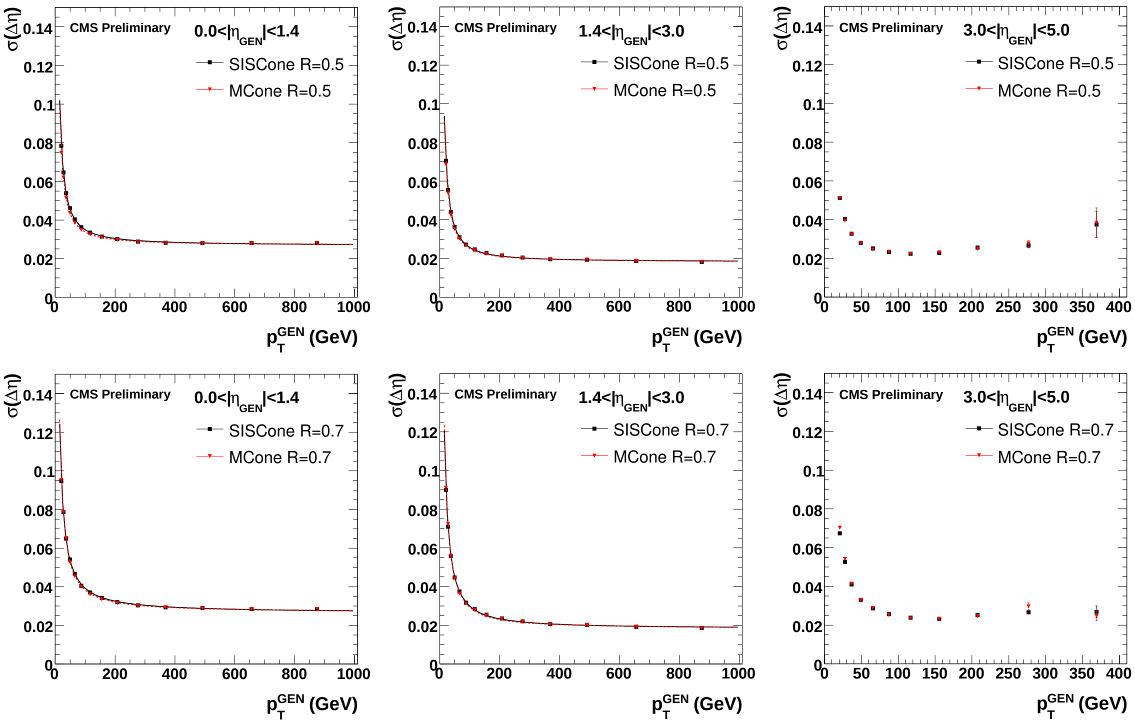

Figure 10: The η resolution for Midpoint (dashed) and SISCone (solid). The top row is for jets with a cone size of 0.5 and the bottom row is for jets with a cone size of 0.7. The columns show the results for different η bins; $|\eta| < 1.4$ (left), $1.4 < |\eta| < 3.0$ (center), and $3.0 < |\eta| < 5.0$ (right).

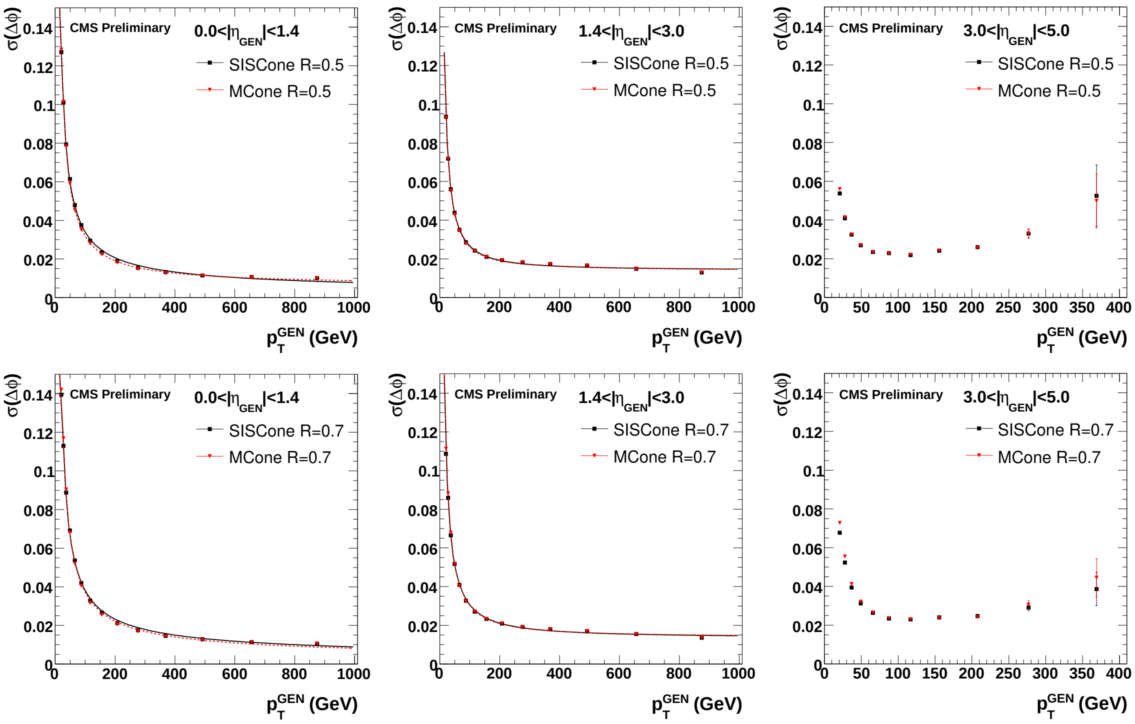

Figure 11: The ϕ resolution for Midpoint (dashed) and SISCone (solid). The top row is for jets with a cone size of 0.5 and the bottom row is for jets with a cone size of 0.7. The different columns show the results for different η bins; $|\eta| < 1.4$ (left), $1.4 < |\eta| < 3.0$ (center), and $3.0 < |\eta| < 5.0$ (right).

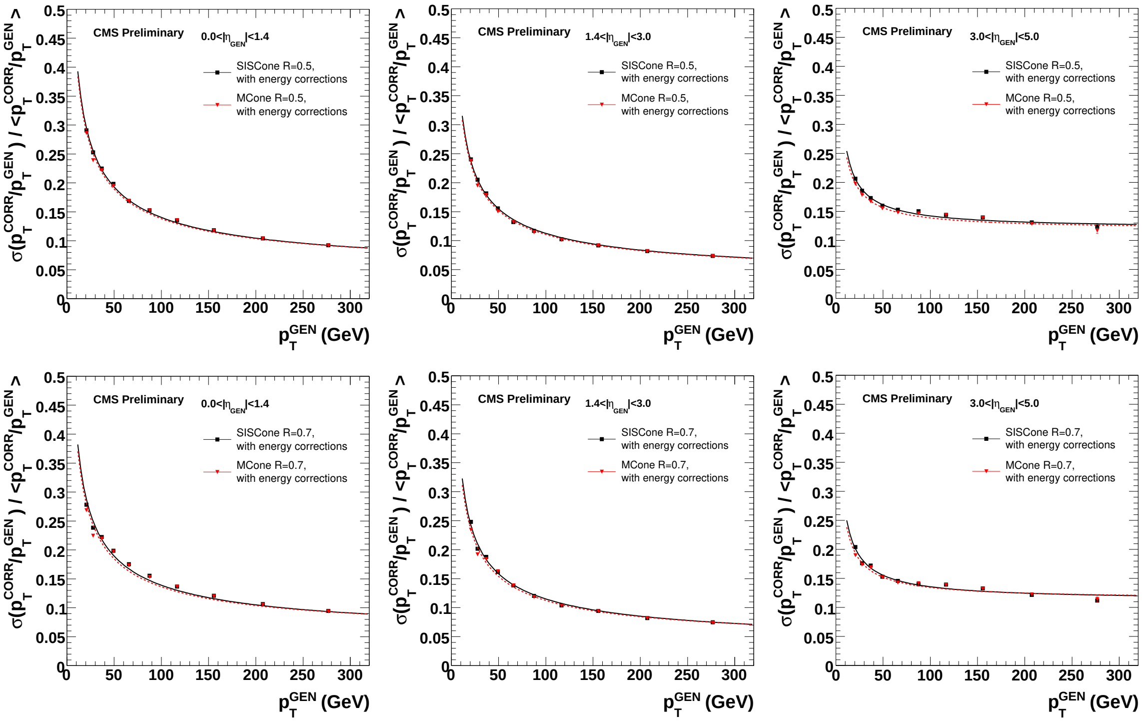

Figure 12: Jet p_T resolution of Midpoint (dashed) and SISCone (solid) for the two leading jets. The top row shows the resolution for jets with $R = 0.5$ and the bottom row is for $R = 0.7$. The columns show the resolution in different η bins; $|\eta| < 1.4$ (left), $1.4 < |\eta| < 3.0$ (center), and $3.0 < |\eta| < 5.0$ (right).

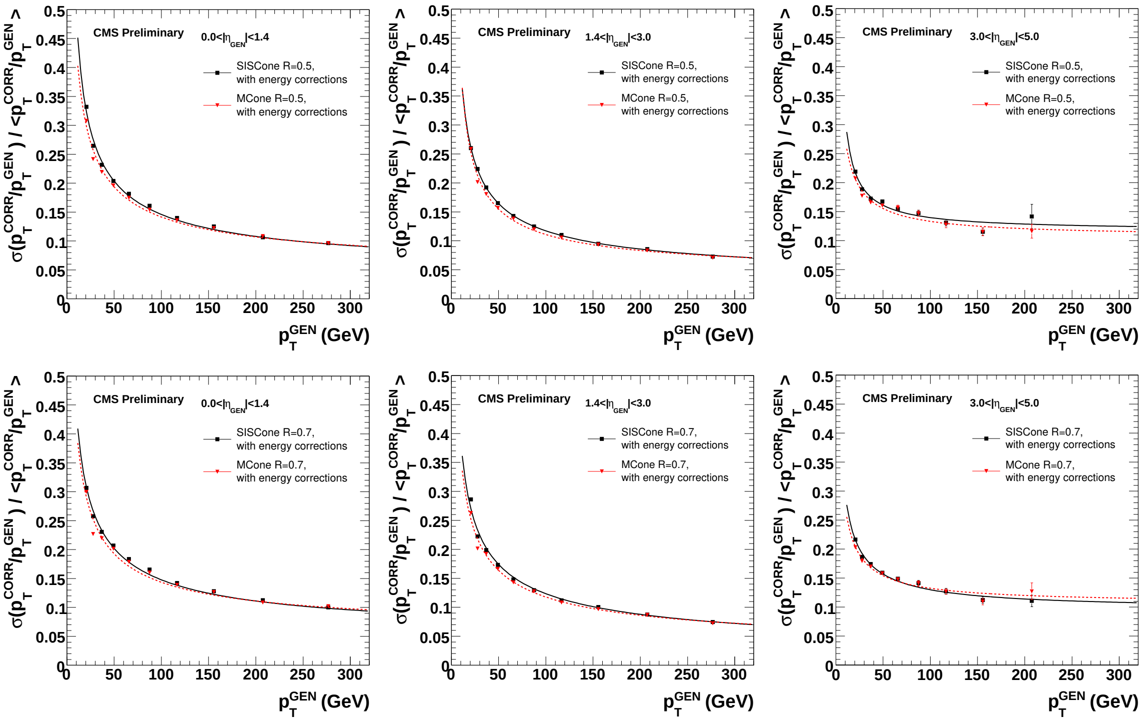

Figure 13: Jet p_T resolution of Midpoint and SISCone for the third leading jet. The top row shows the resolution for jets with $R = 0.5$ and the bottom row is for $R = 0.7$. The different columns show the resolution in different η bins; $|\eta| < 1.4$ (left), $1.4 < |\eta| < 3.0$ (center), and $3.0 < |\eta| < 5.0$ (right).

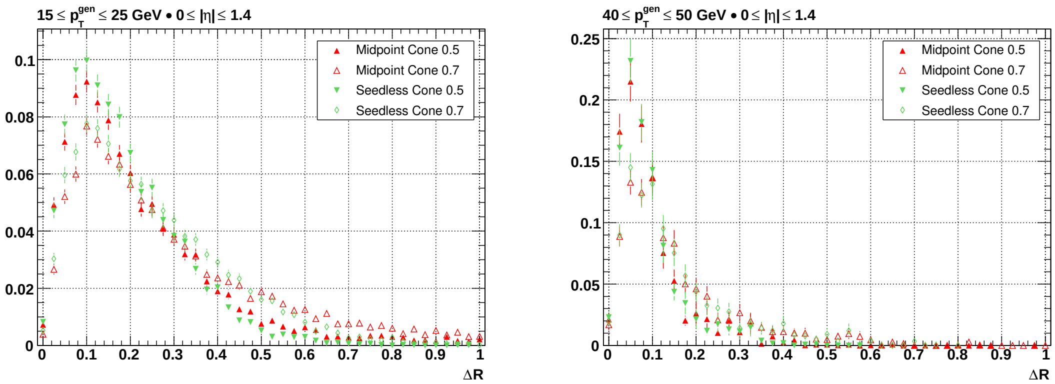

Figure 14: Left: ΔR for jets with $15 < p_T < 25$ and $|\eta| < 1.4$. Right: ΔR for jets with $40 < p_T < 50$ GeV and $|\eta| < 1.4$.

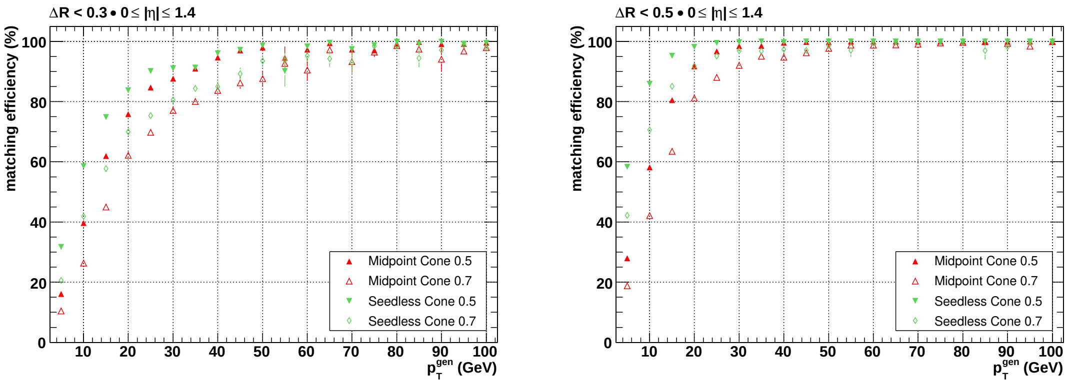

Figure 15: Comparison of the SISCone and Midpoint jet matching efficiency versus p_T for a matching requirement of $\Delta R < 0.3$ (left) and $\Delta R < 0.5$ (right).

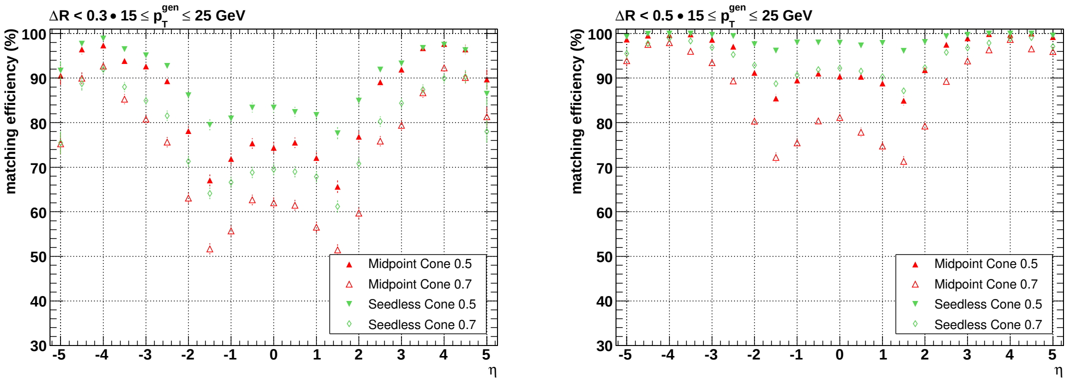

Figure 16: The jet matching efficiency as a function of η for $\Delta R < 0.3$ (left) and $\Delta R < 0.5$.

middle plot shows the total E_T of the unclustered towers. A requirement of $E_T > 0.5$ was placed on the towers. The right plot shows the E_T distribution of unclustered towers. For k_T and SISCone there are no unclustered high E_T towers, while for the Midpoint case there are unclustered towers as high as $E_T \sim 45\mathrm{GeV}$.

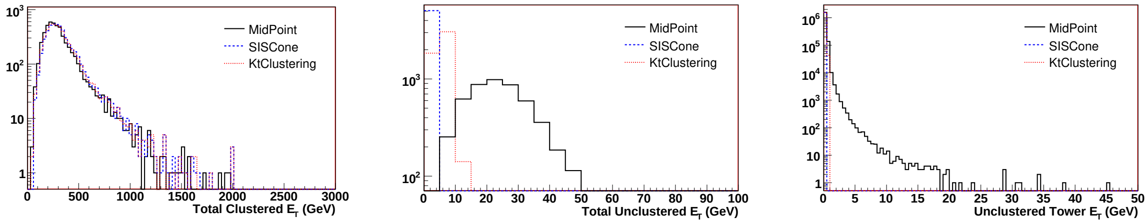

Figure 17: The left plot shows the total E_T of towers clustered in the jets. The middle plot shows the total E_T for towers not included in the jets. The right plot shows the E_T distribution of unclustered towers. A E_T requirement of $E_T > 0.5\mathrm{GeV}$ was placed on the towers.

2.7 Pileup

The Large Hadron Collider (LHC) will collide protons with an instantaneous luminosity of up to $10^{34}\mathrm{cm}^{-2}s^{-1}$ and a bunch spacing of 25 ns. While this high luminosity is essential for many searches of rare new physics processes at high energy scales, it also complicates analyses, because at each bunch crossing there will be of the order of 20 minimum bias pp interactions, which add many soft particles to the interesting event. The beams at LHC will have a longitudinal spread, and it may be possible experimentally to associate charged particles with a distinct primary vertex that corresponds to a single pp interaction and eliminate some fraction of the soft contamination. However, for neutral particles and for jet measurements carried out with calorimeters, this is not possible and kinematic measurements of jets will be adversely affected by pileup (PU).

The effects of pileup have been studied for Midpoint and SISCone using a sample of QCD dijet events generated using PYTHIA with $\hat{p_T}$ in the range 120-170 GeV. Events are processed through the full CMS detector simulation and reconstruction package using CMSSW 1.5.2. To simulate additional proton-proton interactions in a beam crossing, the simulated hits (simhits) of signal events were mixed with the simhits from minimum bias events. The minimum bias events were generated with PYTHIA as inclusive QCD events using a Poisson distribution with an average of five events, corresponding to a luminosity of $2 \times 10^{33}\mathrm{cm}^{-2}s^{-1}$. For simulating in-time PU events, the mixing was done only for the bunch crossing corresponding to the hard scatter; for full PU events the mixing was also done in the adjacent crossings (-5, +3).

Figure 18 and Figure 19 show the shift in jet p_T due to full pileup. The shift in jet p_T was calculated from the difference of jet's p_T with and without full pileup. Jets from the two samples are matched by requiring $\Delta R < 0.1$ between the two jets. The shift is significant (up to 25 and 16 GeV for Midpoint and SISCone respectively) when averaging over the entire η region and varies considerably from jet to jet, both because of variation in jet areas and because the pileup fluctuates from event to event. The negative shifts observed for a small subset of jets are attributable to the pileup having modified the clustering sequence, for example breaking one hard jet into two softer subjets. The mean value of the shift in the jet's p_T due to full PU is 2.2 (1.3) GeV for Midpoint (SISCone) when looking at jets over the entire η region. A similar comparison was done only for in-time PU and the average shift over the entire η range is 0.86 (0.52) GeV for Midpoint (SISCone). Table 2 lists the shift in the jet p_T for the three calorimeter segments and for jets within the range 10 - 60 and $>$ 60 GeV. Jets reconstructed using SISCone algorithm appear to be affected less than when using the Midpoint algorithm.

2.8 Multijet Events

The ability to resolve multijets was studied using 5050 $t\bar{t}$ events (RelVal152TTbar). About 45% or 2280 events decay in the fully hadronic mode for which we expect six reconstructed jets. For this subset of events we count the number of matched reconstructed jets to the parton initiating the jet using the requirement $\Delta R < 0.3$. Figure 20 shows the difference between the reconstructed and parton ΔR, $\Delta\phi$, $\Delta\eta$, and Δp_T. The distributions obtained for the different algorithms are very similar.

The number of matched jets found by the different algorithms is listed in Table 3. The efficiency is defined as

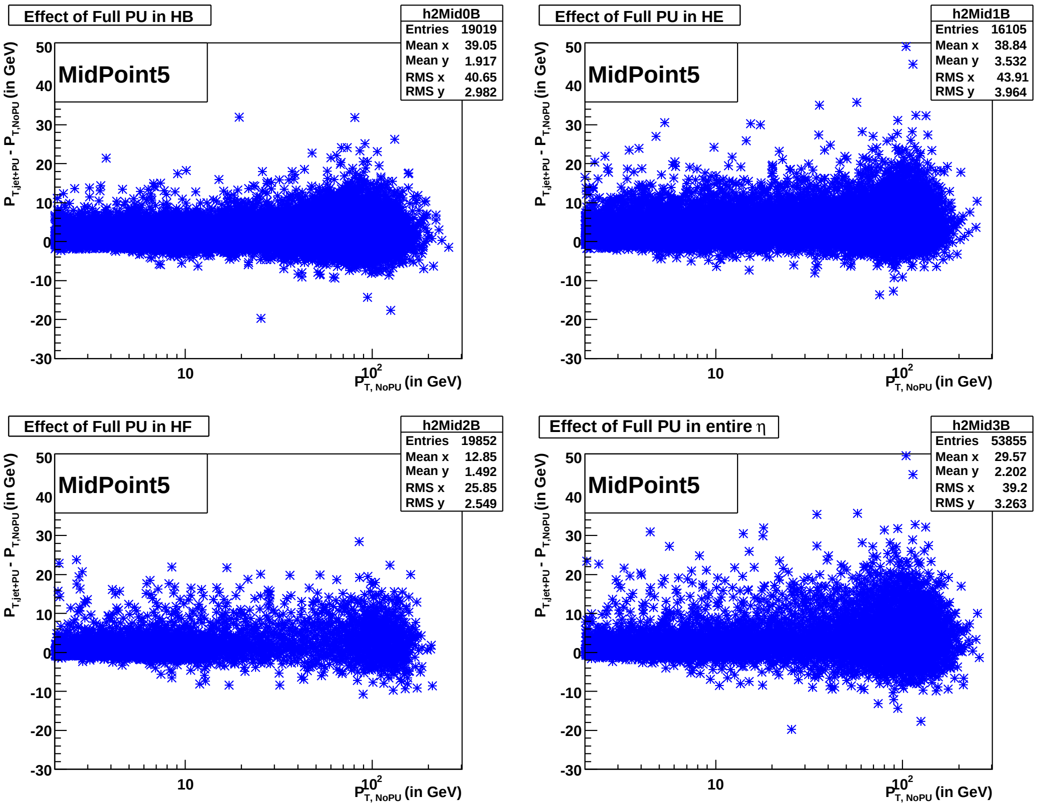

Figure 18: The p_T difference between Midpoint jets with and without full pileup is plotted as a function of the jet p_T. The shifts are shown separately for the Barrel, Endcap and Forward calorimeters.

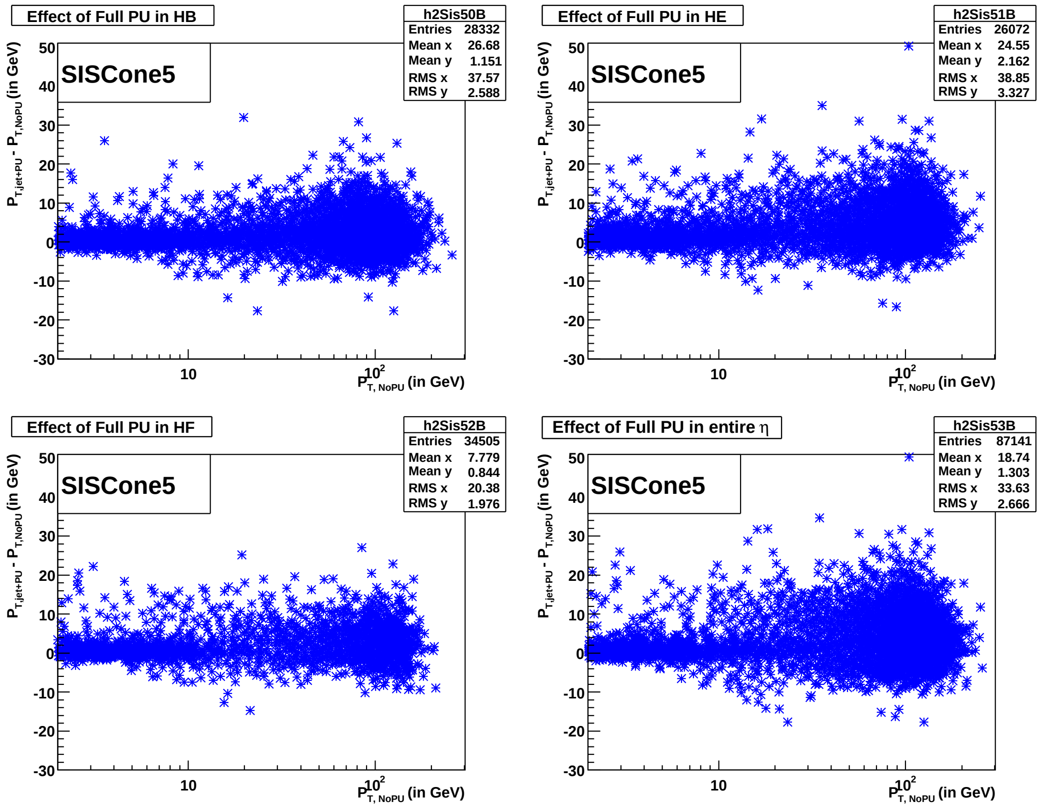

Figure 19: The p_T difference between SISCone jets with and without full pileup is plotted as a function of the jet p_T. The shifts are shown separately for the Barrel, Endcap and Forward calorimeters.

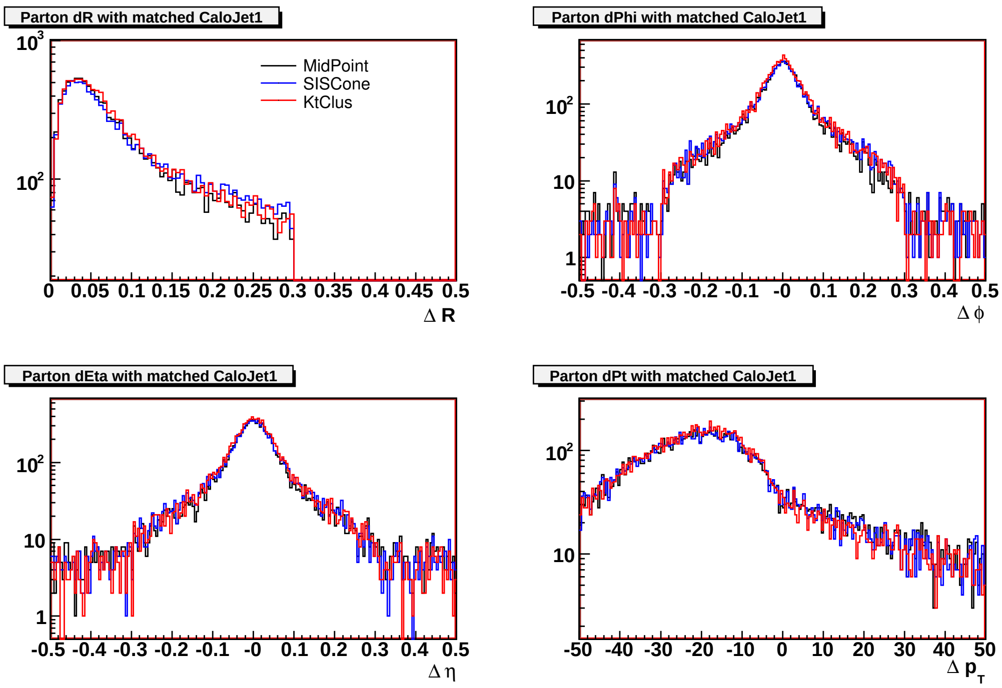

Figure 20: The difference between Calo reconstructed quantities and the parton for matched jets.

Table 2: Shift in the jet p_T (in GeV) due to full PU shown separately for the three calorimeter segments and two bins in p_T.

	Midpoint		SISCone	
p_T bin (GeV)	10 - 60	> 60	10 - 60	> 60
HB	2.2 ± 3.1	2.6 ± 3.8	1.6 ± 3.1	2.1 ± 3.7
HE	4.1 ± 4.0	5.0 ± 4.7	3.1 ± 4.0	4.3 ± 4.6
HF	2.2 ± 3.1	3.2 ± 4.5	1.5 ± 3.1	2.6 ± 4.4
All η	2.8 ± 3.5	3.6 ± 4.4	2.0 ± 3.5	3.0 ± 4.3

the number of events for which six jets are matched to the total number of events that decay in the fully hadronic mode. SISCone is found to perform as well or better than Midpoint.

Table 3: Number of matched jets for $t\bar{t}$ events with fully hadronic decay.

Jets	Midpoint	SISCone	Fast k_T
0	0	1	0
1	0	1	0
2	5	4	0
3	15	18	6
4	89	100	72
5	477	442	399
6	1694	1714	1803
Eff.	74%	75%	79%

2.9 Performance in $t\bar{t}$ Events

The performance of Midpoint and SISCone is compared for $t\bar{t}$ events in which either one ("lepton+jets") or both ("alljets") W bosons decay hadronically into a pair of quarks. The ALPGEN MC sample has been produced and reconstructed with CMSSW 1.5.2 and does not include the production of top pairs in association with additional jets (referred to as "$t\bar{t}$ +0 jets"). After requiring $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ for calorimeter jets, the jet multiplicity, p_{T} and η distributions are shown in Figure 21 both for Midpoint $R = 0.5$ (top) and SISCone $R = 0.5$ (bottom).

Out of a total of 144775 lepton+jets events, the number of events with at least four jets satisfying the p_{T} cut above is 87966 (60.8 %) for Midpoint and 86736 (59.9 %) for SISCone ($R = 0.5$). Of 144800 total alljets events, Midpoint retains 35393 (24.4 %) and SISCone 34499 (23.8 %) events when requiring at least six reconstructed jets with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$. The efficiency ϵ_{top} is defined as the number of hadronic $t \to Wb$ decays for which all three quarks in the final state can be matched to calorimeter jets within $\Delta R < 0.5$, divided by all such decays in events which pass the above described jet selection. In lepton+jets events, ϵ_{top} is found to be 16.0 % for Midpoint and 15.8 % for SISCone. ϵ_{top} is determined to be 40.4 % and 40.2 % for Midpoint and SISCone respectively in alljets events, for which $\epsilon_{t\bar{t}}$ is defined additionally as the fraction of events for which both top decays can be matched: 4.5 % for Midpoint, and 4.3 % for SISCone.

The calorimeter jets belonging to fully matched hadronic top decays are used to form dijet (m_W) and three-jet (m_t) masses in order to compare the mass resolution obtained with both algorithms. The resulting mass distributions are shown in Figure 22 for W bosons (left) and top quarks (right) for Midpoint $R = 0.5$ (top) and SISCone $R = 0.5$ (bottom) at several levels of correction: besides generator and calorimeter level distributions, results after application of MCJet energy corrections and additional flavor corrections ("Level-5") are included as well[10]. The later represents the most accurate level of correction currently available and thus provides the most meaningful measure to compare the resolution obtained with different algorithms. The RMS of the L5-corrected m_W distribution is 13.1 and 13.2 GeV for Midpoint and SISCone reconstructed jets respectively. Similar compatibility is found for the L5-corrected m_t spectra with RMS widths of 22.3 and 22.4 GeV, indicating that both algorithms yield the same mass resolution in $t\bar{t}$ events.

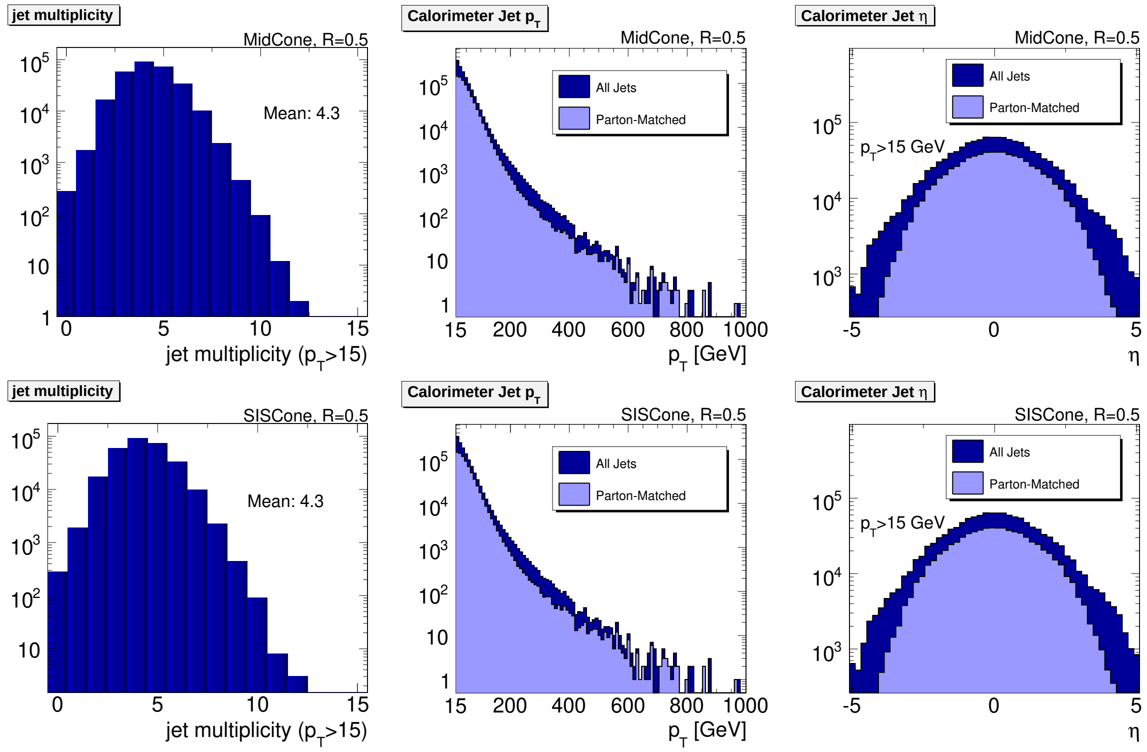

Figure 21: Jet multiplicity (left), p_{T} (middle), and η (right) distributions for jets reconstructed with Midpoint $R = 0.5$ (top) and SISCone $R = 0.5$ (bottom) in $t\bar{t}$ lepton+jets and alljets events. No attempt is made to remove isolated leptons from the list of jets. The parton distributions for hadronic top decays fully matched to calorimeter jets are also shown.

2.10 Dijet Mass Resolution in Z' Events

The dijet mass resolution depends on both the energy and position of the jets. The resolution was studied using a Z' sample generated with three different masses, 700, 2000, and 5000 GeV using CMSSW 1.6.7. The sample used has an intentional 5% channel to channel RMS applied in order to simulate the level of calibration that can be achieved with $100pb^{-1}$ of data. The dijet mass was determined from the two leading jets selected such that they both satisfy $|\eta| < 1.3$. The L2 + L3 factorized corrections[11] were applied as outlined in the WorkBook160JetReco example. The reconstructed dijet mass for GenJet, CaloJet and Corrected CaloJet is shown in Figure 23. The top row shows the results for Midpoint while the bottom row shows the results for SISCone. The reconstructed dijet mass distribution is similar for the two algorithms.

A Gaussian was fit to the distribution in the range from -1.0σ to 1.5σ centered on the mean. The fit procedure was iterated several times such that subsequent fits started from the results of the previous fit. The resolution, defined as σ/mean, was then plotted for the three mass points and presented in Figure 24.

3 Conclusion

In order to limit execution time cone-based jet algorithms use seed towers with a p_T cut making them infrared and collinear unsafe. The Midpoint algorithm is not infrared safe beyond the NLO calculation and using it introduces unnecessary uncertainties when comparing measured results to theory. The SISCone algorithm is both infrared and collinear safe, and the execution time is comparable to Midpoint. The code for the SISCone algorithm is part of the FastJet package maintained in an external repository, HepForge, and allows different experiments to use exactly the same clustering algorithm helping to facilitate the comparison and combination of results from different experiments. The SISCone algorithm has been integrated in the CMSSW framework.

A comparison of reconstructed quantities between Midpoint and SISCone show that the two algorithms give similar results. The effect of pileup was also found to impact SISCone slightly less than for Midpoint. Jet corrections are available for SISCone and provide the same level of accuracy as for Midpoint. SISCone has also been shown to be able to perform as well as Midpoint in resolving multijets in $t\bar{t}$ events and yields a comparable mass resolution when reconstructing the Z', top, or W mass. So far no pathologies have been found when using SISCone.

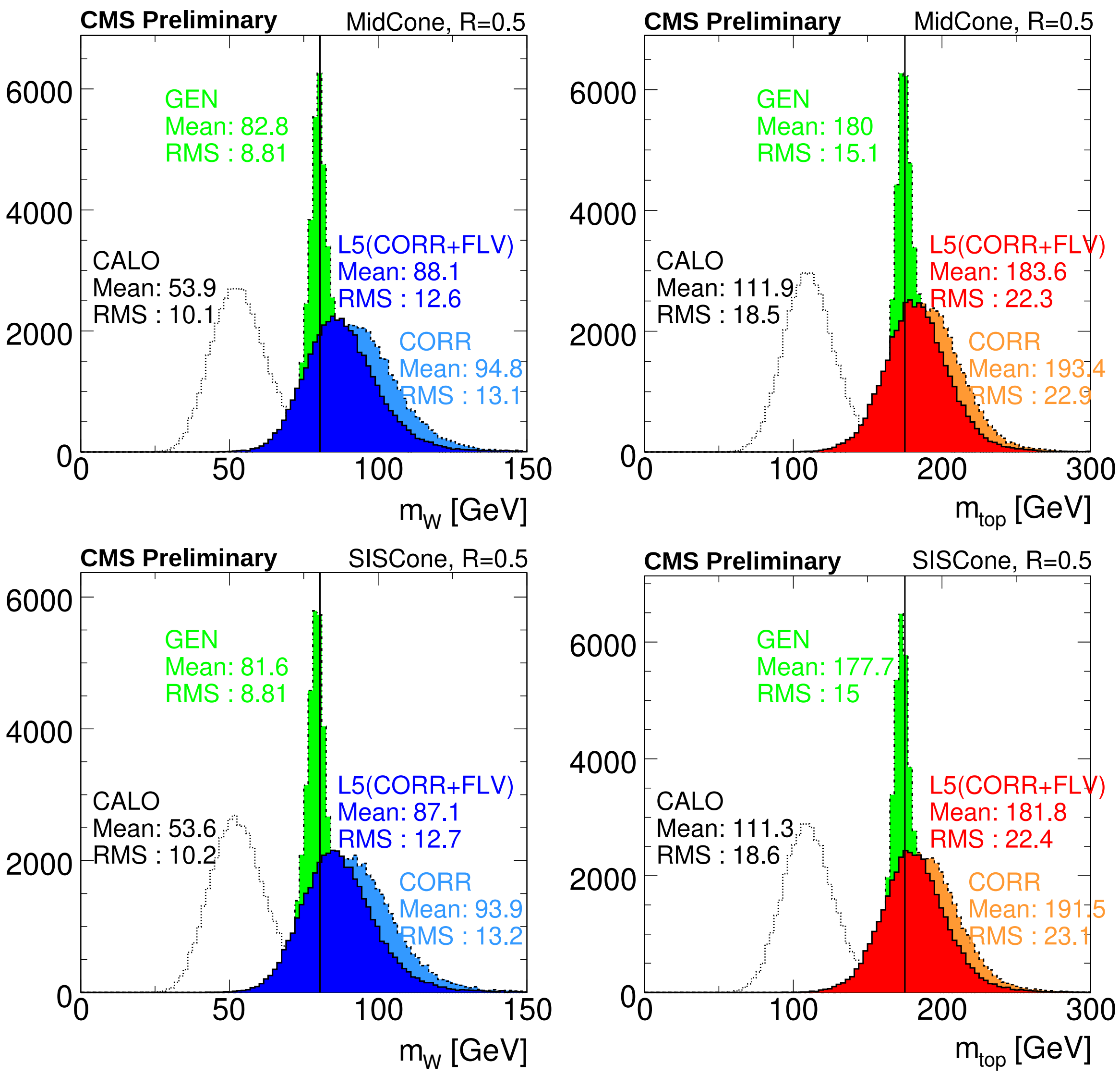

Figure 22: m_W (left) and m_t (right) distributions for hadronic top quark decays reconstructed with Midpoint $R = 0.5$ (top) and SISCone $R = 0.5$ (bottom). Four different correction levels are shown: particle-level ("GEN"), calorimeter-level ("REC"), MCJet-corrected calorimeter-level ("CORR"), and "Level-5", which accounts for the flavor dependence of the MCJet jet energy correction. The black vertical lines indicate the generated W boson and top quark mass of 80.42 GeV and 175 GeV respectively.

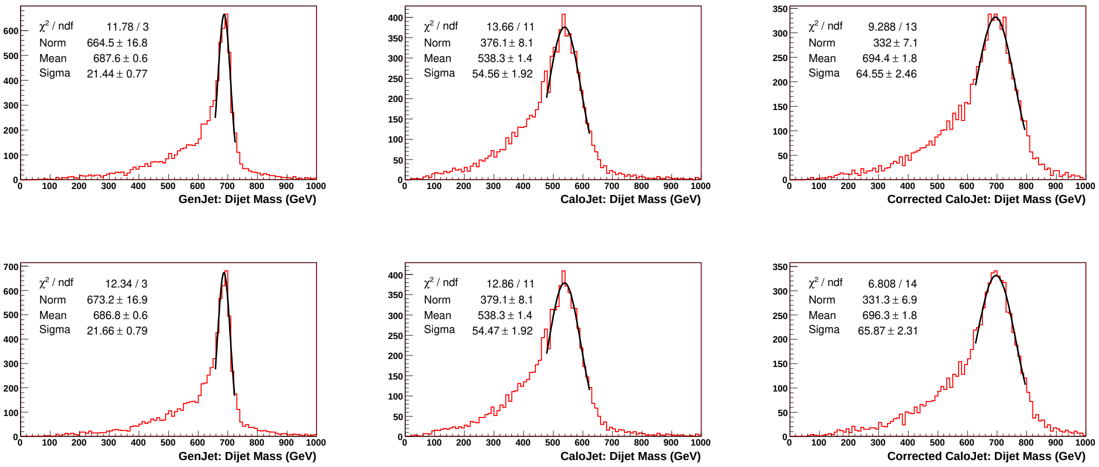

Figure 23: The dijet mass distributions for $Z' \rightarrow q\bar{q}$ events for GenJets, CaloJets, and Corrected CaloJets as determined using Midpoint (top row) and SISCone (bottom row).

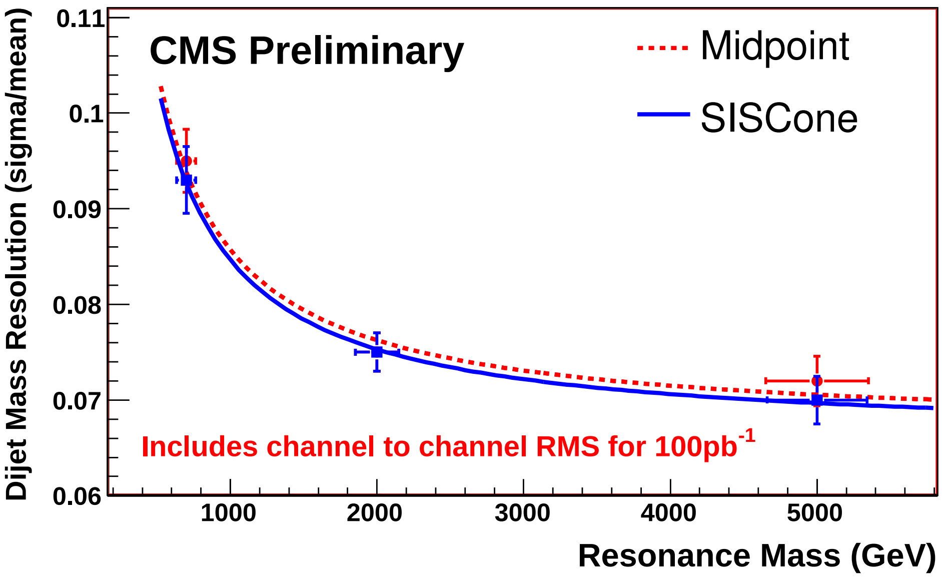

Figure 24: The dijet mass resolution, σ/mean for $Z' \rightarrow q\bar{q}$ events is plotted as a function of the Z' mass. The resolutions obtained for Midpoint and SISCone are similar.

We propose that SISCone be adopted as the default cone-based jet algorithm for CMS and added to the standard reconstruction sequence. It is recommended that Midpoint be maintained so that it can be used by the those interested at the analyses level.

References

[1] A. Abulencia *et al.,* CDF RUN II collaboration, Phys. Rev. D**74** (2006) 0711903, hep-ex/0512020; Tevatron-for-LHC Report of the QCD Working Group, TeV4LHC QCD Working Group, hep-ph/0610012.

[2] G.Soyez and G.Salam, *"A practical Seedless Infrared-Safe Cone jet algorithm"*, **JHEP 05 (2007) 086**.

[3] S.Catani, Y.L.Dokshitzer, M.H.Seymour, and B.R.Webber, Nucl. Phys. **B** 406 (1993) 187.

[4] http://projects.hepforge.org/siscone/

[5] M. Cacciari, G.P. Salam, Phys. Lett. **B** 641 (2006) 57, hep-ph/0512210.

[6] G.C. Blazey *et al.,*, hep-ex/0005012.

[7] J. Campbell and R. K. Ellis, Phys. Rev. D**65** (2002) 113007, hep-ph/0202176.

[8] Z. Nagy, Phys. Rev. Lett. **88** (2002) 122002, hep-ph/0110315; Phys. Rev. D**68** (2003) 094002, hep-ph/0307268.

[9] **CMS IN-2007/053**, M.Vazquez Acosta et al., Jet and MET Performance in CMSSW 1.2.0,

[10] CMS Workbook: "Jet Corrections using MCJet", https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookJetReco.

[11] CMS-AN note in preparation.