# Precision Jet Substructure from Boosted Event Shapes

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1103/PhysRevLett.109.092001">http://dx.doi.org/10.1103/PhysRevLett.109.092001</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Physical Society</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://hdl.handle.net/1721.1/73879">http://hdl.handle.net/1721.1/73879</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
</tbody>
</table>
Precision Jet Substructure from Boosted Event Shapes

Ilya Feige, Matthew D. Schwartz, Iain W. Stewart, and Jesse Thaler

1Center for the Fundamental Laws of Nature, Harvard University, Cambridge, Massachusetts 02138, USA
2Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 25 April 2012; published 30 August 2012)

Jet substructure has emerged as a critical tool for LHC searches, but studies so far have relied heavily on shower Monte Carlo simulations, which formally approximate QCD at the leading-log level. We demonstrate that systematic higher-order QCD computations of jet substructure can be carried out by boosting global event shapes by a large momentum $Q$ and accounting for effects due to finite jet size, initial-state radiation (ISR), and the underlying event (UE) as $1/Q$ corrections. In particular, we compute the 2-subjettiness substructure distribution for boosted $Z \rightarrow q\bar{q}$ events at the LHC at next-to-next-to-next-to-leading-log order. The calculation is greatly simplified by recycling known results for the thrust distribution in $e^+e^-$ collisions. The 2-subjettiness distribution quickly saturates, becoming $Q$ independent for $Q \approx 400$ GeV. Crucially, the effects of jet contamination from ISR/UE can be subtracted out analytically at large $Q$ without knowing their detailed form. Amusingly, the $Q = \infty$ and $Q = 0$ distributions are related by a scaling by $e$ up to next-to-leading-log order.

DOI: 10.1103/PhysRevLett.109.092001 PACS numbers: 12.38.Cy, 13.87.−a

The Large Hadron Collider (LHC) explores a new regime where the collision energy far exceeds the masses of known standard model particles. At such energies, heavy particles such as $W/Z$ bosons and top quarks are often produced with large Lorentz boost factors, which leave their hadronic decay products collimated into a single energetic “fat jet.” Jet substructure techniques extract information from these fat jets to distinguish boosted heavy objects from the QCD background of jets initiated by light quarks and gluons. Examples of variables defined for this purpose include planar flow [1,2], jet angularities [2], pull [3], $N$-subjettiness [4,5], dipolarity [6], and angular correlations [7], with applications to boosted Higgs bosons [8], top quarks [1,9], W’s [10], and quark versus gluon discrimination [11], along with many beyond the standard model applications (see [12,13] for recent reviews). Jet substructure measurements are underway at the LHC [14,15], but to date, studies of the analyzing power of substructure variables have been limited by the use of leading-log shower Monte Carlo simulations. If higher-order QCD computations were available, one could use them to directly compare to experiments or test the accuracy of Monte Carlo simulations.

In this Letter, we develop a framework for performing jet substructure computations analytically in the limit where the boosted object of interest has a large momentum $Q$. We find a mapping between global $e^+e^-$ event shapes—which have been calculated to high precision—and jet substructure variables in the large $Q$ limit, treating the finite jet size, initial-state radiation (ISR), and underlying event (UE) as $1/Q$ corrections. Concretely, we consider the jet substructure observable $N$-subjettiness $T^N$ [4], which is the subject version of the global event shape $N$-jettiness [16]. The ratio $T^N/T^{N-1}$ is a robust probe for $N$-prong decays [17] and compares favorably to other methods for boosted object identification.

Here, we focus on 1- and 2-subjettiness ($T^1$ and $T^2$), which are relevant for LHC searches involving $W/Z$ and Higgs bosons. We compute the distribution for the ratio $T^2/T^1$ from $Z \rightarrow q\bar{q}$ decays to next-to-next-to-next-to-leading-log ($N^3LL$) order, using ingredients from higher-order calculations of the classic $e^+e^-$ thrust event shape [18–24]. From a calculational point of view, the use of this ratio is crucial, since it has a finite limit when $Q \rightarrow \infty$. We will show that our full subjet distribution is equal to the global distribution generated by the $Z$ decay products, up to $1/Q$ power-suppressed corrections. The dominant hadronization corrections cause a shift that is encoded in a single $Q$-independent parameter. We compare our substructure calculation to PYTHIA 8.150 [25] tune 4C and also use PYTHIA to demonstrate that the effects from the jet boundary and external radiation (i.e., ISR and UE) are suppressed by $1/Q$, only entering at the 5% level for $Q \approx 400$ GeV.

We begin by considering a fat jet of size $R$ (clustered with anti-$k_T$ [26]) in a $pp$ collision event. This jet should contain most of the $Z$ decay products as well as some ISR/UE contamination. The jet momentum is $P_j^\mu = \sum_{i \in J} p_i^\mu$, where $j$ runs over the four-vector $p_j^\mu$ within the jet $J$. The jet boost $Q$ is defined as $Q = |P_j|$. To calculate $N$-subjettiness, we must specify a distance measure [4,16,17,27], and we use the geometric measure

$$T_N^\mu = \min_{n_1,n_2,\ldots,n_N} \sum_{j \in J} \min\{n_1 \cdot p_j,n_2 \cdot p_j, \ldots, n_N \cdot p_j\}. \quad (1)$$

Here, $n_j^\mu = (1, \vec{n}_j)$ are lightlike axes defined by the overall minimization. The minimum inside the sum partitions the jet’s constituents into subjet regions $J_1, \ldots, J_N$, defined by

0031-9007/12/109(9)/092001(5) 092001-1 © 2012 American Physical Society
the axes \( n^\mu \). For the \( N \)-jet process, \( J \) is replaced by the entire event.

For 1-jet, the \( T_1 \) is defined as the small component of the fat-jet momentum, \( T_1 = p^+ = n \cdot p_j \), which can also be written as the small component of the fat-jet momentum, \( T_1 = p^+ = n \cdot p_j \). If the jet contained all the \( Z \) decay products and nothing else, \( T_1 \) would depend only on the \( Z \) boson momentum \( P_Z^\mu \) as

\[
\hat{T}_1 = P_Z^+ = \sqrt{Q^2 + m_Z^2} - Q. \tag{2}
\]

Thus, the difference

\[
\Delta \tau = T_1 - \hat{T}_1 \tag{3}
\]

measures how much the \( Z \) is incorrectly reconstructed. We will use \( \Delta \tau \) to correct for ISR/UE contamination.

Turning to 2-jet, we first calculate the ratio \( T_2/T_1 \) including only the \( Z \) decay products and then discuss how other effects can be systematically included. The distribution for the \( Z \) decay products is easily determined by boosting the \( Z \) rest frame distribution. At leading order, \( Z \) decays to a \( q\bar{q} \) pair that goes off back-to-back in the rest frame at an angle \( \theta \) (the helicity angle) with respect to the boost axis, as in Fig. 1. For simplicity, we treat the \( Z \) as unpolarized with a flat \( \theta \) distribution, but one could easily integrate over a different \( \theta \) distribution, for example, for \( W \)'s coming from top decays [9]. In the boosted frame, the \( Z \) momentum \( P_Z^\mu \) and the two daughter-quark momenta \( q_1^\mu \) and \( q_2^\mu \) are

\[
P_Z^\mu = \{E_Z, 0, 0, Q\},
q_1^\mu = \frac{1}{2}(E_Q - Q \cos \theta, -m_Z \sin \theta, 0, Q - \cos \theta E_Q),
q_2^\mu = \frac{1}{2}(E_Q + Q \cos \theta, m_Z \sin \theta, 0, Q + \cos \theta E_Q),
\]

with \( E_Q = \sqrt{m_Z^2 + Q^2} \). The quark energies are \( E_1 = \frac{1}{2}(E_Q - Q \cos \theta) \) and \( E_2 = \frac{1}{2}(E_Q + Q \cos \theta) \).

For the relevant small \( T_2 \) region, the subjet directions from the minimization in Eq. (1) can be aligned with the leading-order quark directions [16]. Thus, we can take

\[
n^\mu = (1, 0, 0, 1), \quad n_1^\mu = \frac{1}{E_1} q_1^\mu, \quad n_2^\mu = \frac{1}{E_2} q_2^\mu, \tag{5}
\]

where \( n^\mu \) is the \( T_1 \) axis and \( n_1^\mu \) and \( n_2^\mu \) are the \( T_2 \) axes. In terms of the subjet masses \( m_i \) and energies \( E_i \),

\[
T_2 = p_1^+ + p_2^+ = \frac{m_i^2}{2E_1} + \frac{m_j^2}{2E_2}. \tag{6}
\]

In the large \( Q \) limit, \( E_1 \sim Q \sin^2(\theta/2) \), \( E_2 \sim Q \cos^2(\theta/2) \), and \( T_1 \sim m_Z^2/(2Q) \), while \( m_i \) are \( Q \)-independent. Thus, the distribution of the ratio \( T_2/T_1 \) asymptotes to a fixed \( Q \)-independent result.

Now let us consider how the scaling with \( Q \) is affected when \( T_2/T_1 \) is considered in a realistic environment, such as at the LHC. A measurement of \( T_2/T_1 \) includes effects from having a finite jet boundary and including radiation from elsewhere in the event. The jet boundary \( R \) identifies a \( Q \)-independent phase space region about the jet axis. As \( Q \to \infty \), the phase space for the \( Z \) decay products to land outside of the cone falls as \( 1/Q \). Hence, the jet boundary \( R \) is at most a \( 1/Q \) correction to \( T_2/T_1 \). The same conclusion holds if \( R \) is defined with a jet algorithm other than anti-\( k_T \).

Next, consider radiation not coming from the \( Z \) decay (i.e., ISR/UE). Since \( T_{1N} \) depends linearly on \( p_T^n \) in Eq. (1), both \( T_1 \) and \( T_2 \) will be distorted by (different) shifts due to this contaminating radiation. If we require the fat-jet mass to be close to \( m_2 \), then the shifts will scale as \( T_{1N} \), giving at most an \( O(1/Q) \) distortion of \( T_2/T_1 \). To turn this into a \( 1/Q \) distortion, note that the distribution of contaminating radiation is smooth over the fat jet, and at large \( Q \),

\[
n_1^\mu = n^\mu + \frac{m_Z}{Q} \left[-\cos \theta \tan \frac{\theta}{2} e^\mu + O(1/Q)\right], \tag{7}
\]

where \( e^\mu = (0, 1, 0, 0) \). Comparing \( n \cdot p_j \) and \( \min \{ n_1 \cdot p_j, n_2 \cdot p_j \} \), both \( T_1 \) and \( T_2 \) will be shifted in the same way up to \( 1/Q \) corrections. Hence, we can remove the leading effect of contamination with \( \Delta \tau \) from Eq. (3) by defining

\[
\tau_{21} = T_2 - \Delta \tau / T_1 - \Delta \tau. \tag{8}
\]

Here, \( \tau_{21} \) has two important properties: first, it is close to \( T_3/T_1 \) since \( \tau_{21} = T_2/T_1 \) if only the exact \( Z \) decay products are included; second, it is insensitive to jet contamination up to \( 1/Q \) corrections. It is crucial that the \( \Delta \tau \) correction be made experimentally on an event-by-event basis; if only the \( T_2/T_1 \) distribution is measured, then the contamination will not be a \( 1/Q \) correction. The subtraction can be improved further by replacing \( \Delta \tau \) with \( \Delta \tau' = \Delta \tau (1 - \frac{m_Z}{Q}) \) in the numerator of Eq. (8); the additional factor accounts for the average fractional difference between \( T_2 \) and \( T_1 \) for uncorrelated soft radiation. The above logic is also appropriate for event pileup.

To compute the \( \tau_{21} \) spectrum at leading order in \( 1/Q \), we calculate \( T_2/T_1 \) assuming that only the \( Z \) decay products are included in the fat jet. We then average over the angle \( \theta \). Using the correspondence with 2-jetness, the factorization formula for \( T_2/T_1 \) is [16]

FIG. 1 (color online). Kinematics of boosted \( Z \) decay.
where $\sigma_0$ is the tree-level cross section given by the $Z$ decay rate. Here, $H = H(m_2, \mu)$, $J(s_1, \mu)$, and $S(k_1, k_2, \{n_i\}, \mu)$ are, respectively, the $Z \to q \bar{q}$ hard function, inclusive jet function, and 2-jettiness soft function. $H$ and $J$ are known at $O(\alpha_s^2)$ [28,29]. For simplicity, we consider the narrow width limit, neglecting $O(1/\sqrt{m_2})$ corrections. We also neglect nonsingular corrections at $O(\alpha_s)$. These contribute less than 5% in the peak of the $\tau_{21}$ distribution and can be included following [23,24].

We now show that the 2-jettiness soft function $S$ can be related to the hemisphere soft function $S_{\text{hemi}}$—relevant for thrust and heavy jet mass—which is known perturbatively to $O(\alpha_s^2)$ [30,31]. The soft function is

$$S(k_1, k_2, n_1 \cdot n_2, \mu, \Lambda) = \frac{1}{N_c} \sum_{X_i} \delta(k_1 - n_1 \cdot P_1^T) \delta(k_2 - n_2 \cdot P_2^T) \times \langle 0 | \bar{P}_{n_2}^T Y_{n_2} | X_i \rangle \langle X_i | Y_{n_1}^T \bar{P}_{n_1}^* | 0 \rangle, \quad (10)$$

where the $Y$'s are lightlike Wilson lines and $P_{1,2}^T$ are the momenta of the subjects $J_{1,2}$ in the state $|X_i\rangle$. Rotational invariance implies that the subject directions appear only in the combination $n_1 \cdot n_2$, and the argument $\Lambda \equiv \Lambda_{\text{QCD}}$ is a reminder of nonperturbative corrections contained in $S$. The hemisphere case corresponds to $n_1 \cdot n_2 = 2$, so that $S_{\text{hemi}}(k_1, k_2, \mu, \Lambda) = S(k_1, k_2, 2, \mu, \Lambda)$. From Eq. (10), the partitioning into regions of 2-subjettiness is invariant under

$$\frac{1}{\sigma_0} \frac{d\sigma}{d\tau_{21}} = \frac{1}{N_c} H(m_2, \mu_H) U_{H}(m_2, \mu_H, \mu_J) \int d\tau_{21} \int \frac{d\cos \theta}{2} H(m_2, \mu_H) J(s_1, \mu_J) J(s_2, \mu_J) S_{\text{hemi}} \left( \frac{\mu_S}{\beta_0}, \alpha_s(\mu_S) \right) \left( \frac{\mu_J}{\beta_0} \right) \left( \frac{\mu_H}{\beta_0} \right), \quad (13)$$

Here, $S_{\text{h}}$ is the perturbative thrust soft function, and $H$, $J$, and $S_2$ are fixed-order expansions in $\alpha_s(\mu_H)$, $\alpha_s(\mu_J)$, and $\alpha_s(\mu_S)$, respectively. $U_H$ and $U_2$ are evolution kernels that sum $\alpha_s^N \ln^N \tau_{21}$ terms. See Ref. [23] for details.

The natural scale choices are

$$\mu_H = m_2, \quad \mu_J = \mu_Q \sqrt{\tau_{21}}, \quad \mu_S = \mu_Q \tau_{21}. \quad (14)$$

Here, $\mu_Q = \sqrt{1 + Q^2/(2m_2^2)}$ is an average over $\theta$ of $\hat{T}_1 \beta_0$ which appears in the large logarithms. For $Q = 0$, one has $\mu_Q = m_2$, while for $Q \to \infty$, one has $\mu_Q = m_2/\sqrt{2}$. We perform the $s_{1,2}$ and $z$ integrations in Eq. (13) analytically and the $\theta$ integral numerically.

The results for the $\tau_{21}$ distribution for various $Q$ are shown in Fig. 2. As anticipated, the curves rapidly approach a fixed distribution at large $Q$.

In Fig. 3, we show a comparison to a “baseline” PYTHIA distribution, where the effects of hadronization are included but the $Z$ width, finite cone size, and ISR/UE contamination have been turned off. For this comparison, we fix $\Phi = 700$ MeV to match the peak of the $Q = 0$ PYTHIA distribution, which allows us to compute the distribution for all $Q \neq 0$. In the tail of the distribution, there is excellent quantitative agreement. The accuracy of PYTHIA’s tail is somewhat artificial since it was tuned to closely related $e^+ e^-$ thrust data at $Q = 0$. Predictions in the peak region require additional nonperturbative corrections, which could be included following [24].

In Fig. 4(a), we show the effect of a finite $R = 1.0$ cone and jet contamination in PYTHIA, restricting our attention to jets whose mass is within a 10 GeV window of $m_2$. At large $Q$, the effect of an $R = 1.0$ cone is quite mild. While ISR/UE gives a large distortion to $T_2/T_1$, this is successfully corrected in $\tau_{21}$ by the $\Delta \tau$ in Eq. (8). With the $\Delta \tau \to \Delta \tau'$ replacement, we do even better. Using $\Delta \tau'$ for $Q = 1000$ GeV, the PYTHIA $\tau_{21}$ distribution with
$R = 1.0/\text{ISR/UE}$ is indistinguishable at the 2% level from the baseline distribution shown in Fig. 3. Thus, our analytic result agrees very well with the full PYTHIA distribution.

We use PYTHIA to verify that the effects we have neglected in our calculation are indeed $1/Q$ suppressed. In Fig. 4(b), we plot the Kolmogorov-Smirnov $D$ statistic between the baseline PYTHIA distribution and PYTHIA as the finite cone and ISR/UE effects are reinstated as a function of $Q$. The $D$ statistic measures the maximum fractional difference between the cumulant $\tau_{21}$ distributions. Both finite cone and the ISR/UE effects fall off as $1/Q$, and the corrections are $\lesssim 5\%$ for $Q \gtrsim 400$ GeV.

In the above calculation, we neglected the finite width of the $Z$ boson, which leads to $O(\Gamma_Z/m_Z)$ corrections that are independent of $Q$. As shown in Fig. 4(c), finite width has only a small effect on the baseline distribution. Including $\Delta\tau$ yields a larger effect, since Eq. (3) assumed that all deviations from the $Z$ pole were due to jet contamination and not $\Gamma_Z$. Nevertheless, we see in Fig. 4(c) that $\Delta\tau'$ still mitigates the effect of ISR/UE. Though beyond the scope of this Letter, one can directly calculate $\tau_{21}$ with finite width effects.

$\text{FIG. 2 (color online).}$ Results of the N$^3$LL analytic calculation for $\tau_{21}$ with $\Phi = 0$. The distribution saturates for $Q \gtrsim 400$ GeV.

$\text{FIG. 3 (color online).}$ Comparison of theory prediction (bands) for $\tau_{21}$ to baseline PYTHIA (histograms). The heavier (lighter) band is N$^3$LL (NNLL), with widths given by the factor of two variations of the hard, jet, and soft scales. Here, $\Phi = 700$ MeV. The arrows indicate the approximate range of validity of Eq. (13).

$\text{FIG. 4 (color online).}$ (a) Effects of the finite jet cone and ISR/UE in PYTHIA. The $\Delta\tau'$ correction mitigates ISR/UE jet contamination. (b) Fractional effect of adding the finite cone and ISR/UE to the PYTHIA baseline distribution. With the $\Delta\tau'$ correction, these effects are smaller than 5% for $Q \gtrsim 400$ GeV and scale as $1/Q$, as expected. (c) Effect of finite $Z$ width. (d) $e$ scaling between $Q = 0$ (thrust) and $Q = \infty$. 
It is interesting to explore analytically the $Q$ dependence of our $d\sigma/d\tau_{21}$ (dropping cone and ISR/UE effects and taking $\Phi = 0$) by considering two extreme cases. In the $Z$ rest frame $Q = 0$, $d\sigma/d\tau_{21}$ is equal to thrust $d\sigma/d\tau$. In the $Q \to \infty$ limit, $d\sigma/d\tau_{21}$ depends logarithmically on $\tau_{21}$ multiplied by various functions of the helicity angle $\theta$. Isotropically averaging over $\theta$, these logarithms behave as

$$\int d\cos \theta \frac{1}{2} \log^a \left( \frac{\tau_{21}}{\tau} \right) = \log^a \frac{\tau}{e} + O(\log^{a-2} \tau).$$

Thus, up to NLL order, the $Q \to \infty$ distribution is related to thrust by scaling by a factor of $e = 2.718 \ldots$

$$\frac{d\sigma}{d\tau_{21}}|_{Q \to \infty} = \frac{1}{e} \frac{d\sigma}{d\tau} (\tau = \tau_{21}/e).$$

This is demonstrated in Fig 4(d).

Our technique of treating the jet boundary and external radiation as $1/Q$ corrections can be readily generalized to color neutral objects with $N$-prong decays, and the known NNLL ingredients for the $N$-jettiness event shape [27] are a starting point for the calculation of $N$-subjettiness. It can also be used to compute the distribution of individual subjet masses $m_i$, which are directly accessible with the $N$-jettiness factorization theorem. Another straightforward generalization would be to incorporate massive final-state quarks as in $H \to b\bar{b}$. To treat colored objects like boosted top quarks (or to calculate the QCD background from light quark and gluon jets) requires understanding the effect of final-state radiation on substructure observables, and we anticipate that expanding about the $Q \to \infty$ limit will be fruitful in that context as well.

We thank M. Baumgart and F. Tackmann for collaboration at an early stage of this work. This work was supported by the U.S. Department of Energy (DOE) under Contracts No. DE-FG02-94ER40818, No. DE-FG02-05ER-41360, No. DE-FG02-11ER-41741, and No. DE-SC003916. I.F. is supported by NSERC of Canada.