Telescoping jets: Probing hadronic event structure with multiple R's

Yang-Ting Chien

Phys. Rev. D **90**, 054008 — Published 9 September 2014

DOI: 10.1103/PhysRevD.90.054008

Telescoping Jets: Probing Hadronic Event Structure with Multiple R's

Yang-Ting Chien
Center for the Fundamental Laws of Nature, Harvard University, Cambridge, MA 02138, USA

Jets at high energy colliders are complicated objects to identify. Even if jets are widely separated, there is no reason to define jets with the same size. A single reconstruction of each event can only extract a limited amount of information. While the Qjet algorithm gives multiple interpretations for each event using nondeterministic jet clustering, here we propose a simple, fast and powerful *deterministic* method to probe each event with multiple angular scales R's in the jet definition. Though qualitatively a different approach, the information we get from reconstructions at different angular scales can be analyzed in the framework of multiple event interpretations. We show that the statistical power of an analysis can be dramatically increased. In particular, we can have a 46% improvement, as compared to 28% using Qjets, in the statistical significance of the Higgs search with a Z boson ($ZH \to \nu\bar{\nu}b\bar{b}$) at the 8 TeV LHC. Our work implies that the wide angle radiation contains useful information, which is a key observation that is overlooked in both the conventional and the Qjet analyses.

Jets are manifestations of the underlying colored partons in hard scattering processes. In order to reconstruct hard processes and uncover physics at high energy, jets are key objects to identify in high energy collider experiments. The conventional way to identify jets is to use clustering algorithms [1–5], where a parameter R sets an *artificial* jet size: the constituents of each reconstructed jet are those particles within an angular scale R away from the jet direction. This is particularly true for the anti-k_T algorithm because it gives almost perfect cone jets in the calorimeter pseudorapidity-azimuthal angle (η-ϕ) plane. On the other hand, a jet is a distinct structure in its own right with many collinear particles. The width of the localized energy distribution of the jet in the η-ϕ plane is an independent quantity and should be distinguished from the parameter R (FIG. 1).

Because the formation of jets is quantum mechanical and probabilistic, the widths of jets are always different (FIG. 2). To reconstruct partonic kinematics we should pick a large enough R so that most of the radiation emitted by the partons is enclosed. However, with a large R more radiation contamination will be included. Many useful jet grooming techniques [7–10] have been introduced to get rid of contamination. Algorithms with a large R may also fail to resolve jets in some events. Multiple hard partons may be in a fat jet which potentially has substructure. Without looking into jet substructure we may incorrectly include irrelevant jets in event reconstruction. In the end an R is chosen for all events to optimize an analysis (see [11] for jets with variable R). A fixed R defines a single set of constituents for each jet and a single reconstruction for each event. A single choice of

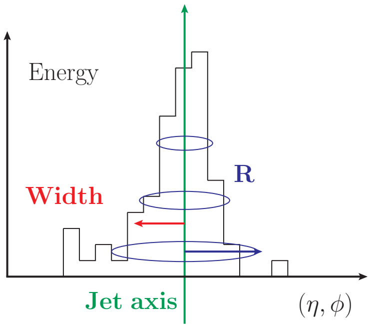

FIG. 1: A cartoon calorimeter plot distinguishing the width of the localized energy distribution of a jet (red) from the parameter R (blue) in the anti-k_T algorithm. R is an artificial distance scale introduced to define the calorimeter region we want to look at. The jet axis points in the direction of the dominant energy flow, and the precise direction is not essential here.

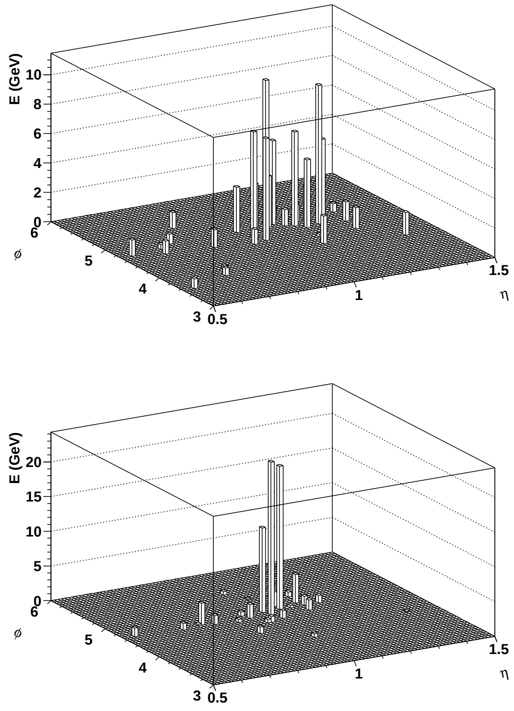

FIG. 2: Two b jets with the same partonic kinematics but different widths, wider (top) and narrower (bottom).

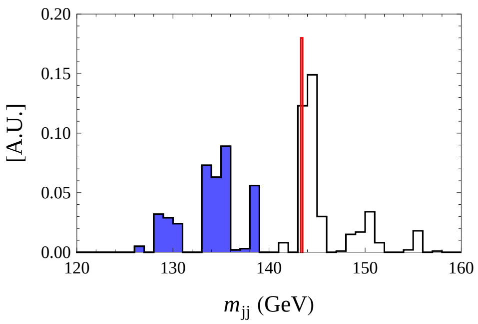

FIG. 3: The invariant mass distribution of the two b jets for a ZH event with multiple reconstructions using the telescoping jet algorithms (black). Using the anti-k_T algorithm with R=0.7, m_{jj}=143.4 GeV (red) which is outside the mass window of 110 GeV $< m_{jj} <$ 140 GeV in a conventional analysis. Multiple reconstructions reveal the ambiguity of this event and 37% of the reconstructions pass the cuts (blue).

R in conventional clustering algorithms may not be able to resolve jets and get most of the relevant radiation for *all* events.

Multiple event interpretations can provide extra information and increase the statistical power of an analysis. The Qjet algorithm [12] gives multiple event interpretations using nondeterministic jet clustering. Unlike conventional clustering algorithms, Qjets merge pairs of particles probabilistically according to an exponential weight, resulting in different clustering histories. An event may have a wide range of interpretations, and the non-deterministic nature of Qjets allows the correct event structure to emerge. It was shown that jet sampling with Qjets [13] can improve considerably the statistical significance $S/\delta B$ –the expected size of the signal divided by the background uncertainty– in many classes of analyses.

Instead in this paper we propose a simple, fast and powerful way to extract more information out of an event by probing jets with multiple R's in the jet definition. This idea is referred to as **telescoping jets** (Tjets). A straightforward application of Tjets is to use conventional clustering algorithms in each event multiple times with different R's. The range of R determines how different the event reconstructions can be. Note that, in this simplest approach, with a too-small R we may resolve an event in too much detail that miss its overall jet structure: in the $R \to 0$ limit particles are all jets. On the other hand, with a too-large R we may fail to resolve close jets: in the $R \to \infty$ limit the whole event is a jet. To deal with these issues and improve the above version of Tjets, we can try first identifying the dominant energy flows in an event. For example, by using the anti-k_T algorithm with a "suitable" R (usually smaller than the optimal R in a conventional physics analysis) to reveal the jet structure of an event and determine the jet axes from those reconstructed jet "cores". These axes should point in the directions of the dominant energy flow in an event, and the precise directions are not essential. We can also use the axes determined through a minimization procedure [20] and bypass using clustering algorithms. Then we define the jet constituents by the particles within a distance R away from the predetermined jet axes in the η-ϕ plane. That is, we telescope around some predetermined axes.

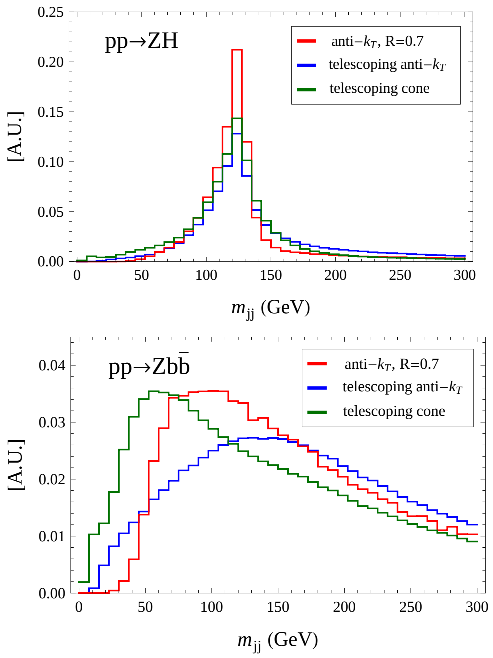

FIG. 4: The signal (top) and background (bottom) m_{jj} distributions reconstructed using the anti-k_T algorithm with R=0.7 (red), as well as the telescoping anti-k_T (blue) and cone (green) algorithms (definitions explained in text). With multiple event reconstructions we get a wider signal Higgs mass peak, but it reduces the statistical fluctuations of the m_{jj} distributions.

Another way of thinking about the above telescoping cone algorithm is that, we essentially move *down* the clustering sequence in the anti-k_T algorithm to build up jets after trying to identify the branch structure. This is complimentary to moving *up* the reclustered tree and looking for mass drops to identify the branches [6, 7]. Using different R's allows us to probe the energy distribution within each jet. The information can be analyzed in the framework of multiple event interpretations, and every observable of each event turns from a single number to a distribution (FIG. 3). Each event interpretation is now defined concretely and is labeled by R.

In the following we present the detailed procedure of the Tjets algorithm and apply it in a search for associated production of a Higgs and a Z where the Higgs decays to two b jets and the Z decays to $\nu\bar{\nu}$ ($ZH \to \nu\bar{\nu}b\bar{b}$). The background is $Z+b\bar{b}$ from $g \to b\bar{b}$. We require the events to pass a $\not{E}_T >$ 120 GeV cut which is among the experimentally available triggers. The $b\bar{b}$ system is slightly boosted so that the two b jets are closer to each other and more difficult to resolve. We define the signal window (specified later) by imposing cuts on the invariant

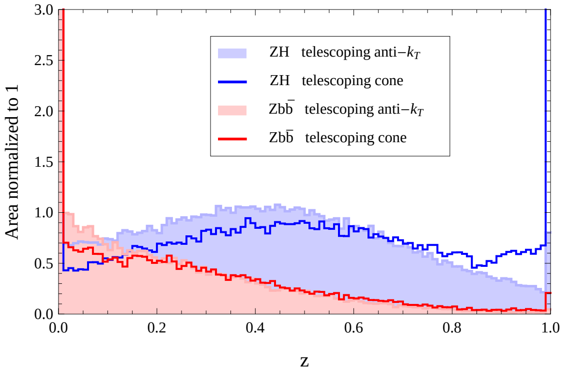

FIG. 5: The signal and background z distributions $\rho_S(z)$ and $\rho_B(z)$ using the telescoping anti-k_T and cone algorithms. z is the fraction of reconstructions of an event passing the experimental cuts. A large fraction of both signal and background events can be reconstructed differently.

mass of the two b jets m_{jj} (FIG. 4) and the transverse momentum of each b-jet in our analysis. Each event is counted by the fraction of reconstructions passing the cuts, instead of 0 or 1 in a conventional analysis. This increases the statistical stability of observables so that background fluctuations shrink considerably, which is the key for $S/\delta B$ improvement.

In the context of Higgs search in the $ZH \to \nu\bar{\nu}b\bar{b}$ channel, we first apply the simplest Tjet idea to reconstruct the two hardest jets using the anti-k_T algorithm with N different R's (referred to as telescoping anti-k_T algorithm in FIG. 4). The scaled-up computation time is tiny compared to using nondeterministic clustering algorithms [13]. Here we take N=100 which is more than enough. The value of R ranges from 0.2 to 1.5 with an increment $\frac{1.3}{N}$. The range of R is chosen because with the $\not{E}_T > 120$ GeV cut the angular separation between the two b jets will be roughly $\lesssim 2$. Here each reconstruction is weighted uniformly for simplicity.

The modified Tjets algorithm which telescopes around two predetermined axes (referred to as telescoping cone algorithm in FIG. 4) goes as follows:

- Use the anti-k_T algorithm with R=0.4 (can be further optimized in the analysis) to reconstruct the cores of the two hardest jets and determine the jet axes n_1 and n_2.

- Define the i-th jet to be the particles within a distance R away from n_i in the η-ϕ plane:

$$\text{jet}^i_R = \{\, p \mid (\eta_p - \eta_{n_i})^2 + (\phi_p - \phi_{n_i})^2 < R^2\}. \quad (1)$$

- In the case of overlapping jets, assign particles to the jet with the closer jet axis. This step is to avoid ambiguity and is not crucial when reconstructing the invariant mass of the two hardest jets m_{jj}.

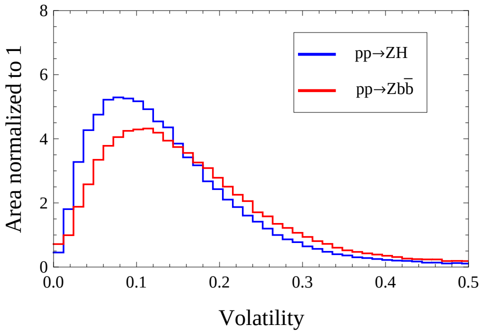

FIG. 6: The signal (blue) and background (red) volatility distributions using the telescoping cone algorithm.

Here we use the same R for both b jets in an event. However, for generic beyond the standard model physics searches with both quark and gluon jets in the final state, one can exploit the full idea of using different R's for different jets. We will leave these for future studies.

Our signal and background events were generated at the parton level using Madgraph 5 [15] and then showered with Pythia 6.4 [16] for the 8 TeV LHC. We impose the $\not{E}_T > 120$ GeV cut at the Madgraph level and the following cuts in the analysis to define the signal window:

- 110 GeV $< m_{jj} <$ 140 GeV

- Both p_Ts of the two hardest jets $>$ 25 GeV.

We use the anti-k_T algorithm implemented in Fastjet v3.0.0 [17, 18], and we perform the conventional analysis with R at the optimized value of R=0.7 (to be compared with the Tjet analysis). We then study how the statistical significance of the Higgs search is improved using Tjets and the framework of multiple event interpretations. With N event reconstructions m_{jj} turns from a single number to a distribution for each event. We define z to be the fraction of reconstructions passing the above cuts. FIG. 5 shows the z distributions $\rho_S(z)$ and $\rho_B(z)$ for signal and background. This is in contrast to the conventional analysis in which an event either passes the cuts or does not. With multiple event reconstructions we can gain more information about the degree of certainty of an event being signal-like. Weighting each event by z in the counting experiment helps improve the significance of the analysis.

Let ϵ and σ^2 be the mean and variance of the z distribution, and N_S and N_B be the expected numbers of signal and background events produced at the 8 TeV LHC. Then the significance is equal to

$$\frac{S}{\delta B} = \frac{N_S\ \epsilon_S}{\sqrt{N_B(\epsilon_B^2 + \sigma_B^2)}}. \quad (2)$$

A more detailed discussion about statistics can be found in [13, 14]. The *volatility* (FIG. 6) of each event is defined by $\mathcal{V}$=$\Gamma/\langle m\rangle$, where Γ and $\langle m\rangle$ are the standard deviation and mean of the m_{jj} distribution of each

R range	N	algorithm	weight	$S/\delta B\uparrow$
0.4 and 1.0	2	cone	z	14%
0.4 to 1.0	7	cone	z	20%
0.4 to 1.5	12	cone	z	26%
0.2 to 1.5	100	anti-k_T	z	20%
0.2 to 1.5	100	cone	z	**28**%
0.4 to 1.5	12	cone	ρ_S/ρ_B	38%
0.2 to 1.5	100	cone	ρ_S/ρ_B	**46**%

TABLE I: $S/\delta B$ improvements using telescoping jets with different ranges of R, numbers of interpretations N, jet algorithms for each reconstruction and weights in the counting experiment.

event with multiple reconstructions. Volatility is useful in distinguishing boosted W jets from their QCD background [12], and we will leave exploiting volatility in Higgs searches for future studies.

The performances of various Tjet implementations are summarized in TABLE I. The key for the $S/\delta B$ improvement is the shrink of background fluctuations, which comes from the rapid decrease of σ_B. For experimental studies with jet energy calibration depending on the parameter R, we try different ranges of R's and fewer reconstructions using the telescoping cone algorithm. Note that we can get half the improvement by using just two R's, and using 12 R's between 0.4 and 1.5 performs almost as good as using 100 R's between 0.2 and 1.5.

With $\rho_S(z)$ and $\rho_B(z)$ we can get an even larger improvement with the optimized weight $\frac{\rho_S(z)}{\rho_B(z)}$ [13] in the counting experiment. Then the significance is equal to

$$\frac{S}{\delta B}=\frac{N_S}{\sqrt{N_B}}\sqrt{\int_0^1\frac{\rho_S^2(z)}{\rho_B(z)}dz}\ , \tag{3}$$

and we get a 46% improvement compared to the conventional analysis. For R=0.4 to 1.5 with increment 0.1 we can get a 38% improvement with just 12 R's.

To conclude, the width of the localized energy distribution of a jet may not match well with the parameter R in jet algorithms. The situation is even more complicated for events with close jets because resolving jets becomes an issue when the parameter R and the distance between jets confront with each other. We explore a simple and promising way of extracting more information out of each event by probing it with multiple R's. The information at different angular scales can be analyzed in the framework of multiple event interpretations. The approach increases the statistical stabilities of observables which leads to remarkable improvement in the significance of a refined counting experiment. Telescoping jets open up the possibility of refining and improving jet physics analysis in high energy experiments.

Also, we only look at the transverse momenta and invariant mass of the two b jets, which are observables at high energy scales. It would be interesting to see how much more we can improve the significance of Higgs searches in hadronic channels by combining the analysis with other jet substructure [19–21] and color flow [22, 23] observables, which probe softer sectors of QCD and color connections in an event. The Tjets approach in this paper could potentially be extended and combined with likelihood ratio test and multivariate analysis, and in the presence of pile up our method will have to combine with jet grooming techniques. Applications of telescoping jets beyond physics searches, for example observable measurements, are also worth investigating. Probing jets with multiple R's may also allow us to construct jet observables more reliably.

The author would like to thank David Farhi, Marat Freytsis, Dilani Kahawala, David Krohn, Matthew Schwartz and Jessie Shelton for helpful discussions and comments on the manuscript. The work is supported by DOE grant DE-SC003916. All the computations were performed on the Odyssey cluster at Harvard University.

[1] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804**, 063 (2008) [arXiv:0802.1189 [hep-ph]].

[2] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP **9708**, 001 (1997) [hep-ph/9707323].

[3] M. Wobisch and T. Wengler, In *Hamburg 1998/1999, Monte Carlo generators for HERA physics* 270-279 [hep-ph/9907280].

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

[5] S. D. Ellis and D. E. Soper, Phys. Rev. D **48**, 3160 (1993) [hep-ph/9305266].

[6] D. E. Kaplan, K. Rehermann, M. D. Schwartz, B. Tweedie and , Phys. Rev. Lett. **101**, 142001 (2008) [arXiv:0806.0848 [hep-ph]].

[7] J. M. Butterworth, A. R. Davison, M. Rubin, G. P. Salam and , Phys. Rev. Lett. **100**, 242001 (2008) [arXiv:0802.2470 [hep-ph]].

[8] S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **81**, 094023 (2010) [arXiv:0912.0033 [hep-ph]].

[9] S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **80**, 051501 (2009) [arXiv:0903.5081 [hep-ph]].

[10] D. Krohn, J. Thaler and L. -T. Wang, JHEP **1002**, 084 (2010) [arXiv:0912.1342 [hep-ph]].

[11] D. Krohn, J. Thaler and L. -T. Wang, JHEP **0906**, 059 (2009) [arXiv:0903.0392 [hep-ph]].

[12] S. D. Ellis, A. Hornig, T. S. Roy, D. Krohn and M. D. Schwartz, Phys. Rev. Lett. **108**, 182003 (2012) [arXiv:1201.1914 [hep-ph]].

[13] D. Kahawala, D. Krohn and M. D. Schwartz, arXiv:1304.2394 [hep-ph].

[14] S. D. Ellis, A. Hornig, T. S. Roy, D. Krohn and M. D. Schwartz, In preparation.

[15] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, JHEP **1106**, 128 (2011) [arXiv:1106.0522

[hep-ph]].

[16] T. Sjostrand, S. Mrenna and P. Z. Skands, JHEP **0605**, 026 (2006) [hep-ph/0603175].

[17] M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012) [arXiv:1111.6097 [hep-ph]].

[18] M. Cacciari, G. P. Salam and G. Soyez, http://fastjet.fr/.

[19] S. D. Ellis, C. K. Vermilion, J. R. Walsh, A. Hornig, C. Lee and , JHEP **1011**, 101 (2010) [arXiv:1001.0014 [hep-ph]].

[20] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, Phys. Rev. Lett. **105**, 092002 (2010) [arXiv:1004.2489 [hep-ph]].

[21] J. Thaler and K. Van Tilburg, JHEP **1103**, 015 (2011) [arXiv:1011.2268 [hep-ph]].

[22] J. Gallicchio and M. D. Schwartz, Phys. Rev. Lett. **105**, 022001 (2010) [arXiv:1001.5027 [hep-ph]].

[23] A. Hook, M. Jankowiak and J. G. Wacker, JHEP **1204**, 007 (2012) [arXiv:1102.1012 [hep-ph]].