Available on the CERN CDS information server **CMS PAS JME-09-001**

CMS Physics Analysis Summary

Contact: cms-pog-conveners-jetmet@cern.ch 2009/07/03

A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging

The CMS Collaboration

Abstract

A new top-jet-tagging algorithm is presented. This algorithm uses the Cambridge-Aachen jet clustering algorithm to decompose highly boosted top jets into subjet components and examine kinematics of these subjets. With this algorithm, an efficiency of 46% for top-jets with $p_{\mathrm{T}} = 600$ GeV/c is obtained, together with a rejection of 98.5% for non-top jets with $p_{\mathrm{T}} = 600$ GeV/c.

1 Introduction

Top quarks play an important role in electroweak symmetry breaking scenarios because of the large coupling to the Higgs field compared to other quarks. Moreover, many theoretical extensions of the Standard Model contain new particles that decay into top quarks with a large branching fraction. These scenarios include Randall-Sundrum KK gluons [1] and any Z′ with Standard-Model-like couplings. If these new particles are sufficiently massive, the resulting top quarks are highly boosted, and may collapse into a single jet. It is therefore useful to develop reconstruction algorithms that attempt to distinguish these boosted top quarks from those produced in the generic QCD background from proton-proton collisions.

At the LHC, top quarks are usually tagged with their semi-leptonic decays, $t \to Wb \to \ell\nu b$. The W from the top cascade decay, however, most often decays hadronically (68% of the time). This note therefore addresses this difficult channel, so as to possibly benefit from the pertaining increased statistics.

The general strategy for tagging boosted top quarks decaying hadronically is to identify jet substructure in top-quark jets, and to use this substructure to impose kinematic cuts that discriminate against non-top jets of the same p_T. In particular, the masses involved in the process (the top mass and the W mass) provide powerful discrimination. This approach is possible because, while the top quark jet is highly boosted such that all of the resultant particles end up within one jet, the individual components of the top cascade decay may be still discernible (i.e. the b quark as well as the quarks from the W decay).

Many possible methods exist to reconstruct hadronic jets. In the CMS detector [2], jet energies are collected by two calorimeters, the electromagnetic calorimeter (ECAL) and the hadron calorimeter (HCAL), in towers of size (0.087,0.087) in the (η,ϕ) coordinates in the central barrel region. These energy deposits are usually clustered in jets by various algorithms, the most popular currently being the iterative cone algorithm [3] with a cone size of 0.5. This cone size is sufficiently large to include all the towers from a top cascade decay if the top momentum exceeds 800 GeV/c. Figure 1 shows a typical top quark jet (with $p_T = 800$ GeV/c) as seen by the CMS calorimeter towers. The substructure of this jet is clearly visible.

The algorithm developed in Ref. [4] is implemented to discern the jet substructure. This approach uses the Cambridge-Aachen (C-A or CA) jet algorithm [5] to reconstruct highly boosted top jets and decompose them into subjets. This decomposition is done by examining the cluster sequence of the final jets in the C-A algorithm to find intermediate clusters (defined as the "subjets") from the algorithm, and attempting to identify the jets from the top and W decays.

1.1 The Cambridge-Aachen Jet Clustering Algorithm

To perform the jet clustering, the FASTJET package is used [6]. This package provides an interface to many algorithms, including cone and sequential recombination (k_T-like) algorithms [3, 5, 7–10].

The C-A algorithm itself is a k_T-like algorithm. These algorithms examine four-vector inputs pairwise and construct jets hierarchically. To do so, they construct the quantities [6]

$$d_{ij} = \min(k^n_{T,i}, k^n_{T,j}) \frac{\Delta R^2_{ij}}{R^2} \quad (1)$$

$$d_{iB} = k^n_{T,i} \quad (2)$$

where $k_{T,i}$ is the transverse momentum of the i-th particle with respect to the beam axis, ΔR_{ij} is the distance between particles i and j in (y, ϕ) space (where y is rapidity, and ϕ is the azimuthal angle), and R is a distance parameter taken of order unity. For the k_T algorithm, $n = 2$. For the Cambridge-Aachen algorithm, $n = 0$ and $d_{iB} = 1$. For the anti-k_T algorithm, $n = -2$. The quantity d_{iB} is referred to as the "beam distance".

The algorithm then finds the minimum $d_{\min}$ of all the d_{ij} and d_{iB}. If $d_{\min}$ is a d_{ij}, the two particles are merged (by default, via a four-vector summation). If it is a d_{iB}, then the particle i is a final jet, and is removed from the list. This process is repeated until there are no particles left.

Physically, the differences between the three algorithms are contained in the momentum weighting. For the k_T algorithm, the weighting $(\min(k_{T,i}^2, k_{T,j}^2))$ is done so as to preferentially merge constituents with low transverse momentum with respect to their nearest neighbours [8],[9]. For the anti-k_T algorithm, the weighting $(\min(1/k_{T,i}^2, 1/k_{T,j}^2))$ is done so as to preferentially merge constituents with high transverse momentum with respect to their nearest neighbors. The approach of the anti-k_T algorithm is subtly different from the k_T approach, and results in jets that are roughly circular in the (y, ϕ) plane [10]. The C-A algorithm relies only on distance weighting with no k_T weighting at all [5].

Figures 2-3 show the jet transverse momentum and the jet mass, for generic QCD dijets with $600 < \hat{p}_T < 800$ GeV/c, simulated with PYTHIA[11] and the GEANT-based simulation of the CMS detector [12, 13]. The R parameter is taken to be 0.8 for all three algorithms. The jets that are selected have nearly identical transverse momenta and rapidity, however the masses differ very slightly.

Figure 4 shows the ΔR of the highest p_T subjet (Section 1.2) to the hard jet axis. The C-A algorithm selects the subjets closest to the hard jet axis. The anti-k_T has the second closest subjets, and the k_T algorithm has the furthest subjets.

Because the C-A algorithm is capable of discerning the components closest to the hard jet, it is therefore well-suited to discriminating softer subjets within harder jets.

1.2 The Top Tagging Algorithm

To construct boosted top jets, the C-A algorithm is used. These final C-A jets are hereby referred to as the "hard jets". The hierarchical clustering sequence of the construction of this jet is extracted, and the "grandparents" of the hard jet in the C-A algorithm are selected. In the declustering, soft clusters are ignored. These four grandparents are hereby referred to as "subjets".

In detail, the algorithm is as follows.

1. The input particles are clustered with C-A with a distance parameter of $R = 0.8$ into hard jets.
2. The hard jets are required to have $p_T > 250$ GeV/c, and rapidity $|y| < 2.5$.
3. The C-A clustering sequence for these jets is then used for decomposition as follows.
   (a) Primary decomposition : parent clusters
      i. If the two parent clusters satisfy the criterion $p_T^{\text{cluster}} > 0.05 \times p_T^{\text{hardjet}}$, then the clusters are considered as subjets, and the decomposition succeeds. The p_T cut on the cluster serves to remove low-p_T clusters from consideration. The 0.05 parameter is chosen to match that of Reference [4].

ii. If only one of the parent clusters satisfies the criterion $p_{\mathrm{T}}^{\mathrm{cluster}} > 0.05 \times p_{\mathrm{T}}^{\mathrm{hardjet}}$, then the decomposition process is repeated on the passed cluster, ignoring the constituents from the failed cluster. This decomposition is repeated until both clusters pass, both clusters fail, or the cluster consists of a single constituent.

iii. If, after this iterative process, there is no cluster with $p_{\mathrm{T}}^{\mathrm{cluster}} > 0.05 \times p_{\mathrm{T}}^{\mathrm{hardjet}}$, or the cluster is a single constituent, the decomposition fails and the jet is no longer considered.

iv. If two parent clusters are found that satisfy the criterion $p_{\mathrm{T}}^{\mathrm{cluster}} > 0.05 \times p_{\mathrm{T}}^{\mathrm{hardjet}}$ (define them as clusters A and B), then a secondary decomposition is performed to further resolve them.

(b) Secondary decomposition: grandparent clusters

i. The process described in the primary decomposition is repeated for each of the two parent clusters (clusters A and B). The fractional p_{T} cut on the clusters is, however, still taken with respect to the p_{T} of the total (hard) jet.

ii. If either cluster A or cluster B is decomposed into two clusters as described in the primary decomposition (clusters A' and A'', or B' and B'', respectively), then the hard jet is considered to be fully decomposed.

iii. A successfully decomposed jet yields either three or four clusters (the three possibilities are (A, B', B''), (A', A'', B), or (A', A'', B', B'').

iv. These three or four clusters are referred to as "subjets".

4. Finally, the following kinematic cuts are applied on the three or four subjets.

- The mass of the four-vector sum of the calorimeter towers of the hard jet (hereby referred to as the "jet mass") is required to be roughly consistent with the top mass: $100\ \mathrm{GeV}/c^2 < m_{\mathrm{jet}} < 250\ \mathrm{GeV}/c^2$.
- The three highest p_{T} subjets are taken pairwise, and the minimum invariant mass of those six pairwise candidates (hereby referred to as the "minimum pairwise mass" or m_{min}) is required to have $m_{\mathrm{min}} > 50\ \mathrm{GeV}/c^2$.

A view of the calorimeter energy deposits of a top jet and the corresponding subjets is displayed in Fig. 5 for a particular event with one top boosted with a p_{T} of around 250 GeV/c and an energy of around 660 $\mathrm{GeV/c^2}$.

2 Kinematic Discriminators

The motivation for the kinematic selection criteria applied in Section 1 is that highly boosted top jets, in some p_T range, are boosted enough to be reconstructed into a single jet, but still have resolvable components from the b and $W \to q\overline{q'}$.

The selection on the jet mass is justified because, in the case of true top jets, the jet mass tends toward the top mass, while for generic QCD non-top jets, the jet mass does not reconstruct to the top mass but instead approximately scales by the jet p_T over a constant of order 10.

The minimum pairwise mass of the subjets often reconstructs to the W mass. Figure 6 shows the true minimum mass pairing of the three partons from the $t \to Wb \to qqb$ decay for the Z' sample. It is most often the case that the minimum mass pairing of the "true" partons results in the W mass, which means that the b quark is most often the hardest parton in the event. Despite the fact that the lowest mass pairing of the subjets is not always the W mass after hadronization and reconstruction, the minimum mass pairing selection criterion is nonetheless exploited. The minimum mass pairing provides good discrimination against non-top jets, where there is no on-shell W and instead the minimum mass pairing of the subjets reconstructs to a low-mass falling spectrum. Figure 7 shows the minimum mass pairing versus the jet mass, to show the correlations. Due to the fact that the two variables are largely uncorrelated, two one-dimensional cuts are applied.

Figure 8 shows the number of subjets for top jets (solid lines) versus non-top jets (dashed lines) for $Z' \to t\bar{t}$ and generic QCD, respectively. The samples are chosen such that the reconstructed jets have approximately the same p_T. A Z' mass of 2 TeV/c^2 is used, to compare with generic QCD with $\hat{p}_T = 600$-800 GeV/c. The possible values are one, two, three, or four subjets, due to the fractional p_T cut of the subjets with respect to the hard jet ($p_T^{\text{subjet}} / p_T^{\text{hardjet}} > 0.05$).

Figures 9 and 10 show the jet mass, and minimum pairwise mass of the three subjets with the highest p_T, for top jets (solid lines) versus non-top jets (dashed lines) for the $Z' \to t\bar{t}$ and generic QCD, respectively. Also shown for the minimum mass distribution is the quantity $S/\sqrt{B}$ (dashed lines, in arbitrary units) which correspond to the right-hand-side axis. The jet minimum mass cut described in Section 1.2 is chosen to optimize $S/\sqrt{B}$ in Figure 10. The jet mass cut in Section 1.2 is chosen to be very loose and is not optimized.

Different jet algorithms are also examined that are related to C-A. Figures 10, 11 and 12 show the minimum invariant mass pairing for the Cambridge-Aachen, k_T and anti-k_T algorithms, respectively. The dashed line (with corresponding axis on the right) shows the quantity $S/\sqrt{B}$ for the three algorithms, again in arbitrary units. The discrimination in the C-A case is superior, with $S/\sqrt{B} = 2.4$ for C-A, 1.6 for k_T, and 1.3 for anti-k_T.

3 Algorithm Characterization

3.1 Data Samples and Event Selection

The following events have been produced:

- Several samples of QCD multi-jet events are generated with PYTHIA[11] in thirteen $\hat{p}_T$ bins, from 230 to 5000 GeV/c, for a total of 350k events.
- Samples of Z′ with mass 1, 2, 3, and 4 TeV/c^2, and widths of 1% and 10% of the mass, are generated with PYTHIA[11], for a total of 180k events.
- A total of one million events from the continuum $t\bar{t}$-plus-jets process is generated with MADGRAPH[14].

Unless otherwise noted, the events are generated with the GEANT-based simulation of the CMS detector, and the systematic studies are conducted with the fast simulation of the CMS detector [13].

The following event selection is applied

- Two jets with $p_T > 250$ GeV/c and $|y| < 2.5$ are required, corrected with absolute plus relative energy corrections [15].
- Both jets must satisfy the selection criteria described in Section 1.2.

3.2 Fake Tag Rate

Non-top decays may pass the selection defined in the previous section and thus fake a boosted top tag. In order to derive a parameterization of the fake tag rate, a data-driven method is proposed that makes use of a high statistics sample, and uses an "anti-tag and probe" method. This method is expected to provide over a thousand fake tags for a data sample of 100 pb^{-1}, allowing for a robust data driven determination of the fake background.

In detail, the following selection- orthogonal to that described in Section 1.2- is made for the fake tags:

- Two jets are required to have $p_T > 250$ GeV/c, and $|y| < 2.5$.
- Events are required to have one jet "anti-tagged". To "anti-tag", jets are selected that have two subjets or less, or to have more than two subjets, with jet mass and jet minimum mass outside the signal window described in Section 1.2.
- The other jets in the sample are referred to as the "probe" jets. The contamination from continuum $t\bar{t}$ production is subtracted based on an estimate from simulation, and the amount of that subtraction is taken as a systematic uncertainty. This "probe jet" selection constitutes an almost entirely signal-depleted sample.
- The tag rates are then parameterized with respect to the jet p_T using these "probe jets". The prediction from the simulation is taken as the central value and scaled to 100 pb^{-1}, assuming Poisson statistics and taking a binomial uncertainty.

Figure 13 shows the number of events and the fake tag parameterization as function of p_T for a 100 pb^{-1} data sample. These plots should be taken as a proxy for the real data. The results are fully data-driven in the real analysis with data, with the sole exception of the correction for the $t\bar{t}$ contamination. Even for a sample as low as 100 pb^{-1}, it is possible to reliably estimate the fake tag rate directly from the data, with an approximately 33% statistical uncertainty for jets with $p_T = 800$ GeV/c.

3.3 Efficiency

The efficiency of this boosted top algorithm is difficult to compute from data. There are several ideas of how to do so, however they all rely on continuum semileptonic $\mathrm{t\bar{t}}$ as a sample from which to estimate the efficiency, which has very low statistical precision in the kinematic region of interest ($p_\mathrm{T} > 250$ GeV/c) for 10 TeV collisions.

The total number of expected semileptonic $\mathrm{t\bar{t}}$ events in the muon channel is

$$N_{1\mu} = \sigma_{\mathrm{t\bar{t}}} \times \epsilon \times A \times \mathrm{BR}(\mathrm{W} \to \mu\nu) \times \int \mathcal{L}dt \tag{3}$$

where $\sigma_{\mathrm{t\bar{t}}}$ is the cross section, ϵ is the trigger times selection efficiency, A is the kinematic acceptance, BR is the branching ratio, and $\int \mathcal{L}dt$ is the integrated luminosity. The acceptance of the p_T cut is estimated to be 2.5%, and the efficiency to select jets of that p_T range with the top tagging algorithm is estimated to be 0.65% (both numbers are taken from simulation). Equation 3 therefore results in

$$N_{1\mu} = 443\ \mathrm{pb} \times 0.025 \times 0.0065 \times 0.11 \times 100\ \mathrm{pb}^{-1} = 0.8 \tag{4}$$

To have a statistical uncertainty of around 10%, at least 100 $\mu+$ tag events are needed. Thus, a luminosity around 12.5 fb^{-1} would be required to estimate the efficiency from data in 10 TeV collisions.

Instead, the efficiency is estimated from the simulation, for the early data. This approach is not entirely robust, and should therefore be taken as indicative of the performance. As more luminosity is accumulated (and as the centre-of-mass energy is increased), the data-driven method of estimating the efficiency will become more usable.

Several systematic effects in this simulation study can affect the estimate of the tagging efficiency, by changing the shower development, hence the profile of the subjets. Several effects are studied:

- effects from initial and final state radiation;
- effects from renormalization scale;
- effects from fragmentation.

Some of the parameters describing each of these effects are varied as indicated below, and the results on the efficiencies are shown in Table 1. For this systematic study, the fast simulation of the CMS detector [13] is used.

The following parameters were varied as per the suggestion of Ref.[16]:

- Λ_QCD is varied up and down by a factor 2;
- the maximum parton virtuality for space-like showers (representing initial state radiation) is varied up and down by a factor 2;
- the maximum parton virtuality for time-like showers (representing final state radiation) is varied up and down by a factor 4;
- the "a" and "b" parameters of the Lund model, the minimum Q^2 to radiate a gluon, and the width of the hadron p_T distribution (representing light parton fragmentation) are varied up and down by one standard deviation;

- the b and c quark fragmentations in the Peterson function are varied up and down by one standard deviation.

The total uncertainty from these effects is estimated to be 3.8%.

In order to account for the detector-based systematic uncertainties on the efficiency, the resolution of the subjets within the hard jets is derived from a simulation of $Z' \rightarrow t\bar{t}$ events with Z' mass values of 1000 and 3000 GeV/c^2. The partons from the $t\bar{t} \rightarrow W + b \rightarrow b + q + q'$ decay (i.e. the b, q, and q′) are matched to the closest reconstructed subjet. The response of the simulated calorimeter is then parameterized as function of the subjet transverse momentum. This parameterization is done to study the resolution of the transverse momentum, rapidity, and azimuthal angle. The jet resolutions are conservatively estimated to have an uncertainty of 10% for p_T[17] and 50% for angular resolution (from the differences between the RMS and Gaussian width of simulation fits).

This variation leads to an additional 3.3% systematic uncertainty from the p_T resolution smearing, and 2.9% systematic uncertainty from each of the angular resolution smearings.

Figure 14 shows the efficiency with simulation statistical uncertainties, as well as the total 6.5% systematic uncertainty, obtained from combining the theoretical (3.8%) and detector-based (5.3%) systematic uncertainties. Table 1 summarizes the systematic uncertainties from the various contributions.

Table 1: Effects of variation of several systematic uncertainties on the estimated efficiency from simulation.

Effect	Systematic Uncertainty (%)
Initial State Radiation	1
Final State Radiation	2
Renormalization Scale	3
Light Quark Fragmentation	< 1
Heavy Quark Fragmentation	< 1
Theoretical Uncertainty	3.8
Momentum Smearing + 10%	3.3
Azimuthal Smearing + 50%	2.9
Rapidity Smearing + 50%	2.9
Detector-Based Uncertainty	5.3
Total Systematic Uncertainty	6.5

4 Conclusions

The algorithm described in Ref. [4] has been implemented in CMS and has achieved similar rejection of non-top backgrounds as described in that paper.

The algorithm deals exclusively with hadronic decays of the W boson in the cascade decays of top quarks, and has made this channel accessible experimentally, due to its high rejection ($\approx$ 98% of jets with $p_{\mathrm{T}} = 600\,\mathrm{GeV}/c$) of non-top-quark boosted jets while retaining a high fraction of top-quark boosted jets ($\approx$ 46% of jets with $p_{\mathrm{T}} > 600\,\mathrm{GeV}/c$). This performance is comparable to that for bottom-quark jet-tagging algorithms at hadron colliders.

This algorithm can be used to examine resonance decays into $\mathrm{t\bar{t}}$ in the all-hadronic mode with comparable efficiency as a semileptonic decay analysis [18], while benefiting from the much higher branching ratio of the all-hadronic decay channel.

5 Acknowledgments

We would like to thank the authors of the theory paper in question, David E. Kaplan, Keith Rehermann, Matthew D. Schwartz, and Brock Tweedie, for helpful conversations.

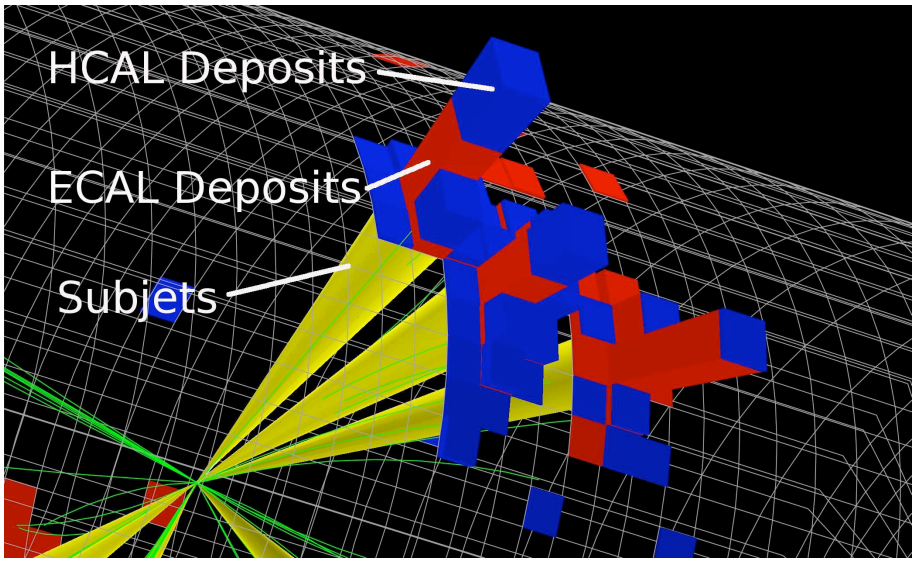


Figure 1: Reconstructed top-quark jet in cylindrical view with $p_{\mathrm{T}} = 800$ GeV/c. The cones represent the subjets. The HCAL and ECAL deposits, and the subjets are indicated on the figure.

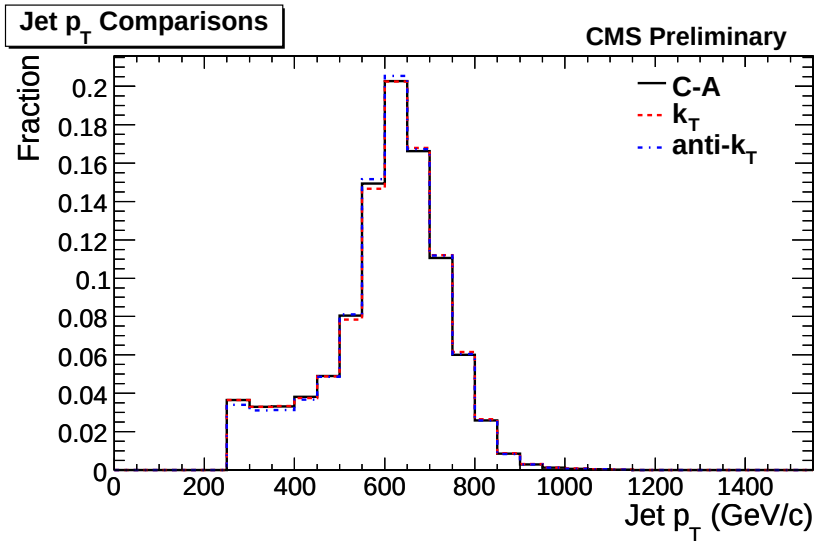


Figure 2: Comparison of the jet transverse momentum when using C-A, k_T, and anti-k_T jet clustering algorithms.

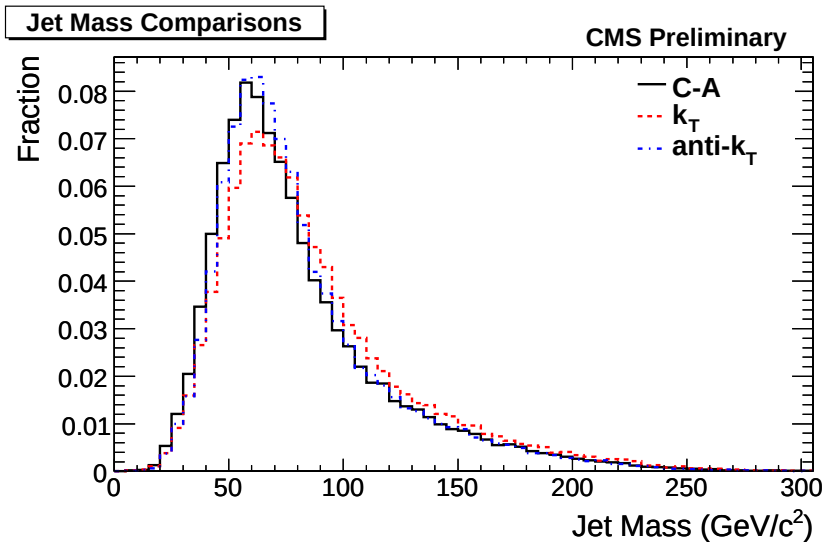


Figure 3: Comparison of the jet mass when using C-A, k_T, and anti-k_T jet clustering algorithms.

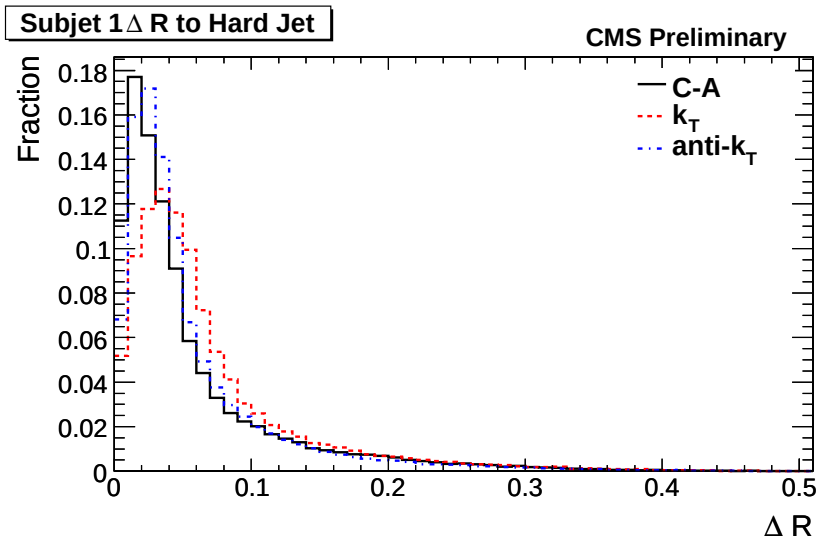


Figure 4: Comparison of the ΔR to the hard jet axis of the highest-p_{T} subjet when using C-A, k_{T}, and anti-k_{T} jet clustering algorithms.

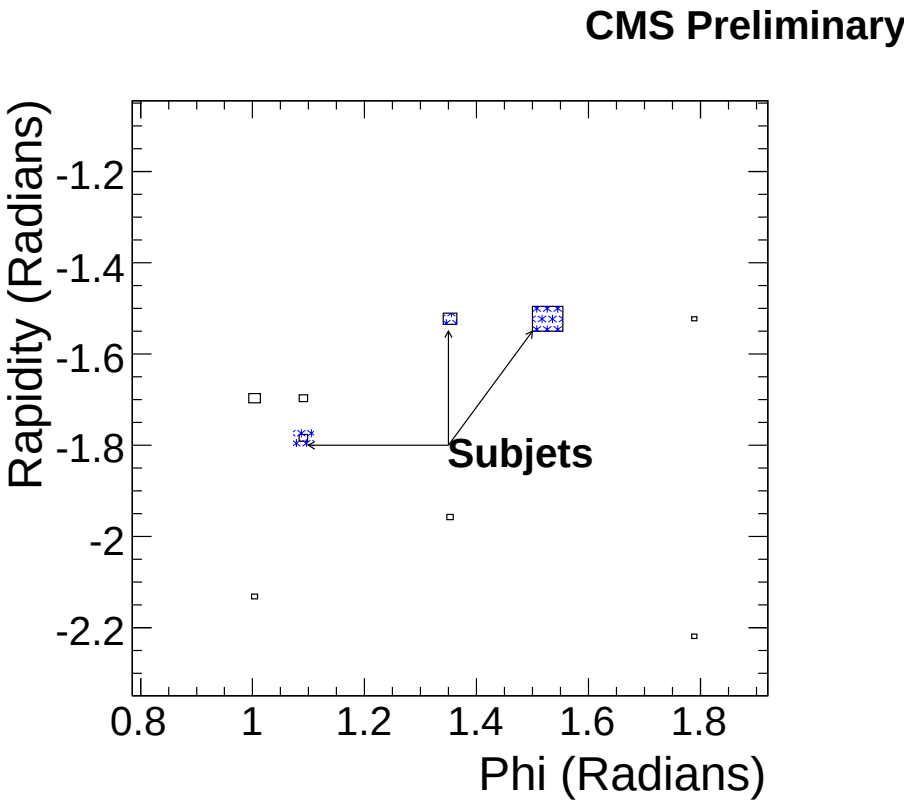


Figure 5: Subjet decomposition of a typical boosted top jet. The hollow boxes are the calorimeter towers in the jet. The starred boxes are the subjets found as described in Section 1.

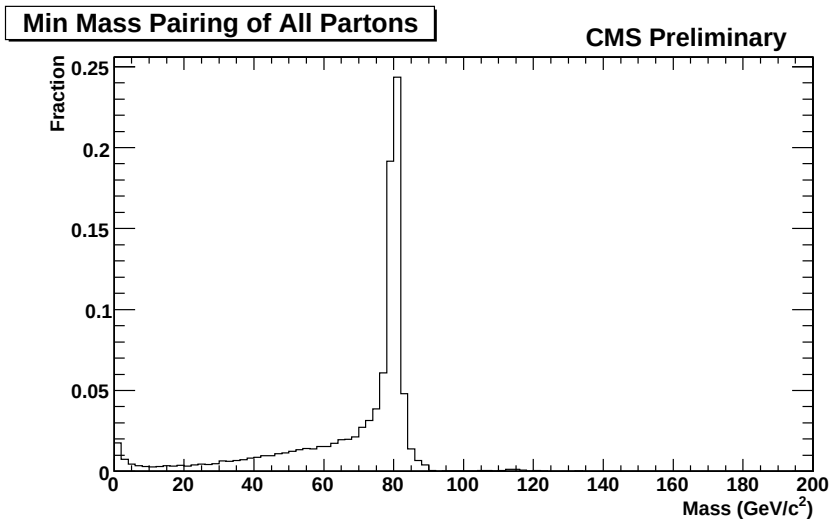


Figure 6: True minimum invariant mass pairing of partons from top cascade decays from a Z' with a mass of 2000 GeV/c^2 and width 20 GeV. Most often this reconstructs to the W mass.

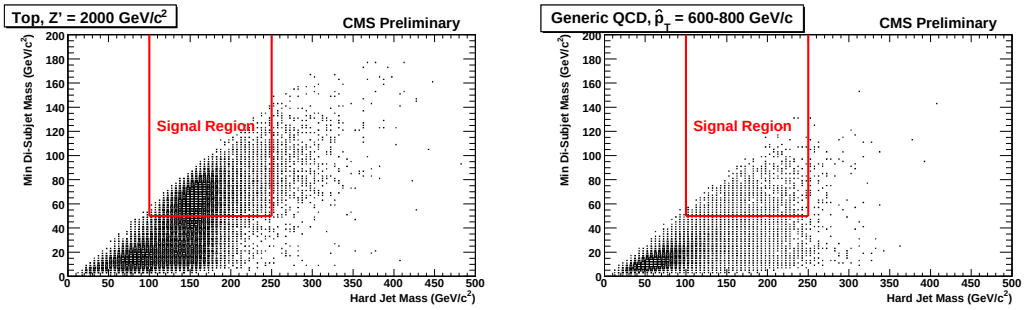


Figure 7: Jet mass (x axis) versus minimum di-subjet invariant mass (y axis) for the Z' sample with $M = 2000$ GeV/c^2 (left) and the QCD dijet sample with $\hat{p}_{\mathrm{T}} =$ 600-800 GeV/c (right).

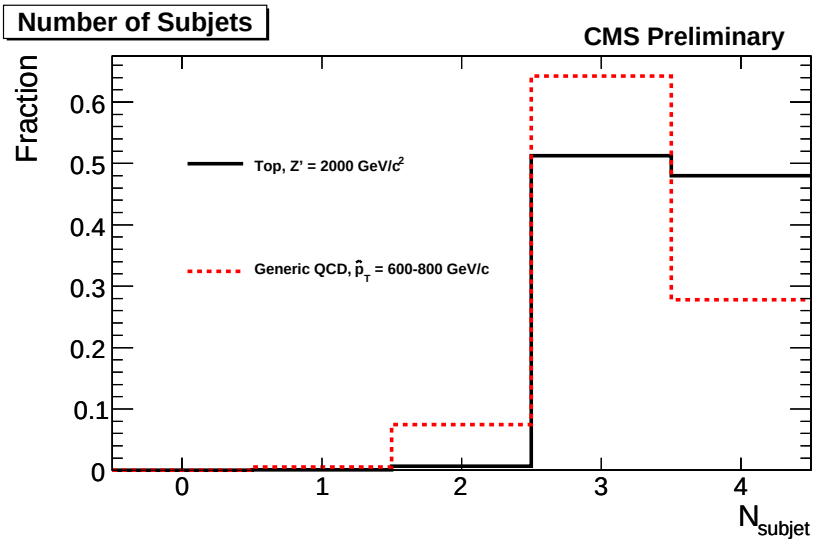


Figure 8: Number of subjets for top jets from $Z' \to t\bar{t}$ with $M = 2$ TeV$/c^2$ (solid line) versus non-top jets from generic QCD with $\hat{p}_T = 600$-800 GeV$/c$ (dashed lines). The samples are chosen such that the reconstructed jets have approximately the same p_T.

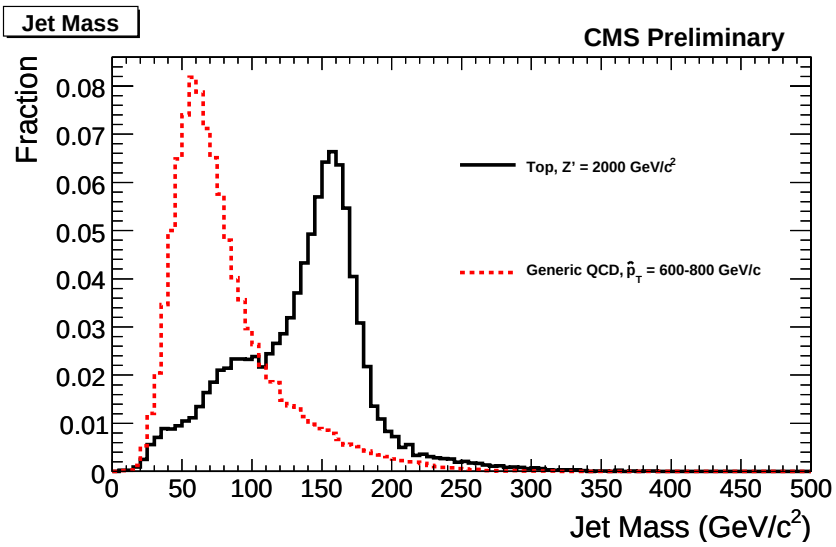


Figure 9: Jet mass for top jets from $Z' \to t\bar{t}$ with $M = 2$ TeV$/c^2$ (solid line) versus non-top jets from generic QCD with $\hat{p}_T = 600$-800 GeV$/c$ (dashed lines). The samples are chosen such that the reconstructed jets have approximately the same p_T.

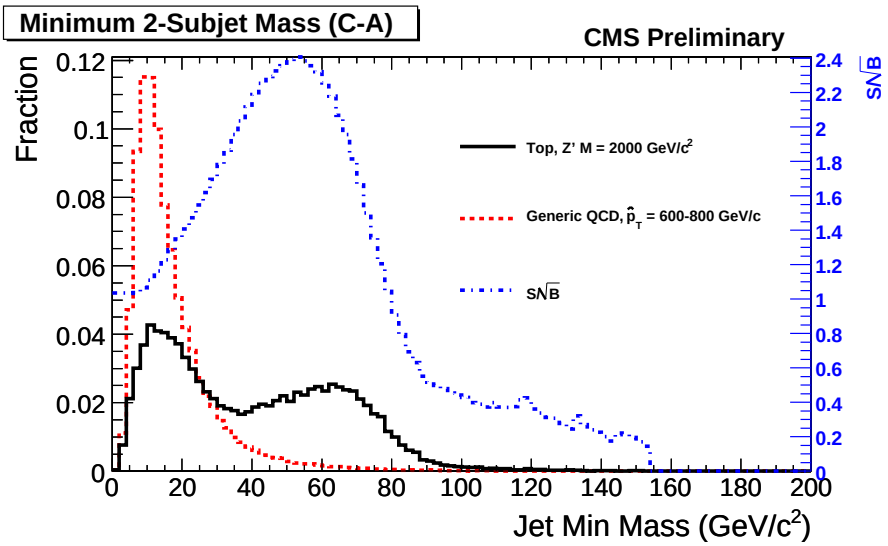


Figure 10: Minimum two-subjet invariant mass constructed with the Cambridge-Aachen clustering algorithm, plotted for top jets from $Z' \rightarrow t\bar{t}$ with $M = 2$ TeV$/c^2$ (solid line) versus non-top jets from generic QCD with $\hat{p}_T = 600$-800 GeV$/c$ (dashed lines). The samples are chosen such that the reconstructed jets have approximately the same p_T. Also shown is the quantity $S/\sqrt{B}$ in dashed lines, which corresponds to the right hand axis.

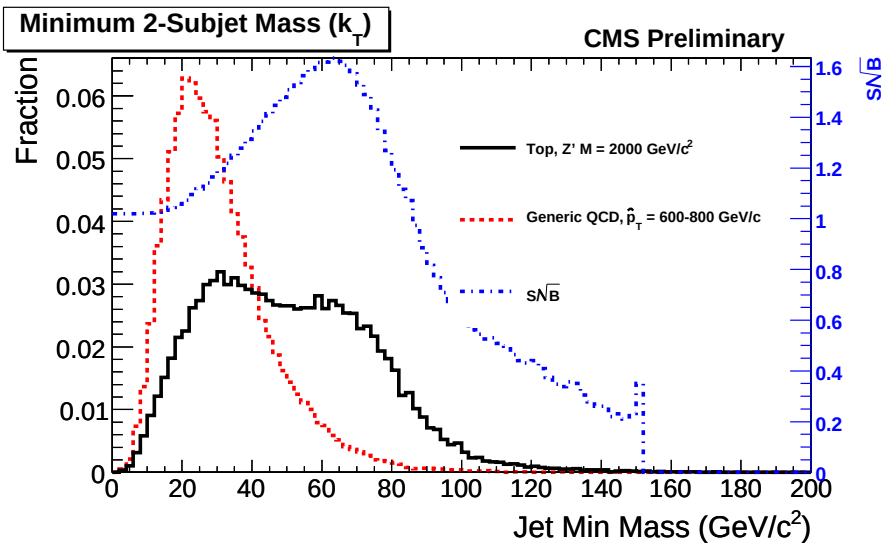


Figure 11: Minimum two-subjet invariant mass constructed with the k_T clustering algorithm, plotted for top jets from $Z' \rightarrow t\bar{t}$ with $M = 2$ TeV$/c^2$ (solid line) versus non-top jets from generic QCD with $\hat{p}_T = 600$-800 GeV$/c$ (dashed lines). The samples are chosen such that the reconstructed jets have approximately the same p_T. Also shown is the quantity $S/\sqrt{B}$ in dashed lines, which corresponds to the right hand axis.

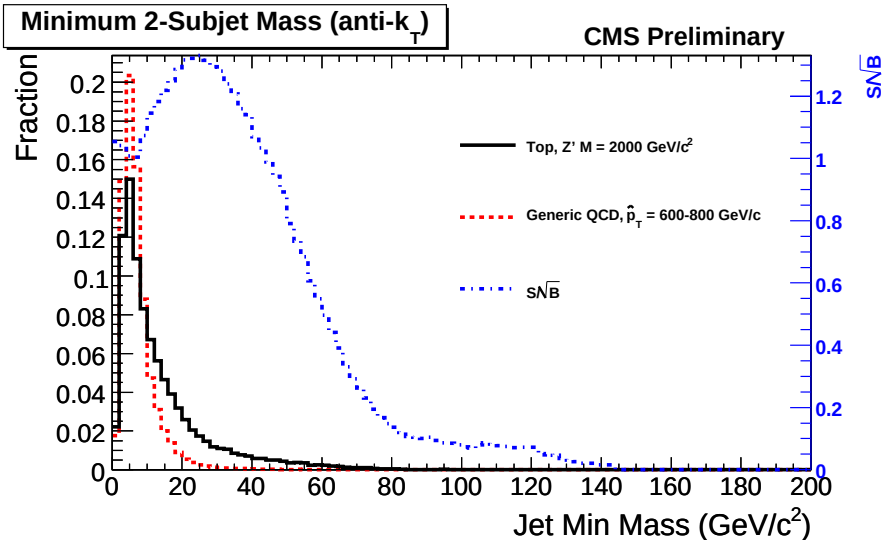
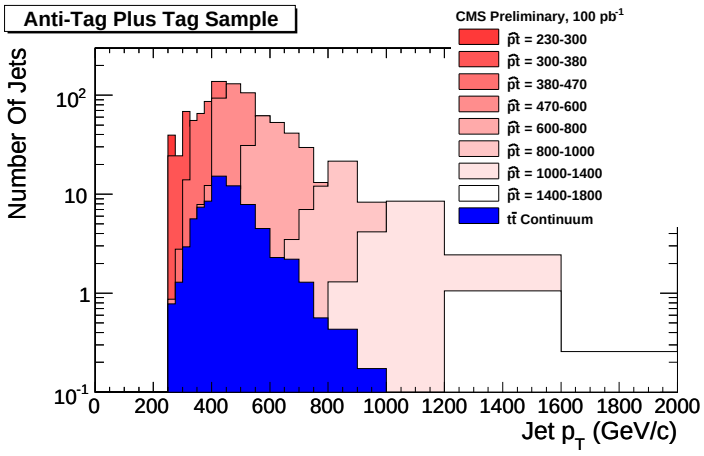
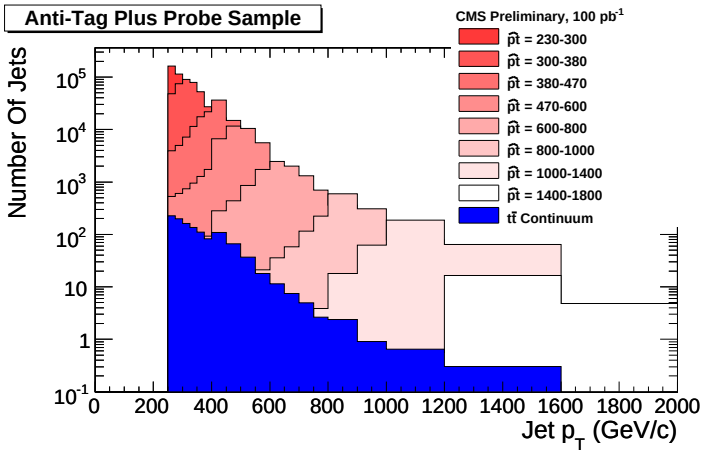


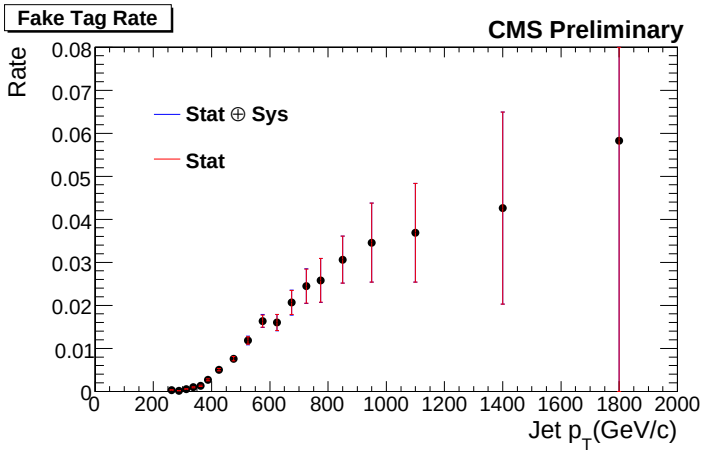
Figure 12: Minimum two-subjet invariant mass constructed with the anti-k_T clustering algorithm, plotted for top jets from $Z' \rightarrow t\bar{t}$ with $M = 2$ TeV$/c^2$ (solid line) versus non-top jets from generic QCD with $\hat{p}_T = 600$-800 GeV$/c$ (dashed lines). The samples are chosen such that the reconstructed jets have approximately the same p_T. Also shown is the quantity $S/\sqrt{B}$ in dashed lines, which corresponds to the right hand axis.



(a) Anti-Tag Plus Tag Sample



(b) Anti-Tag Plus Probe Sample



(c) Mistag Rate

Figure 13: Figure 13(a) shows the "anti-tag plus tag" sample (i.e. the numerator of the mistag rate), Figure 13(b) shows the "anti-tag plus probe" sample (i.e. the denominator of the mistag rate), and Figure 13(c) shows the fake Tag Parameterization versus jet p_{T} for a 100 pb^{-1} scenario. The central value is taken as the prediction from simulation, and the uncertainties are given as the expected statistical plus systematic uncertainties. The continuum $\mathrm{t\bar{t}}$ contribution is subtracted and this contribution taken as the systematic uncertainty.

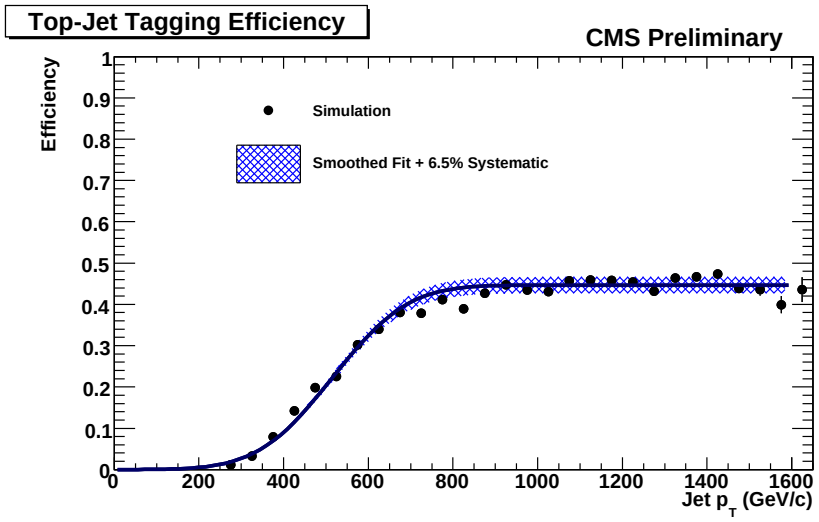


Figure 14: Efficiency for matched top-jets, including the 6.5% systematic uncertainty. The efficiency is fit to a functional form of $0.45 \times \frac{1}{2}\left(\mathrm{Erf}(\frac{p_T - 516}{197})\right)$.

References

[1] L. Randall and R. Sundrum, "*A Large Mass Hierarchy from a Small Extra Dimension* ," *Phys.Rev.Lett* **83:3370-3373** (1999).

[2] **CMS** Collaboration, S. C. et al, "*The CMS experiment at the CERN LHC*," *Journal of Instrumentation* **3** (2008), no. 08, S08004.

[3] G. C. Blazey et al., "*Run II jet physics*," *hep-ex/0005012* (2000) arXiv:hep-ex/0005012.

[4] D. Kaplan, K. Rehermann, M. Schwartz, and B. Tweedie, "*Top-tagging: A Method for Identifying Boosted Hadronic Tops*," *Phys.Rev.Lett* **101:142001** (2008).

[5] Y. Dokshitzer, G. Leder, S. Moretti, and B. Webber, "*Better Jet Clustering Algorithms*," *JHEP* **9708:001** (1997).

[6] G. S. Matteo Cacciari and G. Soyez, "*FastJet 2.3 User Manual*," *Phys. Lett. B* **641:57** (2006).

[7] G. Soyez, "*The SISCone and anti-kt jet algorithms*," (2008) arXiv:0807.0021.

[8] S. Cataniand, Y. L. Dokshitzerand, M. H. Seymour, and B. R. Webber, "*Longitudinally invariant K(t) clustering algorithms for hadron hadron collisions*," *Nucl. Phys. B* **406:187** (1993).

[9] S. D. Ellis and D. E. Soper, "*Successive combination jet algorithm for hadron collisions*," *Phys. Rev. D 48* **3160** (1993).

[10] M. Cacciari, G. Salam, and G. Soyez, "*The anti-kt jet clustering algorithm*," *JHEP* **0804:063** (2008).

[11] T. Sjostrand, S. Mrenna, and P. Skands, "*PYTHIA 6.4 Physics and Manual*," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

[12] **GEANT4** Collaboration, S. Agostinelli et al., "*GEANT4: A simulation toolkit*," *Nucl. Instrum. and Methods* **A506** (2003) 250–303.

[13] **CMS** Collaboration, "*CMS physics: Technical Design Report*". Technical Design Report CMS. CERN, Geneva, 2006.

[14] J. Alwall et al., "*MadGraph/MadEvent v4: The New Web Generation*," *JHEP* **09** (2007) 028, arXiv:0706.2334.

[15] **CMS** Collaboration, "*Plans for Jet Energy Corrections at CMS*," **CMS PAS JME-07-002** (2007).

[16] **CMS** Collaboration, "*Guidelines for the Estimation of Theoretical Uncertainties at the LHC*," **CMS Note 2005/013** (2005).

[17] **CMS** Collaboration, "*Data-driven Determination of the Jet Energy Resolution and Jet Reconstruction Efficiency at CMS*," **CMS PAS JME-09-007** (2009).

[18] **CMS** Collaboration, "*Search for TeV top resonances into jets plus muon*," **CMS PAS EXO-09-008** (2009).