Substructure of high-pT Jets at the LHC

Leandro G. Almeida, Seung J. Lee, Gilad Perez, Ilmo Sung & Joseph Virzi

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

YITP-SB-08-31

Substructure of high-p_T Jets at the LHC

Leandro G. Almeida[a], Seung J. Lee[a], Gilad Perez[a], George Sterman[a], Ilmo Sung[a], Joseph Virzi[b]

[a] *C. N. Yang Institute for Theoretical Physics*
Stony Brook University, Stony Brook, NY 11794-3840, USA

[b] *Lawrence Berkeley National Laboratory*
Physics Division, 1 Cyclotron Road, Berkeley, CA 94720, USA

Abstract

We study high-p_T jets from QCD and from highly-boosted massive particles such as tops, W, Z and Higgs, and argue that infrared-safe observables can help reduce QCD backgrounds. Jets from QCD are characterized by different patterns of energy flow compared to the products of highly-boosted heavy particle decays, and we employ a variety of *jet shapes*, observables restricted to energy flow within a jet, to explore this difference. Results from Monte Carlo generators and arguments based on perturbation theory support the discriminating power of the shapes we refer to as *planar flow* and *angularities*. We emphasize that for massive jets, these and other observables can be analyzed perturbatively.

Introduction. At the Large Hadron Collider (LHC), events with highly-boosted massive particles, tops [1], W, Z and Higgs, h [2] may be the key ingredient for the discovery of physics beyond the standard model [3, 4, 5]. In many decay channels, these particles would be identified as high-p_T jets, and any such signal of definite mass must be disentangled from a large background of light-parton ("QCD") jets with a continuous distribution. This background far exceeds such signals, and relying solely on jet mass as a way to reject QCD background from signal would probably not suffice in most cases [6], even using a narrow window for dijets in the search for massive particles such as tops, produced in pairs.

In this paper, we argue that for massive, high-p_T jets, infrared (IR) safe observables may offer an additional tool to distinguish heavy particle decays from QCD background, perhaps even on an event-by-event basis. We will refer to inclusive observables dependent on energy flow within individual jets as *jet shapes*. A jet within a cone of radius 0.4, for example, may be identified from the energy recorded in roughly fifty 0.1×0.1 calorimeter towers. It thus contains a great deal of information. Perturbatively-calculable, infrared safe jet shapes combined, of course, with IR-safe jet finding algorithms [7], may enable us to access that information systematically, and to form a bridge between event generator output and direct theory predictions.

Essentially, any observable that is a smooth functional of the distribution of energy flow among the cells defines an IR-safe jet shape [8]. The jet mass is one example, but a single jet may be analyzed according to a variety of shapes. In particular, the jet mass distribution has large corrections when the ratio of the jet mass to jet energy is small [9], but can be computed at fixed order when the logarithm of that ratio has an absolute value of order unity. Once the jet mass is fixed at a high scale, a large class of other jet shapes become perturbatively calculable with nominally small corrections. Indeed, a jet whose mass exceeds $\mathcal{O}(100\,\mathrm{GeV})$ becomes, from the point of view of perturbation theory, much like the final state in leptonic annihilation at a similar scale. At such energies, event shapes, which in the terminology of this paper are jet shapes extended over all particles, have been extensively studied in perturbation theory [10]. In this study we explore the possibility that perturbative predictions for jet shapes differ between those jets that originate from the decay of heavy particles, and those which result from the showering of light quarks and gluons. Very interesting related studies have recently appeared in [4, 11, 12].

Jet Shapes and Jet Substructure. We would like to identify jet shapes for which perturbative predictions differ significantly between QCD and other high-p_T jets, focusing on relatively narrow windows in jet mass. In our companion paper [6] we have discussed how to calculate the jet mass distribution for the QCD background. We now extend this argument to the computation of other jet shape observables.

We emphasize that, because the observables under consideration are IR-safe, we may calculate them as power series in α_s, starting at first order for the QCD background, and zeroth order for an electroweak decay signal.

Our approximation for the jet cross sections is based on factorization for the relatively-collinear partons that form a jet from the remainder of process [9]. For a jet of

cone size R, contributions that do not vanish as a power of R are generated by a function that depends only on the flavor of the parent parton, its transverse momentum, and the factorization scale. Denoting an jet shape by e, we then have,

$$\frac{d\sigma}{dm_J\, de} = \sum_c \int_{{p_T}^{min}}^{\infty} dp_T \frac{d\hat{\sigma}_c(p_T)}{dp_T} \frac{dJ_c(e, m_J, p_T, R)}{de}\,, \tag{1}$$

where $d\hat{\sigma}/dp_T$ includes the hard scattering and the parton distributions of the incoming hadrons, and where the jet function for partons c in the final state is defined formally as in Refs. [6, 13].

In Ref. [6], we have found that the distribution of QCD jet masses in the range of hundreds of GeV is fairly well described by the jet function in Eq. (1) at order α_s, based on two-body final states. It thus seems natural to anticipate that for QCD jets, energy flow inside the cone would produce a *linear* deposition in the detector [4, 5, 11, 12, 14]. While this is certainly the case for an event consisting of two sub-jets, it is a simpler condition, and more easily quantified. Indeed, such a linear flow should also characterize jets from the two-body decay of a highly-boosted, massive particle. We will see below that relatively simple jet shapes can help distinguish QCD jets from many top-decay jets that involve three-body decay. We will also see that jet shapes can help separate samples that contain both QCD jets and jets from two-body decays, such as those of the W, Z or h. We emphasize that a single event may be analyzed by a variety of jet shapes, so that the resolution associated with each one need not be dramatic, so long as they are effectively independent.

Top decay and planar flow. The linear flow of QCD jets at leading order should be compared with a $\geq$ 3-body decay where the energy deposition tends to be *planar*, covering a two-dimensional region of the detector. An IR-safe jet shape, which we denote as *planar flow*, a two-dimensional version of the "D parameter" [15, 16, 17, 18], distinguishes planar from linear configurations. The utility of a closely-related observable was emphasized in Ref. [12].

Planar flow is defined as follows. We first construct for a given jet a matrix I_w as

$$I_w^{kl} = \frac{1}{m_J} \sum_i w_i \frac{p_{i,k}}{w_i} \frac{p_{i,l}}{w_i}\,, \tag{2}$$

where m_J is the jet mass, w_i is the energy of particle i in the jet, and $p_{i,k}$ is the k^{th} component of its transverse momentum relative to the axis of the jet's momentum. Given I_w, we define Pf for that jet as

$$Pf = \frac{4\det(I_w)}{\mathrm{tr}(I_w)^2} = \frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}\,, \tag{3}$$

where $\lambda_{1,2}$ are the eigenvalues of I_w. Pf vanishes for linear shapes and approaches unity for isotropic depositions of energy.

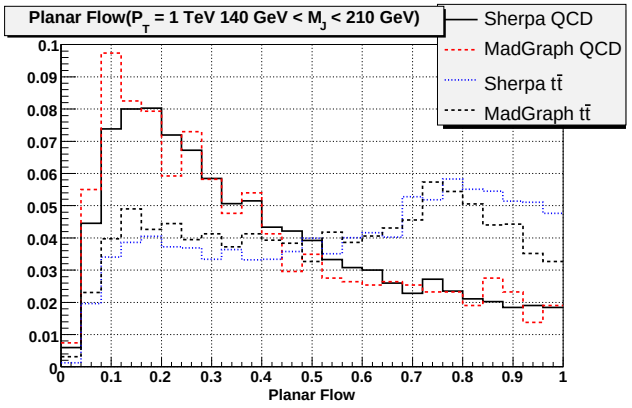

FIG. 1: The planar flow distribution for QCD and top jets obtained from MadGraph and Sherpa. Distributions are normalized to unit area.

Jets with pure two-body kinematics have a differential jet function fixed at zero planar flow,

$$\frac{1}{J}\left(\frac{dJ}{dPf}\right)_{\rm 2body} = \delta(Pf)\,. \tag{4}$$

This would apply to leading order for events with highly boosted weak gauge boson, Higgs and QCD jets. On the other hand, events that are characterized by $\geq$ 3-body kinematics have a smooth distribution.

Realistic QCD jets have, of course, nonzero Pf. Because Pf is an IR safe observable, however, its average value can depend only on the hard momentum scales of the jet, that is, m_J and p_T. This suggests an average Pf of order $\alpha_s(m_J) \sim 0.1$ for high jet masses, times at most logarithms of that are order unity for these heavy jets. Correspondingly, higher orders corrections should, by analogy to two-jet event shapes [10], replace the delta function of Eq. (4) with a differential distribution that peaks near the origin and then falls off. For jets resulting from three-body decay, on the other hand, we anticipate that corrections in α_s shift the already-smooth distribution modestly, without affecting its overall shape. Finally, for the vast majority of high-p_T QCD jets, with masses $m_J \ll p_T$, planarity corrections associated with multi-gluon emission may be expected to be large, and to shift Pf to order unity.

These considerations are confirmed in Fig. 1, where we show the Pf distribution for QCD jet and hadronic $t\bar{t}$ events, for $R = 0.4$, $p_T = 1000\,\mathrm{GeV}$ and $m_J = 140 - 210\,\mathrm{GeV}$ as obtained from MadGraph [19] and Sherpa [20] with jet reconstruction via (the IR-safe algorithm) SISCone [7]. We see that QCD jets peak around small values of Pf, while the top jet events are more dispersed. A planar flow cut around 0.5 would clearly remove a considerably larger proportion of QCD jets than top jets. Correspondingly, we have confirmed by event

generator studies that low-mass QCD jets have much larger planar flows that those in Fig. 1.

Two-body decay. While planar flow can help enrich samples with characteristically three- and higher-body kinematics, other jet shapes can also provide additional information on events with relatively low Pf. Here, we will still wish to distinguish the QCD background from highly boosted electroweak gauge bosons or Higgs [14] as well as from top jets whose Pf happens to be relatively low. We begin with jets that are linear at lowest order, and identify a set of jet shapes that have some power to distinguish between the two. Fixing p_T, R and m_J leaves only one free parameter to characterize the shape.

The QCD jet function for two-body kinematics is defined as a matrix element in [6] and is readily expressed as an integral over θ_s, the angle of the softer particle relative to the jet momentum axis. For a quark jet, for example, the integrand is therefore the differential jet function,

$$\frac{dJ^{QCD}}{d(c\theta_s)} = \frac{\alpha_s C_F \beta_z z^2}{4\pi m_J^2 (1-\beta_z c\theta_s)(2(1-\beta_z c\theta_s)-z^2)} \times \left[\frac{(2(1+\beta_z)(1-\beta_z c\theta_s)-z^2(1+c\theta_s))^2}{z^2(1+c\theta_s)(1-\beta_z c\theta_s)} + 3(1+\beta_z) + \frac{z^4(1+c\theta_s)^2}{(1-\beta_z c\theta_s)(2(1+\beta_z)(1-\beta_z c\theta_s)-z^2(1+c\theta_s))} \right], \quad (5)$$

where $z \equiv m_J/p_T$, $\beta_z \equiv \sqrt{1-z^2}$ and $c\theta_s \equiv \cos\theta_s$. The jet mass function is obtained by the integral $\int_{\theta_m}^{R} d\theta_s \, (dJ/d\theta_s)$, where θ_m is the angle with the smallest possible value of the softest particle, $\theta_m = \cos^{-1}\left(\sqrt{1-z^2}\right)$, at which both particles have the same energy and angle to the axis.

For signal events from a highly-boosted massive gauge bosons, we consider separately the cases when it is longitudinal and when it has helicity ($h = \pm 1$),

$$\frac{dJ^{\rm Long}}{d(c\theta_s)} = \frac{C}{(1-\beta_z c\theta_s)^2}, \quad (6)$$
$$\frac{dJ^{h=\pm 1}}{d(c\theta_s)} = \frac{C}{(1-\beta_z c\theta_s)^2}\left(1 - \frac{(zs\theta_s)^2}{2(1-\beta_z c\theta_s)^2}\right),$$

where $s\theta_s \equiv \sin\theta_s$ and C is a proportionality coefficient, totally determined from the two-body decay kinematics. We can interpret the appropriately normalized differential jet functions, $P^x(\theta_s) = (dJ^x/d\theta_s)/J^x$ as the probability to find the softer particle at an angle between θ_s and $\theta_s + \delta\theta_s$. As the ratio z decreases, the decay products become boosted and the cone shrinks. For QCD jets from light partons, however, this shrinkage is much less pronounced. Plots of these jet functions show that the gauge boson distributions of Eq. (6) fall off with θ_s faster than do QCD jets, Eq. (5). This observation suggests that the signal (vector boson-jet) and background (QCD jets) have different shapes for fixed p_T, R and jet mass. This may be used to obtain an improved rejection power against background events. We now consider a class of jet shapes, *angularities*, originally introduced in Ref. [13, 21] for two-jet events in e^+e^- annihilation. A natural generalization of these jet shapes to single-cone jets of large mass m_J is

$$\tilde{\tau}_a(R, p_T) = \frac{1}{m_J}\sum_{i\in jet}\omega_i \, \sin^a\left(\frac{\pi\theta_i}{2R}\right)\left[1-\cos\left(\frac{\pi\theta_i}{2R}\right)\right]^{1-a}, \quad (7)$$

with m_J the jet mass. The arguments of the trigonometric functions vary from zero to $\pi/2$ as θ increases from zero to R, that is, over the size of the cone. These weights revert to the angularities as defined in for leptonic annihilation in [13, 21] when $R = \pi/2$, so that the cone is enlarged to a hemisphere and m_J is replaced by the center-of-mass energy in a two-jet event. For massive jets, the angularities are clearly non-zero at lowest order, in contrast to the lowest order planar flow, Eq. (4). Then, precisely because their IR safety, higher-order corrections to the τ_a distributions should be moderate.

As the parameter a varies, the angularities give more or less weight to particles at the edge of the cone compared to those near the center. From the differential jet

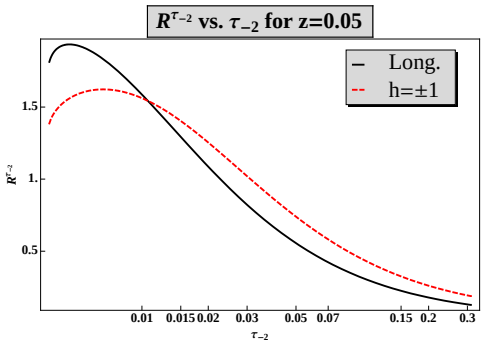

FIG. 2: The ratio between the signal and background probabilities to have jet angularity $\tilde{\tau}_{-2}$, $R^{\tilde{\tau}_{-2}}$.

distribution functions in Eqs. (5) and (6) and the definition of $\tilde{\tau}_a$ we can obtain the expression for $P^x(\tilde{\tau}_a)$ [as before x= sig (signal) or QCD] the probability to find a jet with with an angularity value between $\tilde{\tau}_a$ and $\tilde{\tau}_a + \delta\tilde{\tau}_a$ at fixed p_T, R, m_J and a. Our focus is not on the form of the individual distributions but rather on the ratio of the signal to background

$$R(\tilde{\tau}_a) = \frac{P^{\rm sig}(\tilde{\tau}_a)}{P^{\rm QCD}(\tilde{\tau}_a)}. \quad (8)$$

In Fig. 2 we show $R^{\tilde{\tau}_a}$ for $a = -2$ and $z = 0.05$, for the different vector boson polarizations. In Fig. 3 we show the corresponding angularity distributions at the event generator level, comparing the output of MadGraph for longitudinal Z boson production to QCD jets in the same mass window. The pattern suggested by the lowest-order

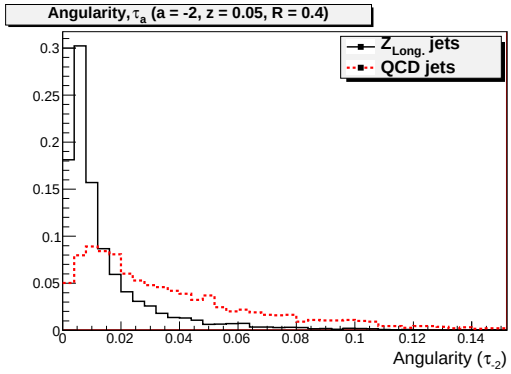

FIG. 3: The angularity distribution for QCD (red-dashed curve) and longitudinal Z (black-solid curve) jets obtained from MadGraph. Both distributions are normalized to unit area.

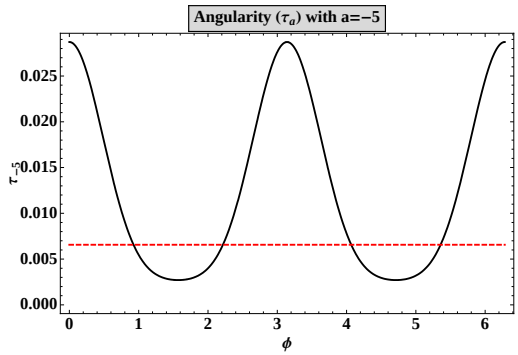

FIG. 4: Angularity, $\tilde{\tau}_{-5}$ as a function of the azimuthical angle of the $W(q\bar{q})$ pair, ϕ_q, for a typical top jet event, compared to the typical case two-body kinematics.

prediction of Fig. 2 is confirmed by the output of the event generator, with signal and data curves crossing in Fig. 2 near $\tau_{-2} = 0.02$, where $R(\tilde{\tau}_{-2}) \sim 1$.

Linear three-body decay. The leading-order differential top jet function can be obtained by considering its three-body decay kinematics. The analytic expression is similar to Eq. (6) for the two-body case, although a bit more elaborate. In the following we simply point out a few features that may help angularities to distinguish top jets from background, even when they have relatively linear flow.

The lowest-order three-body distribution is fully characterized by three angles. The first, θ_b, is the angle between the b quark and the jet axis. The second, θ_{Wq}, is the angle of the quark (from W-decay) relative to the W. The third, ϕ, is the angle of the same quark relative to the plane defined by the W and the b. For an on-shell W, the distributions peak around $\theta_b = \theta_m$ (as in two-body kinematics) and $\theta_{Wq} = \theta_{m(W)}$ the minimal angle relative to the W momenta in the W rest frame. Because it is massive, the W's decay products move in somewhat different directions, even in the boosted frame, and their relative orientation induces the ϕ-dependence. Clearly, planar flow has maxima for odd multiples of $\phi = \pi/2$, and vanishes at lowest order at multiples of π. To tag top events at zero planar flow, angularities can be of use. In Fig. 4 we plot $\tilde{\tau}_{-5}$ as a function of the azimuthical angle of the $W(q\bar{q})$ pair, ϕ, for a typical top jet event. We also show the corresponding value for the two-body case (clearly ϕ independent). For illustration we choose the kinematical configuration that maximizes the corresponding differential jet distributions. We notice that this top angularity has maxima with ϕ at zero and π at values far above the most likely two-body configuration. The reason is simply that angularities with large negative values of a tend to emphasize flow at the edge of the cone. Other values of a weight individual jets differently in general. We consider this simple plot, along with the forgoing examples from event generators, as strong evidence for the potential of jet shape analysis.

In summary, planar flow, angularities, and jet shapes that are as yet to be invented, may afford a variety of tools with which to distinguish the quantum mechanical histories of jets, whether resulting from heavy particle decay, or strong interactions.

Acknowledgements

The work of L.A., S.L., G.P., G.S. and I.S. was supported by the National Science Foundation, grants PHY-0354776, PHY-0354822, PHY-0653342 and PHY-06353354. The work of JV was supported by the Director, Office of Science, Office of High Energy Physics, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank I. Hinchliffe and M. Shapiro for comments on the manuscript.

[1] For recent studies see *e.g*: K. Agashe *et al.*, Phys. Rev. D **77**, 015003 (2008) [arXiv:hep-ph/0612015]; B. Lillie, L. Randall and L. T. Wang, JHEP **0709**, 074 (2007) [arXiv:hep-ph/0701166].

[2] See *e.g*: K. Agashe *et al.*, Phys. Rev. D **76**, 115015 (2007) [arXiv:0709.0007 [hep-ph]]; K. Agashe, H. Davoudiasl, G. Perez and A. Soni, Phys. Rev. D **76**, 036006 (2007) [arXiv:hep-ph/0701186]; J. Hirn, A. Martin and V. Sanz, JHEP **0805**, 084 (2008) [arXiv:0712.3783 [hep-ph]];

[3] D. Benchekroun, C. Driouichi, A. Hoummada, SN-ATLAS-2001-001, ATL-COM-PHYS-2000-020, EPJ Direct 3, 1 (2001); J. Conway, *et al.*, LPC Workshop on Early Physics at CMS, UC Davis (2007). G. Brooijmans *et al.*, arXiv:0802.3715 [hep-ph]; M. Vos, talk given in the *ATLAS Flavour Tagging Meeting*, CERN (2008).

[4] G. Brooijmans, ATLAS note, ATL-PHYS-CONF-2008-008;

[5] J. M. Butterworth, B. E. Cox and J. R. Forshaw, Phys. Rev. D **65**, 096014 (2002) [arXiv:hep-ph/0201098],

[6] L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, I. Sung, J. Virzi, in preparation.

[7] G. P. Salam and G. Soyez, JHEP **0705**, 086 (2007) [arXiv:0704.0292 [hep-ph]].

[8] G. Sterman, Phys. Rev. D **19**, 3135 (1979).

[9] N. Kidonakis, G. Oderda and G. Sterman, Nucl. Phys. B **525**, 299 (1998) [arXiv:hep-ph/9801268]; D. de Florian and W. Vogelsang, Phys. Rev. D **76**, 074031 (2007) [arXiv:0704.1677 [hep-ph]].

[10] M. Dasgupta and G. P. Salam, J. Phys. G **30**, R143 (2004) [arXiv:hep-ph/0312283]; C. F. Berger, Mod. Phys. Lett. A **20**, 1187 (2005) [arXiv:hep-ph/0505037]; M. Dasgupta, L. Magnea and G. P. Salam, JHEP **0802**, 055 (2008) [arXiv:0712.3014 [hep-ph]].

[11] D. E. Kaplan, K. Rehermann, M. D. Schwartz and B. Tweedie, arXiv:0806.0848 [hep-ph].

[12] J. Thaler and L. T. Wang, arXiv:0806.0023 [hep-ph].

[13] C. F. Berger, T. Kucs and G. Sterman, Phys. Rev. D **68**, 014012 (2003) [arXiv:hep-ph/0303051].

[14] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, arXiv:0802.2470 [hep-ph].

[15] G. Parisi, Phys. Lett. B **74**, 65 (1978).

[16] J. F. Donoghue, F. E. Low and S. Y. Pi, Phys. Rev. D **20**, 2759 (1979).

[17] R. K. Ellis, D. A. Ross and A. E. Terrano, Nucl. Phys. B **178**, 421 (1981).

[18] A. Banfi, Y. L. Dokshitzer, G. Marchesini and G. Zanderighi, JHEP **0105**, 040 (2001) [arXiv:hep-ph/0104162].

[19] F. Maltoni and T. Stelzer, JHEP **0302**, 027 (2003) [arXiv:hep-ph/0208156]. T. Stelzer and W. F. Long, Comput. Phys. Commun. **81**, 357 (1994) [arXiv:hep-ph/9401258]. JHEP **0709**, 028 (2007) [arXiv:0706.2334 [hep-ph]].

[20] T. Gleisberg et al., JHEP **0402** (2004) 056 [arXiv:hep-ph/0311263].

[21] C. F. Berger and L. Magnea, Phys. Rev. D **70**, 094010 (2004) [arXiv:hep-ph/0407024].