Factorization for substructures of boosted Higgs jets

Joshua Isaacson a,*, Hsiang-nan Li b, Zhao Li c, C.-P. Yuan a

a Dept. of Physics & Astronomy, Michigan State University, E. Lansing, MI 48824, USA
b Institute of Physics, Academia Sinica, Taipei 115, Taiwan, Republic of China
c Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

Article history:
Received 4 January 2016
Received in revised form 29 December 2016
Accepted 31 May 2017
Available online 14 June 2017

Keywords:
QCD
Jet substructure

A B S T R A C T
We present a perturbative QCD factorization formula for substructures of an energetic Higgs jet, taking the energy profile resulting from the $H \to b \bar{b}$ decay as an example. The formula is written as a convolution of a hard Higgs decay kernel with two $b$-quark jet functions and a soft function that links the colors of the two $b$ quarks. We derive an analytical expression to approximate the energy profile within a boosted Higgs jet, which significantly differs from those of ordinary QCD jets. This formalism also extends to boosted $W$ and $Z$ bosons in their hadronic decay modes, allowing an easy and efficient discrimination of fat jets produced from different processes.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/). Funded by SCOAP3.

The Higgs boson, which is responsible for the electroweak symmetry breaking mechanism in the Standard Model (SM), has been discovered at the Large Hadron Collider (LHC) with its mass around 125 GeV. Though its couplings to other particles seem to be consistent with the SM, the ultimate test as to whether this observed particle is the SM Higgs boson relies on the measurement of the trilinear Higgs coupling that appears in Higgs pair production. A Higgs boson is predominantly produced via gluon fusion processes at the LHC. It has been shown [1] that the cross section of the Higgs pair production increases rapidly with center-of-mass energy of hadron colliders. With much higher collision energy in the partonic process, preferred for exploring the trilinear Higgs coupling, the Higgs boson and its decay products will be boosted. An energetic Higgs boson can also be associated produced with other SM particles, such as $W$, $Z$ bosons, top quarks and jets [2].

The SM Higgs boson decays into a pair of bottom quark and antiquark dominantly. When the Higgs boson is highly boosted, this pair of bottom quarks may appear as a single jet and cannot be unambiguously discriminated from an ordinary QCD jet. A similar challenge applies to the identification of other boosted heavy particles, e.g., $W$ bosons, $Z$ bosons, and top quarks, when decaying via hadronic modes. Hence, additional information on internal structures of these boosted jets (such as their masses, energy profiles, and configurations of subjets) is required for the experimental identification. Many theoretical efforts were devoted to the exploration of heavy particle jet properties based on event generators [3–5]. As to the study of QCD jets in perturbative QCD (pQCD), the fixed-order calculation for their energy profiles was completed in [6], and the leading-logarithm resummation was performed in [7]. Recently, the pQCD formalism, including fixed-order evaluations [8] and the next-to-leading-logarithm resummation technique [9], was employed to investigate jet substructures. The alternative approach based on the soft-collinear effective theory and its application to jet production at an electron-positron collider were presented in Refs. [10,11]. For the application to jet profiles in proton-proton collisions, see, for example, Ref. [12].

In this Letter we develop a pQCD factorization formula to describe the internal jet energy profile (JEP) of the boosted jet resulting from the $H \to b \bar{b}$ decay, with energy $E_{J_H}$ and invariant mass $m_{J_H}$. The basic idea of our theoretical approach is as follows. A Higgs boson is a colorless particle, while its decay products, the bottom quark and antiquark, are colored objects and dressed by multiple gluon radiations to form a system with mass of $O(m_{J_H})$ and energy of $O(E_{J_H})$. The invariant mass $m_{J_H}$ of the bottom quark and its collimated gluons, with energy of $O(E_{J_H})$, typically satisfies the hierarchy $E_{J_H} \gg m_{J_H} \gg m_J$. Based on the factorization theorem, QCD dynamics characterized by different scales must factorize into soft, collinear, and hard pieces, separately. First, the Higgs jet function $f_{J_H}$ is factorized from a Higgs boson production process at the leading power of $m_{J_H}/E_{J_H}$. Then the $b$-quark jet function is defined at the leading power of $m_{J_H}/m_{J_{bb}}$ [9], soft gluons with energy of $O(m_{J_H})$ are absorbed into a soft function $S$, and the remaining energetic gluons with energy $O(E_{J_H})$ and invariant mass of $O(m_{J_H})$ go into the hard Higgs decay kernel $H$. 

* Corresponding author.
E-mail addresses: isaacs21@msu.edu (J. Isaacson), hsh@phys.sinica.edu.tw (H.-n. Li), zhaoli@ihep.ac.cn (Z. Li), yuan@pa.msu.edu (C.-P. Yuan).
The Higgs JEP is then factorized at leading power of $m_j/m_H$ into a convolution of the hard kernel with two $b$-quark jet functions and a soft function that links colors of the two $b$ quarks. We will demonstrate a simple scheme, in which the soft gluons are absorbed into one of the $b$-quark jets, forming a fat jet, and the soft function reduces to unity. The other $b$-quark jet is a thin jet to avoid double counting of soft radiation. Since a Higgs boson is massive and a color singlet, its JEP dramatically differs from that of ordinary QCD jets. Below, we present the derivation of the JEP of a Higgs boson decaying into a bottom-quark pair. We shall evaluate the hard decay kernel at the leading order (LO) and substitute the light-quark jet function, as obtained in Ref. [9], for the $b$-quark jet functions to predict the Higgs JEP up to the accuracy of next-to-leading logarithms.

The four-momentum of the Higgs jet can be written as $P_J = E_J h_J(1, \beta_{f_J}, 0, 0)$, with $\beta_{f_J} = \sqrt{1 - (m_{f_J}/E_J)^2}$. We define the Higgs jet function at the scale $\mu_J$ as

$$J_H(m_J^2, E_J, R, \mu_J^2) = \frac{2(2\pi)^3}{E_J} \sum_{N_J} \langle 0 | \phi(0) | N_J \rangle \langle N_J | \phi(0) | 0 \rangle \delta^2(\hat{n}_{J} - \hat{n}_{N}) \delta^2(m_{J}^2 - m_{N}^2)$$

where the coefficient has been chosen to satisfy $J_H^{(0)} = \delta(m_{J}^2 - m_{N}^2)$ at the zeroth order in the Yukawa coupling. $m_H$ represents the Higgs boson mass, and $R$ the Higgs jet cone radius. The three $\delta$-functions in the above definition specify the Higgs jet invariant mass, energy, and unit momentum direction of the set $N_J$ of final-state particles, respectively. After applying the aforementioned factorization procedure, $J_H$ is written as

$$J_H(m_J^2, E_J, R, \mu_J^2) = \frac{1}{E_J} \Pi_{i=1}^2 \int dm_{j_i}^2 dE_{j_i} d^2 \hat{n}_{j_i} \times \delta(m_{j_i}^2 - m_{j_i}^2) \times H(P_{j_1}, P_{j_2}, R, \mu_J^2) \delta^2(m_{j_1} - P_{j_1}, P_{j_2} - P_{j_1}, P_{j_2} - \omega)$$

where the factorization scale $\mu_J$ is introduced by the $b$-quark jet functions $J_i$, $m_{j_i}$, $E_{j_i}$, $P_{j_i}$, $R_i$ is the invariant mass (energy, momentum, radius) of the $b$-quark jets, and the soft function takes the form $S^{(0)} = \delta(\omega)$ at LO with the variable $\omega \equiv P_S \cdot P_{j_i}$, where $P_S$ is the soft gluon momentum.

To describe the Higgs JEP, we define the jet energy function $J^E_H(m_J^2, E_J, R, \mu_J^2)$ by including in Eq. (2) a step function $\Theta(r - \theta_J)$ for every final-state particle $j$. The final-state particles with non-vanishing step functions (i.e., emitted within the test cone of radius $r$) and associated with the $b$-quark jet $J$ are grouped into the $b$-quark jet energy function $J^E_H(m_J^2, E_J, R, r, \mu_J^2)$. The energetic final-state particles outside the $b$-quark jets and the test cone are absorbed into the hard kernel $H^E$. The other final-state particles outside the test cone are absorbed back into the jet energy functions. In this work, we will consider only the LO hard kernel, from which the hard gluon contribution to the JEP is absent, i.e., $J_1 \times J_2 \times H^E = 0$. We then arrive at

$$J^E_H(m_J^2, E_J, R, \mu_J^2) = \frac{1}{E_J} \Pi_{i=1}^2 \int dm_{j_i}^2 dE_{j_i} d^2 \hat{n}_{j_i} \int d\omega S(\omega, R, \mu_J^2) \times \sum_{i \neq j} J^E_i(m_{j_i}^2, E_{j_i}, R_i, \mu_J^2) J_H(m_{j_j}^2, E_{j_j}, R_j, \mu_J^2)$$

where the LO hard kernel is

$$H^{(0)}(P_{j_1}, P_{j_2}, R, \mu_J^2)$$

with the number of colors $N_c$, the $b$-quark mass $m_b$, the vacuum expectation value $v$, the Higgs decay width $\Gamma_H$ and the polar angle $\theta_J$.

Next, we integrate out the dependence on the Higgs jet invariant mass by taking the first moment in the Mellin transformation of $J^E_H$, defined as $J^E_H(1, E_J, R, \mu_J^2) \equiv \int J^E_H(m_J^2, E_J, R, \mu_J^2) dm_J^2 / (E_J R)^2$. To perform the integration over $\hat{n}_{J}$, we write the corresponding $\delta$-function as

$$\delta^2(\hat{n}_{j_1} + \hat{n}_{j_2}) = \delta \int \frac{dP_{j_1}}{|P_{j_1}|} \frac{dP_{j_2}}{|P_{j_2}|} \frac{d\hat{n}_{j_1}}{|P_{j_1}|} \frac{d\hat{n}_{j_2}}{|P_{j_2}|}$$

where the ratio is given by $|P_{j_1} + P_{j_2}|/|P_{j_1}| = (E_{j_1} \cos \theta_{j_1} + E_{j_2} \cos \theta_{j_2})/E_J$. The angular relation $E_{j_1} \sin \theta_{j_1} = E_{j_2} \sin \theta_{j_2}$ is then demanded. The integration over $E_{j_1}$ is trivial. The $b$-quark jet masses $m_{j_1}$, whose typical values are much smaller than $m_H$, are negligible in the hard kernel. The integrations over $m_{j_1}^2$ and $m_{j_2}^2$ can then be done trivially, with $\int dm_{j_1}^2 J^E(m_{j_1}^2, E_J, R, \mu_J^2) = 1 + O(\alpha_s)$.

The soft function is defined as a vacuum expectation value of two Wilson lines in the directions $\xi_{j_i}$ and $\xi_{j_i}$ in the $N$-space gives

$$S^{(1)} = \frac{\alpha_s C_F}{\pi (E_J R)^2} \int d(\xi_{j_i}^2 - \xi_{j_i}^4)^2 \left( 1 + \frac{1}{\ln \frac{4 \pi \mu_J^2 N^2}{R^4 E_J^2 \exp(\gamma_E)}} \right)$$

where the color factor $C_F = 4/3$ and the moment $\tilde{N} \equiv N M_\alpha / \gamma_E$, with $\gamma_E$ being the Euler constant. The off-shellness $\xi_{j_i}^2$, associated with the $b$-quark jets implies that $S^{(1)}$ contains the collinear dynamics which has been absorbed into the jet functions. Hence, the subtraction of the collinear divergences from the soft function is
necessary to avoid double counting. The collinear divergences from loop momenta collimated to the b quark (\(b\) quark) can be collected with the b quark (\(b\) quark) line being replaced by the eikonal line in the direction \(n_{j_1} (n_{j_2})\) that appear in the b-quark jet definitions [8]. The NLO subtraction term for the latter with the same cone radius \(R\) is obtained by substituting the vector \(n_{j_2}\) for \(\xi_{j_2}\) in \(S^{(1)}\). After this subtraction, we have

\[
S^{(1)} - S^{(1)}_{\xi_{j_2}} = \frac{\alpha_s C_F}{\pi (R E_{j_H})^2} \ln \frac{\xi_{j_1}^2}{(\xi_{j_1} \cdot n_{j_2})^2} \left( \frac{1}{\epsilon} + \ln \frac{4\pi \mu_0^2 N^2}{R^4 E_{j_H}^2 e^{-\epsilon}} \right),
\]

(7)

to which we can further impose the condition \(4(\xi_{j_1} \cdot n_{j_2})^2/n_{j_2}^2 = R^2\) for defining a quark (or gluon) jet [9].

Because the thin jet \(j_1\) contributes only the overall normalization in \(j_H\), the choice of \(n_{j_1}\) is arbitrary. We then utilize this freedom, and choose \(n_{j_2}\) such that \(S^{(1)}_{\xi_{j_2}}\) has the logarithmic coefficient the same as of \(S^{(1)} - S^{(1)}_{\xi_{j_2}}\). This choice is possible, because \(\xi_{j_1} \cdot \xi_{j_2} \sim (m_H/E_{j_H})^2 \sim O(r)\) in the considered kinematic region. The further collinear subtraction leads to

\[
S^{(1)} - S^{(1)}_{\xi_{j_2}} - S^{(1)}_{\xi_{j_1}} \approx 0,
\]

(8)

so the soft function in this special scheme is given by \(S(\alpha, R, \mu_0^2) \approx \delta(\alpha)\). We stress that the above argument is made only on the NLO level, and that it is likely to get residual soft gluon effects at higher loops, especially as the thin jet has a vanishing radius. For a detailed analysis of the soft gluon correction to jet momenta and jet masses at hadron colliders, refer to [14].

Equation (3) then reduces to

\[
\tilde{j}_H^E (1, E_{j_H}, R, r) = \frac{1}{R^2 (E_{j_H})^3} \int \left( \frac{m_b}{V} \right)^2 \int dE_{j_1} \int d \cos \theta_{j_1} \times (E_{j_1}, \cos \theta_{j_1} + E_{j_2} \cos \theta_{j_2}) \times [E_{j_1}^2 \Theta(a - \theta_{j_1}) + R^2 E_{j_1}^2 \tilde{j}_H^E (1, E_{j_2}, R, r) \Theta(a - \theta_{j_2})] \\
\times \frac{\sqrt{2 E_{j_1} E_{j_2} [1 - \cos(\theta_{j_1} + \theta_{j_2})]} - m_H^2 + \Gamma_H m_H^2}{[2 E_{j_2} (1 - \cos(\theta_{j_1} + \theta_{j_2})) - m_H^2]^2 + \Gamma_H^2 m_H^2},
\]

(9)

where the light-jet quark functions are set at the factorization scale \(\mu_0^2 = (E_{j_H})^2/N\), and the renormalization scale for \(\tilde{j}_H^E\) is chosen to be \(\mu = E_{j_H}/R\) [9]. The Mellin transformation \(\tilde{j}_H^E (1, E_{j_H}, R, r) = \int \tilde{j}_H^E (m_{j_H}^2, E_{j_H}, r) d m_{j_H}^2/(R E_{j_H})^2\), which includes the effect from the resummation of the large logarithms in \(r\) at small \(r\) to all orders [9], has been inserted.

The choice of the merging parameter \(a\) is a matter of factorization schemes, and the difference arising from distinct \(a's\) will be compensated by the corresponding distinct hard kernels \(H^E\).

That is, a larger \(a\) means more contribution to the Higgs JEP from the b-quark jets, and less contribution from \(H^E\). Since we consider only the LO hard decay kernel here, our analysis will be more consistent as a larger \(a\) is chosen. Below, we set \(a = 1.7\), which corresponds to the configuration with the overlap area of the test cone and the thin jet cone being half of the test cone area, and predict the Higgs JEP with \(E_{j_H} = 500\) GeV and \(E_{j_H} = 1000\) GeV, both with a cone radius of \(R = 0.7\). A JEP is defined as

\[
\Psi(E_{j_H}, R, r) = \frac{\tilde{j}_H^E (1, E_{j_H}, R, r)}{\tilde{j}_H^E (1, E_{j_H}, R, R)}.
\]

(10)

It is interesting to note that a simple expression can be derived for the JEP of a boosted Higgs jet after applying the narrow width approximation for the Higgs boson propagator. It yields

\[
\Psi(E_{j_H}, R, r) = \frac{\bar{E}_{j_H} (1, E_{j_H}, R, r)}{\bar{E}_{j_H} (1, E_{j_H}, R, R)},
\]

(11)

where the integration variable \(z = E_{j_H}/E_{j_H}/\Psi_{1}\), the lower limit \(z_m(r) = m^2_{j_H}/m^2_{j_H} + a^2 r^2\), the small parameter \(m_{j_H} \approx m_{j_H}/E_{j_H}\), and \(\Psi_{1}\) denotes the light-quark JEP [9,13].

Our formalism can be readily extended for studying boosted unpolarized W and Z bosons in their hadronic decay modes, by inserting their masses and widths, since Eq. (11) is coupling- and spin-independent. As shown in Fig. 1, the predicted W, Z and Higgs JEPs, for the heavy-boson jet energy \(E_{j_H} = 500\) GeV, agree well with those given by Pythia8 [15]. In the same figure, we also show the theoretical uncertainties of our prediction, which arise from the theoretical uncertainties of the input light-quark JEP \(\Psi_{1}\), in which the resummation scales are varied by a factor of two around their typical values [9,13]. For the Pythia8 comparison, we have used the 4C tune which was shown to agree well with the ATLAS data for the JEPs of QCD jets with energies ranging from 30 GeV to 600 GeV [16]. In the current analysis we include the effects of initial-state and final-state radiations, hadronization, and beam remnants, but turn off the effects from multiple parton interactions, which are supposed to be subleading. Similar agreement is also observed, cf. Fig. 2, for these boosted electroweak bosons at 1 TeV, though the deviation at \(r = 0.1\) is larger. The QCD JEPs [9] for the jet energies 500 GeV and 1 TeV are also exhibited in Figs. 1 and 2, respectively, for comparison. (Here, we consider the inclusive jet production at the 7 TeV LHC as an example, such that the 500 GeV and 1 TeV QCD jets are composed of about 75% and 90% quark jet, respectively.) Note that the invariant mass of 1-TeV QCD jets, being around 100 GeV, is roughly in the same mass range as of Higgs jet. It is clear to see that the Higgs JEPs are lower at small \(r\) due to the finite angular separation between the two b-quark jets set by the large Higgs mass, and increases faster with \(r\) once the energetic b-quark jets start to contribute. This obvious difference allows a discrimination of Higgs jets from QCD jets, in a statistical sense. An illuminating example for discriminating Higgs boson production mechanisms (gluon fusion versus vector-boson fusion) based on the QCD JEPs has been investigated in [17]. From Figs. 1 and 2, we can get a rough estimate of the effective radii needed to capture all of the radiation for these boosted electroweak bosons. We find that the effective radius needed is \(R \approx 0.6\) and \(R \approx 0.4\) for the Higgs boson at 500 GeV and 1 TeV, respectively. For either W or Z boson, they are about 0.5 and 0.3, respectively. Moreover, the W and Z JEPs are thinner than the Higgs JEP due to their smaller masses. This is further demonstrated via the energy dependence of the JEP for a fixed \(r \) value (\(r = 0.2\)) in Fig. 3: the separation of the W, Z, and Higgs jets becomes more difficult as the jet energy in-
increases, because the ratio of the mass to the jet energy becomes smaller. Furthermore, the energy dependence of the JETs of bosons is quite different from that of the QCD jet. This is because the former is governed by the 1−2 decay kinematics, while the latter is by the collinear splitting of quark and gluon partons.

Note that the simple expression in Eq. (11) is derived under the aforementioned approximations, at the LO accuracy for the hard kernels, and with the neglect of soft links among the heavy-boson jet and other subprocesses, such as beams and other final-state particles. The associated theoretical uncertainties can be reduced by taking into account relevant corrections, and will be addressed in a future work. Furthermore, the questions of how techniques like trimming [18], pruning [19,20], mass drop tagger [21,22], soft drop [23], and other similar techniques affect the JET, how subleading corrections affect the JET, and what happens to the jet energy profile if we loose some of the aforementioned approximations will be investigated in a future work. A challenging subject is the study of effects from the non-global logarithms [24−26], although the good agreement between the theoretical results in [9,13] with data hints that such effects are not significant here.

In conclusion, we have applied the pQCD factorization to formulate the JET of a boosted colorless heavy-particle (such as W, Z, Higgs, W' and Z' boson) jet, which is found to differ dramatically from the ordinary QCD JETs with similar energy and jet radius. The formalism is greatly simplified by considering inside the boosted jet a thin jet and a fat jet, which absorbs all the soft-gluon effect, when the heavy-particle decays into a quark and antiquark pair. More interestingly, the analytical expression for the JETs of the electroweak bosons allows for an easy and efficient discrimination of different production processes for these boosted jets. The implementation of this discrimination method can further help suppress QCD background to signals of boosted W and Z jets, after applying conventional kinematic selections.

Acknowledgements

This work was supported by Ministry of Science and Technology of the Republic of China under Grant No. NSC-101-2112-M-001-006-MY3; by the Natural Science Foundation of China under Grant No. 11305179; by Youth Innovation Promotion Association, CAS; and by the U.S. National Science Foundation under Grant No. PHY-1417326.

References