Recent progress on jet substructure theory

Zhong-Bo Kang[1,2,3,★]

[1]*Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA*
[2]*Mani L. Bhaumik Institute for Theoretical Physics, University of California, Los Angeles, CA 90095, USA*
[3]*Center for Frontiers in Nuclear Science, Stony Brook University, Stony Brook, NY 11794, USA*

Abstract. In these proceedings, we review recent theoretical progress on jet substructure at the colliders. Focusing on two observables – jet mass and groomed jet radius, we perform theoretical computations for jets measured in the single inclusive jet production in proton-proton collisions, $pp \rightarrow \text{jet} + X$. We consider both standard ungroomed jets as well as soft-drop groomed jets. Within the Soft Collinear Effective Theory (SCET), we establish QCD factorization theorems which allow for the joint resummation of several classes of logarithmic corrections to all orders in the strong coupling constant. We present numerical results and compare with the available data from the LHC.

1 Introduction

At present day collider experiments such as the Large Hadron Collider (LHC) and the Relativistic Heavy Ion Collider (RHIC), highly energetic jets play an important role as precision probes of the Standard Model and beyond [1]. In the past years, the capability of jet probes at a future Electron Ion Collider (EIC) and their usage of probing different aspects of nuclear dynamics has been realized [2–4]. In addition, jet substructure techniques have become important tools in high energy particle and nuclear physics, where people have realized that the internal structure of the identified jets contains additional valuable information. For example, jet substructure techniques are used to improve our understanding of the QCD hadronization mechanism, to discriminate between quark and gluon jets, or to identify jets originating from the decay of boosted objects. Jet substructure observables are also increasingly being studied in heavy-ion collisions where they provide important tests and insights of the hot and dense QCD medium [5]. More recently it has also been realized that jet substructure can probe both collinear and transverse momentum dependent parton-to-hadron fragmentation functions in ways that traditional inclusive observables do not offer [6–10].

One of the important techniques that have been developed for better studying jet substructure is jet grooming, which is designed to remove soft wide-angle radiation from the identified jets. We focus on the so-called soft-drop grooming algorithm developed in [11]. Such a grooming algorithm removes the soft contamination of jets and thus leads to reduced sensitivity to nonperturbative effects. Because of this, it allows for a more direct comparison of perturbative QCD calculations and experimental data. Both on the experimental and the theoretical side, significant progress has been made recently in improving our understanding of soft drop groomed jet observables. In the heavy-ion community, soft

---

★e-mail: zkang@physics.ucla.edu
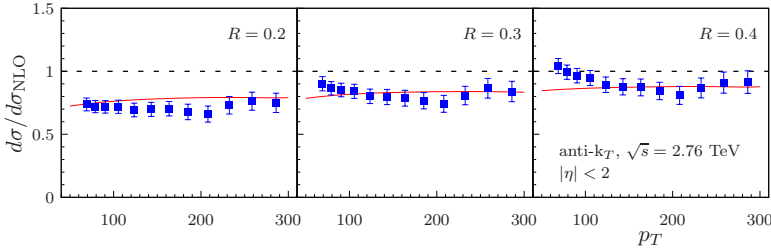
Figure 1. The $\ln R$-resummed jet cross section is compared with the CMS experimental data [18].

drop groomed jet substructure observables have also received an increasing attention from both experiment and theory. Jet grooming techniques can be used to isolate different aspects of jet quenching and may help to discriminate between different model assumptions.

In these proceedings, we mainly review our recent studies on groomed and ungroomed jet mass distribution, and the groomed jet radius.

2 Jet cross section and jet mass

We consider inclusive jet production $pp \rightarrow$ jet + X, where jets are identified with the anti-k_T algorithm [12] with a given radius R. A factorized form has been shown in [13] within the framework of Soft Collinear Effective Theory (SCET) [14–17], where the double differential cross section for inclusive jets with a given transverse momentum p_T and rapidity η is given by

$$\frac{d\sigma^{pp\rightarrow\mathrm{jet}X}}{dp_T d\eta} = \sum_{a,b,c} f_a \otimes f_b \otimes H_{ab}^c \otimes J_c(z, p_T R, \mu) \,. \tag{1}$$

Here, we suppressed the arguments of the various functions for better readability. See [13] for more details. Let us just mention in the above equation that $J_c(z, p_T R, \mu)$ is called semi-inclusive jet functions (siJFs), which describe the formation of a jet with the transverse momentum p_T and jet parameter R originating from a parent parton c at scale μ. They satisfy the same timelike DGLAP equations that govern the scale evolution of fragmentation functions. By solving the DGLAP equations and thus evolving the siJFs from the natural scale $\mu \sim p_T R$ to the hard scale $\mu \sim p_T$, the resummation of logarithms $\ln R$ can be achieved. As shown in Fig. 1, such a $\ln R$-resummed jet cross section leads to better agreement with the CMS experimental data [18]. A further improvement can be achieved with a joint threshold and $\ln R$ resummation [19].

One might generalize the above framework to study jet substructure. One of the most prominent and most often used jet substructure observables is the jet mass distribution, with $m_J^2 = (\sum_{i\in J} p_i)^2$ and p_i the four-momenta of all the particles inside the reconstructed jet. The factorized cross section can be written as

$$\frac{d\sigma}{d\eta dp_T d\tau} = \sum_{abc} f_a \otimes f_b \otimes H_{ab}^c \otimes \mathcal{G}_c(z, p_T R, \tau, \mu) \,, \tag{2}$$

where the dimensionless variable $\tau = m_J^2/p_T^2$, and the functions $\mathcal{G}_c(z, p_T R, \tau, \mu)$ are the semi-inclusive jet mass functions (siJMFs), which describe how a parton c initiates the signal jet that carries a momentum fraction z of that parton, and at the same time the jet mass τ is observed. In the kinematic
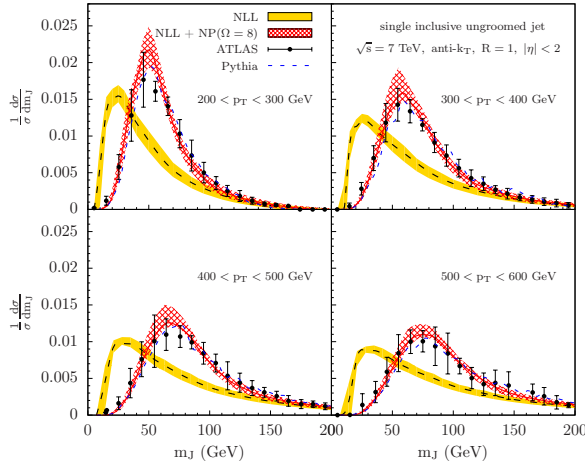
Figure 2. Comparison with the experimental data from the ATLAS collaboration [20] taken at $\sqrt{s}$ = 7 TeV.

region where $\tau \ll R^2$, i.e. $m_J^2 \ll (p_T R)^2$, we obtain the following factorization for the siJMFs

$$\mathcal{G}_c(z, p_T R, \tau, \mu) = \sum_i \mathcal{H}_{c \to i}(z, p_T R, \mu) C_i\left(\tau^{C_i}, p_T, \mu\right) \otimes S_i\left(\tau^{S_i}, p_T R, \mu\right), \tag{3}$$

where $\mathcal{H}_{c \to i}$ are the hard matching functions, and are related to unconstrained radiation outside the jet, C_i are the collinear functions that take into account collinear radiation inside the observed jet, and S_i are the soft functions. In Fig. 2, we show comparison of our theoretical calculations and the experimental data from the ATLAS collaboration [20] taken at $\sqrt{s}$ = 7 TeV. The dashed black lines and the yellow error bands show the purely perturbative results at next-to-leading logarithmic (NLL) accuracy. The red lines and the red hatched error bands show the theoretical results when the non-perturbative shape function is included. For details about the functional form of such a shape function, see Ref. [21]. The characteristic momentum scale of Ω = 8 GeV signals the fact that, as expected, the ungroomed jet mass distribution is very sensitive to non-perturbative physics such as hadronization and the underlying event etc [22, 23].

3 Soft-drop groomed jet mass

The soft drop grooming procedure recursively removes soft wide-angle radiation from an identified jet [11]. The algorithm starts by re-clustering the constituents of an anti-k_T jet [12] with Cambridge-Aachen algorithm [24, 25]. The obtained angular ordered clustering tree is then declustered recursively where at each step the soft drop condition is checked

$$\frac{\min[p_{T1}, p_{T2}]}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R}\right)^\beta. \tag{4}$$

Here $p_{T1,2}$ are the transverse momenta of the two branches obtained at each declustering step and ΔR_{12} is their geometric distance. Soft branches that fail the criterion are removed from the jet. The algorithm terminates when the criterion is met and the particles in the remaining two branches consti-tute the groomed jet. The soft drop algorithm depends on two parameters: a soft threshold z_{cut} and an
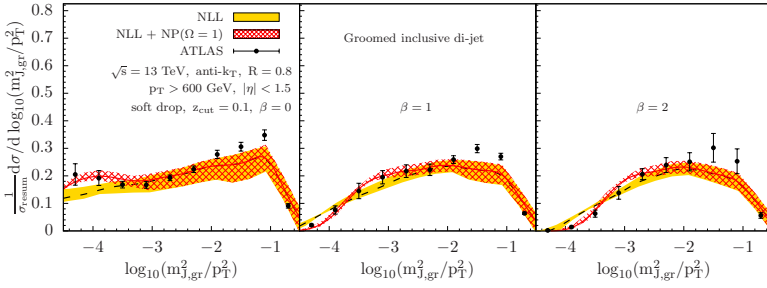
Figure 3. Comparison with the ATLAS data [26] for the soft drop groomed jet mass distribution at $\sqrt{s} = 13$ TeV.

angular exponent β. While z_{cut} sets the energy scale below which soft branches are removed from the jet, β determines the sensitivity of the grooming algorithm to the wide-angle soft radiation. The mass of the resulting groomed jet is usually referred to as the soft drop groomed jet mass denoted as $m_{J,\mathrm{gr}}$.

One can derive a similar factorized form for the groomed jet mass distribution, where one replaces $\mathcal{G}_c(z, p_T R, \tau, \mu)$ in Eq. (2) by the groomed siJMFs $\mathcal{G}_c^{\mathrm{gr}}(z, p_T, R, \tau_{\mathrm{gr}}, \mu; z_{\mathrm{cut}}, \beta)$, where $\tau_{\mathrm{gr}} = m_{J,\mathrm{gr}}^2 / p_T^2$. In practice, the LHC experiments usually choose $z_{\mathrm{cut}} \sim 0.1$ while τ_{gr} can be as low as $\mathcal{O}\left(10^{-4}\right)$, and a typical value for the jet radius parameter is $R = 0.8$. Therefore, we are particularly interested in the factorization of the cross section in the region where $\tau_{\mathrm{gr}}/R^2 \ll z_{\mathrm{cut}} \ll 1$. In this region, we have the following refactorized expression for the groomed siJMFs

$$\mathcal{G}_c^{\mathrm{gr}}(z, p_T, R, \tau_{\mathrm{gr}}, \mu; z_{\mathrm{cut}}, \beta) = \sum_i \mathcal{H}_{c \to i}(z, p_T R, \mu) S_i^{\not{\mathrm{gr}}}(p_T, R, \mu; z_{\mathrm{cut}}, \beta) C_i \otimes S_i^{\mathrm{gr}}(\tau_{\mathrm{gr}}, p_T, R, \mu; z_{\mathrm{cut}}, \beta).$$

Here the hard matching functions $\mathcal{H}_{c \to i}$ and the collinear functions C_i are the same as for the ungroomed case in Eq. (3). However, the soft functions are different where $S_i^{\not{\mathrm{gr}}}$ characterizes the soft wide-angle radiation (failing the soft drop criterion), while S_i^{gr} is associated with soft particles that pass the soft drop criterion and thus remain inside the groomed jet.

The soft drop grooming procedure is designed to eliminate the sensitivity to the underlying event contribution. This is confirmed by our numerical results for the jet mass distributions. In Fig. 3, we compare our theoretical calculations with the experimental data from the ATLAS collaboration [26] at $\sqrt{s} = 13$ TeV at the LHC. The dashed line and the corresponding yellow error band are the purely perturbative results at NLL accuracy, while the red line and the red hatched band are NLL results convolved with the non-perturbative shape function. By choosing a much smaller non-perturbative parameter $\Omega = 1$ GeV in the shape function, we find a very good description of the experimental data.

4 Groomed jet radius

Two variables that characterize important features of the soft drop groomed jet are the momentum sharing fraction z_g and the groomed jet radius R_g. Their values are obtained from the kinematics of the two remaining branches when the soft drop algorithm terminates: $z_g = \frac{\min[p_{T1}, p_{T2}]}{p_{T1} + p_{T2}}$, and $R_g = \Delta R_{12} = R \theta_g$. Often the variable θ_g is used which corresponds to the geometric distance of the two remaining branches normalized by the radius R of the ungroomed initial jet. The radius of a soft drop groomed jet is one of the key observables characterizing the impact of grooming on a jet and is related to the active area of the groomed jet. In [27], we establish a factorization formalism that allows for the resummation of all relevant large logarithms, which is based on demonstrating the all
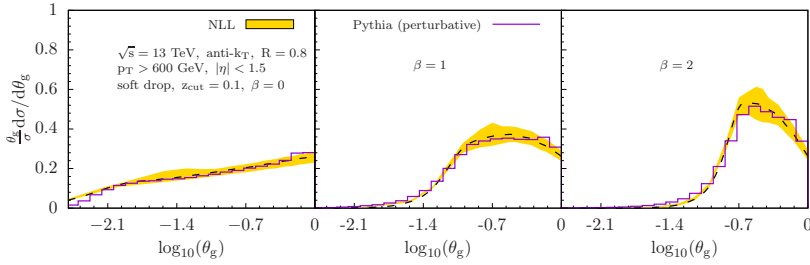
Figure 4. The soft drop groomed jet radius θ_g (dashed black, yellow band) in proton-proton collisions at $\sqrt{s} =$ 13 TeV, compared to Pythia 8 simulations [28] (purple) at the parton level, for $\beta = 0, 1, 2$, respectively.

order equivalence to a jet veto in the region between the boundaries of the groomed and ungroomed jet. Non-global logarithms including clustering effects due to the Cambridge/Aachen algorithm are resummed to all orders using a suitable Monte Carlo algorithm.

We start from the cross section $\Sigma(\theta_g)$ differential in the transverse momentum p_T and rapidity η of the observed jet, but cumulative in the groomed jet radius where any value below θ_g contributes. The distribution differential in θ_g can then be obtained as

$$\frac{d\sigma}{d\eta \, dp_T \, d\theta_g} = \frac{d}{d\theta_g} \frac{d\Sigma(\theta_g)}{d\eta \, dp_T} \, . \tag{5}$$

The factorized formalism for the cumulative cross section $d\Sigma(\theta_g)$ is similar to Eq. (2), but now with a new semi-inclusive jet function $\mathcal{G}_c(z, \theta_g, p_T R, \mu; z_{\text{cut}}, \beta)$. In the kinematic region where $z_{\text{cut}} \ll 1$ and $\theta_g \ll 1$, we develop a refactorization for $\mathcal{G}_c$. For details, see Ref. [27]. In Fig. 4, we present the numerical results (dashed black, yellow band) for the θ_g differential distribution $\theta_g/\sigma_{\text{incl}} \, d\sigma/d\theta_g$ which is obtained by differentiating the cumulative cross section $\Sigma(\theta_g)$, see eq. (5). The three panels show the result for different values of $\beta = 0, 1, 2$ (left to right). For comparison, we also show Pythia 8.230 results using the default tune [28] (purple). Here we do not include the nonperturbative hadronization and the contribution from underlying event (UE). In general, we find very good agreement between the Pythia 8 simulation and our perturbative results concerning both the shape and the overall magnitude. More recently ATLAS released their measurements [29] on groomed jet radius, and our predictions agree well with the experimental data.

5 Conclusions

In these proceedings, we review recent theoretical progress on jet substructure at the colliders. We first reported the impact of the jet radius $\ln R$ resummation on the single inclusive jet cross section, and we then studied groomed and ungroomed jet mass, and the soft-drop groomed jet radius. Within the soft-collinear effective theory, we establish QCD factorization theorems for the observables that are measured for single inclusive jet production in proton-proton collisions, $pp \rightarrow$ jet $+ X$. We find that the $\ln R$ resummation leads to better agreement with the experimental measurements on the jet cross section. For the groomed and ungroomed jet substructure observables, in general, we found that our theoretical calculations lead to a very good description of the experimental data in the regions where the factorization theorems hold. In addition, we found that the ungroomed/standard jet mass distribution is very sensitive to the non-perturbative contributions such as underlying events and hadronization. However, with soft-drop grooming applied, such strong sensitivity is much reduced for the groomed jet mass distribution.

6 Acknowledgements

References

[1] A.J. Larkoski, I. Moult, B. Nachman, Phys. Rept. **841**, 1 (2020), `1709.04464`

[2] E.C. Aschenauer, S. Fazio, J.H. Lee, H. Mantysaari, B.S. Page, B. Schenke, T. Ullrich, R. Venugopalan, P. Zurita, Rept. Prog. Phys. **82**, 024301 (2019), `1708.01527`

[3] E.C. Aschenauer, K. Lee, B.S. Page, F. Ringer (2019), `1910.11460`

[4] M. Arratia, Y. Song, F. Ringer, B. Jacak (2019), `1912.05931`

[5] X.N. Wang, ed., *Quark-Gluon Plasma 5* (World Scientific, New Jersey, 2016), ISBN 9789814663700

[6] Z.B. Kang, F. Ringer, I. Vitev, JHEP **11**, 155 (2016), `1606.07063`

[7] D.P. Anderle, T. Kaufmann, M. Stratmann, F. Ringer, I. Vitev, Phys. Rev. **D96**, 034028 (2017), `1706.09857`

[8] Z.B. Kang, J.W. Qiu, F. Ringer, H. Xing, H. Zhang, Phys. Rev. Lett. **119**, 032001 (2017), `1702.03287`

[9] R. Bain, L. Dai, A. Leibovich, Y. Makris, T. Mehen, Phys. Rev. Lett. **119**, 032002 (2017), `1702.05525`

[10] Z.B. Kang, X. Liu, F. Ringer, H. Xing, JHEP **11**, 068 (2017), `1705.08443`

[11] A.J. Larkoski, S. Marzani, G. Soyez, J. Thaler, JHEP **05**, 146 (2014), `1402.2657`

[12] M. Cacciari, G.P. Salam, G. Soyez, JHEP **04**, 063 (2008), `0802.1189`

[13] Z.B. Kang, F. Ringer, I. Vitev, JHEP **10**, 125 (2016), `1606.06732`

[14] C.W. Bauer, S. Fleming, D. Pirjol, I.W. Stewart, Phys. Rev. **D63**, 114020 (2001), `hep-ph/0011336`

[15] C.W. Bauer, I.W. Stewart, Phys. Lett. **B516**, 134 (2001), `hep-ph/0107001`

[16] C.W. Bauer, D. Pirjol, I.W. Stewart, Phys. Rev. **D65**, 054022 (2002), `hep-ph/0109045`

[17] M. Beneke, A.P. Chapovsky, M. Diehl, T. Feldmann, Nucl. Phys. **B643**, 431 (2002), `hep-ph/0206152`

[18] V. Khachatryan et al. (CMS), Phys. Rev. **C96**, 015202 (2017), `1609.05383`

[19] X. Liu, S.O. Moch, F. Ringer, Phys. Rev. Lett. **119**, 212001 (2017), `1708.04641`

[20] G. Aad et al. (ATLAS), JHEP **05**, 128 (2012), `1203.4606`

[21] Z.B. Kang, K. Lee, X. Liu, F. Ringer, JHEP **10**, 137 (2018), `1803.03645`

[22] M. Dasgupta, K. Khelifa-Kerfa, S. Marzani, M. Spannowsky, JHEP **1210**, 126 (2012), `1207.1640`

[23] I.W. Stewart, F.J. Tackmann, W.J. Waalewijn, Phys. Rev. Lett. **114**, 092001 (2015), `1405.6722`

[24] Y.L. Dokshitzer, G.D. Leder, S. Moretti, B.R. Webber, JHEP **08**, 001 (1997), `hep-ph/9707323`

[25] M. Wobisch, T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999* (1998), pp. 270–279, `hep-ph/9907280`

[26] M. Aaboud et al. (ATLAS), Phys. Rev. Lett. **121**, 092001 (2018), `1711.08341`

[27] Z.B. Kang, K. Lee, X. Liu, D. Neill, F. Ringer, JHEP **02**, 054 (2020), `1908.01783`

[28] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, Comput. Phys. Commun. **191**, 159 (2015), `1410.3012`

[29] G. Aad et al. (ATLAS) (2019), `1912.09837`