Letter

Field-theoretic analysis of hadronization using soft drop jet mass

Anna Ferdinand,[1,2,*] Kyle Lee,[3,4,†] and Aditya Pathak[1,5,‡]

[1]*University of Manchester, School of Physics and Astronomy, Manchester M13 9PL, United Kingdom*
[2]*DAMTP, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom*
[3]*Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[4]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[5]*Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany*

(Received 16 February 2023; revised 30 May 2023; accepted 15 November 2023; published 11 December 2023)

One of the greatest challenges in quantum chromodynamics is understanding the hadronization mechanism, which is also crucial for carrying out precision physics with jet substructure. In this paper, we bring together recent advancements in our understanding of nonperturbative structure of the soft drop jet mass based on field theory, with precise perturbative calculations at next-to-next-to-leading logarithmic accuracy of its multidifferential variants. This allows for a model-independent analysis of power corrections associated with hadronization in a systematic manner. We test and calibrate hadronization models and their interplay with parton showers by comparing our universality predictions with various event generators for quark and gluon initiated jets in both lepton-lepton and hadron-hadron collisions. Our findings reveal that hadronization models perform better for quark jets relative to gluon jets. Our results provide a valuable toolbox for precision studies with the soft drop jet mass and pave the way for future analyses using real-world collider data. The stringent constraints derived in our framework are useful for improving the modeling of hadronization and its interplay with parton showers in next-generation event generators.

DOI: 10.1103/PhysRevD.108.L111501

Introduction. The study of jets and their substructure in high-energy particle colliders has become a highly active research area over the past decade [1,2]. One significant development in this field is the application of jet grooming techniques [3–10] which allow for theoretical control by eliminating the wide-angle soft radiation, including contamination from the underlying event and pileup, as well as by reducing hadronization effects. In particular, the soft drop (SD) grooming [6–8] has received widespread attention, inspiring many theoretical calculations both for jets in vacuum [11–39] and in medium [40–44], as well as several experimental analyses [45–54]. Of the various groomed observables, the SD jet mass has been extensively studied in both theoretical [6,55–61] and experimental communities [62–72] and has been utilized in numerous phenomenological applications, such as quantifying medium modification [40,62,69] and conducting precise measurements of the top-quark mass [73–75] and strong coupling constant [76,77].

To realize the full potential of the SD jet mass for precision measurements, it is crucial to account for the impact of hadronization power corrections, which are of comparable magnitude to the uncertainty of state-of-the-art perturbative calculations at next-to-next-to-leading logarithmic (NNLL) accuracy (see Ref. [77] for a recent overview). In Ref. [78], a field-theory-based formalism [79–85] was developed for describing nonperturbative (NP) effects in the groomed jet mass. Using the soft collinear effective theory (SCET) [86–89] including boosted nonperturbative modes, this work gained an analytical handle on the power corrections, expressing them in terms of universal $\mathcal{O}(\Lambda_{\mathrm{QCD}})$ nonperturbative constants and perturbatively calculable Wilson coefficients. Recent studies [90,91] have computed the Wilson coefficients to NNLL accuracy as an essential step toward this goal. Moreover, it was revealed that these corrections exhibit a strongly constrained functional dependence on jet-flavor, kinematics, and grooming parameters, rendering them a powerful tool for testing hadronization models [92–94] in event generators. This subtle nature of nonperturbative power corrections in the groomed jet mass intertwined with perturbative dynamics thus presents us with a unique

*annaferd@mit.edu
†kylel@mit.edu
‡aditya.pathak@desy.de

opportunity to deepen our understanding of the complex physics of hadronization.

Building on this formalism, in this paper, we take a crucial step forward in our quest to describe the intricate effects of hadronization within jets. We present the first-ever proof-of-principle analysis that investigates the universality of hadronization effects in jet substructure under LHC kinematics, employing multiple parton shower generators, hadronization models, and state-of-the-art perturbative calculations. Our technique offers a unique approach to assess the compatibility between hadronization models and parton showers. This is particularly important when striving to maintain consistency with perturbative calculations. Furthermore, our analysis also demonstrates the effectiveness of this approach for conducting precision studies, such as determination of the strong coupling constant [77] and top mass measurement [73,95]. These findings underscore the significance of utilizing real-world collider data for future follow-up studies.

Hadronization corrections to groomed jet mass. Soft drop [7] proceeds by declustering a Cambridge-Achen clustered tree that defines a given jet. At each stage, the groomer tests the pair of subjets with transverse momenta $p_{T,i}$ and $p_{T,j}$ for the following condition:

$$\frac{\min(p_{T_i}, p_{T_j})}{p_{T_i} + p_{T_j}} > z_{\text{cut}} \left(\frac{\Delta R_{ij}}{R_0}\right)^{\beta}. \tag{1}$$

If the above condition is not satisfied, then the softer subjet of the pair (with $p_{T,i} < p_{T,j}$) is groomed away, and the groomer then proceeds by declustering and applying the above test on the harder subjet. The SD condition depends on the energy cut z_{cut} and the angular modulation parameter β that determine the strength of the groomer.

The SD jet mass, as compared to the ungroomed jet mass, has a much larger perturbative regime known as the soft drop operator expansion (SDOE) region, which is depicted between the vertical lines in Fig. 1 and defined below. This region allows for studying hadronization effects in a systematic expansion. In Ref. [78], it was demonstrated using SCET [86–89] that the leading hadronization corrections in the SDOE region are characterized by three $\mathcal{O}(\Lambda_{\text{QCD}})$ NP universal constants $\{\Omega_{1\kappa}^{\infty}, \Upsilon_{1,0\kappa}^{\circledcirc}, \Upsilon_{1,1\kappa}^{\circledcirc}\}$, which solely depend on the parton $\kappa = q, g$ that initiates the jet. These constants are completely independent of the jet's kinematic properties, such as the jet p_T (or E_J), rapidity η_J, radius R, and the grooming parameters [7], such that

$$\frac{1}{\sigma_\kappa}\frac{\mathrm{d}\sigma_\kappa}{\mathrm{d}m_J^2} = \frac{1}{\hat{\sigma}_\kappa}\frac{\mathrm{d}\hat{\sigma}_\kappa}{\mathrm{d}m_J^2} - Q\Omega_{1\kappa}^{\infty}\frac{\mathrm{d}}{\mathrm{d}m_J^2}\frac{1}{\hat{\sigma}_\kappa}\frac{\mathrm{d}\hat{\sigma}_\kappa^{\infty}}{\mathrm{d}m_J^2} + \frac{\Upsilon_{1,0\kappa}^{\circledcirc} + \beta\Upsilon_{1,1\kappa}^{\circledcirc}}{Q}\frac{1}{\hat{\sigma}_\kappa}\frac{\mathrm{d}\hat{\sigma}_\kappa^{\circledcirc}}{\mathrm{d}m_J^2} + \cdots, \tag{2}$$

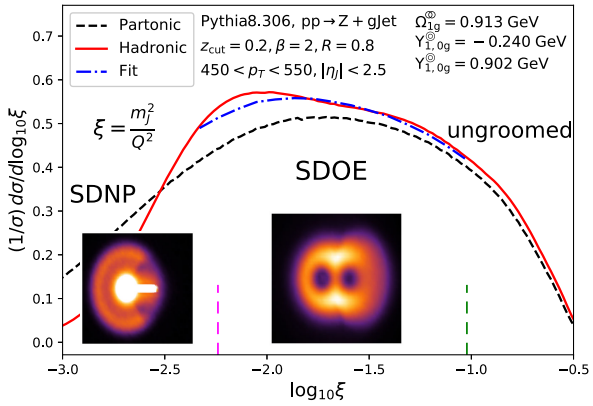

FIG. 1. An example of fit for nonperturbative parameters in PYTHIA8.3 simulation of groomed jet mass. The insets show distribution of low-energy particles as heat maps around the soft drop stopping subjets in the transverse plane.

Here, $\mathrm{d}\sigma_\kappa$ and $\mathrm{d}\hat{\sigma}_\kappa$, respectively, refer to hadron and parton level groomed jet mass cross sections for flavor κ and Q characterizing the hard scale of the jet. The weights $\mathrm{d}\hat{\sigma}_\kappa^{\infty,\circledcirc}$ are perturbatively calculable.

In the SDOE region, the dominant hadronization corrections arise from a two-pronged dipole structure, which consists of an energetic collinear subjet at the center of the jet and a collinear-soft (c-soft) subjet responsible for stopping the grooming algorithm. The power corrections beyond the leading terms in Eq. (2) arise from effects suppressed by higher powers of Λ_{QCD} and from configurations that distort the two-pronged catchment area. The latter correction is a next-to-leading-logarithmic effect, and hence Eq. (2) can be regarded as a factorization of nonperturbative effects at leading-logarithmic accuracy, where the strong ordering of angles ensures the two-pronged geometry. As the jet mass decreases, we enter the soft drop nonperturbative region, where the c-soft mode becomes nonperturbative and the associated nonperturbative effects become $\mathcal{O}(1)$. The transition between these two regions is clearly visible in Fig. 1, where the insets show the distribution of low-energy nonperturbative particles in the transverse plane of the jet [78]. It is worth noting that, unlike the analytical hadronization models used in previous studies [6,61,76,84], Eq. (2) is a model-independent statement and takes into account hadron mass effects.

The NP factorization formula given by Eq. (2) offers a unique opportunity to investigate hadronization effects in jets in a comprehensive way. The formalism imposes stringent constraints on the three NP constants, which must provide a good description of collider data spanning a wide range of energies. The remarkable structure of Eq. (2), with its $\mathcal{O}(\Lambda_{\text{QCD}})$ constants, a β-proportional coefficient $\Upsilon_{1,1\kappa}^{\circledcirc}$, and z_{cut} independence, among other features, makes this task far from trivial, but it also makes it a valuable tool for testing hadronization models. In this study, we show how the universality of the NP parameters strongly

constrains their values, enabling us to determine them with high precision using various combinations of soft drop and kinematic parameters. This has important implications, for instance, for improving the prospects of the strong coupling constant α_s determination at the LHC.

Calculation of perturbative weights. To characterize the two-pronged configuration of the collinear and the c-soft subjet in the SDOE region, additional measurements of the groomed jet radius R_g and soft subjet energy fraction z_g [7,35,36,96] are needed. After marginalization, the resulting expressions are given by [90]

$$\frac{1}{\hat{\sigma}_\kappa}\frac{\mathrm{d}\hat{\sigma}_\kappa^{\infty}}{\mathrm{d}m_J^2} \equiv \int \mathrm{d}r_g r_g \frac{1}{\hat{\sigma}_\kappa}\frac{\mathrm{d}^2\hat{\sigma}_\kappa}{\mathrm{d}m_J^2 \mathrm{d}r_g},$$
$$\frac{1}{\hat{\sigma}_\kappa}\frac{\mathrm{d}\hat{\sigma}_\kappa^{\circledcirc}}{\mathrm{d}m_J^2} \equiv \int \frac{\mathrm{d}r_g \mathrm{d}z_g \delta(z_g - z_{\mathrm{cut}} r_g^\beta)}{r_g} \frac{1}{\hat{\sigma}_\kappa}\frac{\mathrm{d}^3\hat{\sigma}_\kappa}{\mathrm{d}m_J^2 \mathrm{d}r_g \mathrm{d}z_g}. \quad (3)$$

As described above, the NP constants in Eq. (2) are independent of the jet kinematics and grooming parameters. Therefore, all these dependencies are included in $\mathrm{d}\hat{\sigma}_\kappa^{\infty,\circledcirc}$. In Eq. (3), $r_g = R_g/R$ appears analogously to how jet radius R appears in hadronization corrections in the tail of ungroomed jet mass and in the jet p_T spectrum [84,85],

$$m_{J,\mathrm{no\ sd}}^2 = \hat{m}_{J,\mathrm{no\ sd}}^2 + p_T R \Omega_{1\kappa}, \qquad p_T = \hat{p}_T + \frac{1}{R}\Upsilon_{1\kappa}, \quad (4)$$

where $\Omega_{1\kappa}, \Upsilon_{1\kappa} \sim \Lambda_{\mathrm{QCD}}$ are NP parameters and hatted variables are parton level values. In the case of the SD jet mass, the dynamically determined groomed jet radius R_g plays the role of R. The term in Eq. (2) with $\mathrm{d}\hat{\sigma}_\kappa^{\infty}$ is analogous to the ungroomed jet mass shift correction in the tail but is now described by a *different constant* $\Omega_{1\kappa}^{\infty}$ as $m_J^2 = \hat{m}_J^2 + p_T R_g \Omega_{1\kappa}^{\infty}$.

The term in the second line in Eq. (2) with $\mathrm{d}\hat{\sigma}_\kappa^{\circledcirc}$ is called the boundary correction. This effect is similar to the migration of events across p_T bins due to hadronization. Near the "boundary" of the c-soft subjet passing/failing soft drop, i.e., when $z_g \approx z_{\mathrm{cut}} r_g^\beta$, the partonic values $\hat{z}_g$ and $\hat{r}_g$ are modified due to hadronization as

$$z_g = \hat{z}_g + \frac{1}{r_g}\frac{\Upsilon_{1,0\kappa}^{\circledcirc}}{p_T R}, \qquad r_g = \hat{r}_g - \frac{\Upsilon_{1,1\kappa}^{\circledcirc}}{p_T R}. \quad (5)$$

Here, $\Upsilon_{1,0\kappa}^{\circledcirc}$ characterizes the shift in the p_T of the c-soft subjet analogous to jet p_T shift in Eq. (4), and $\Upsilon_{1,1\kappa}^{\circledcirc}$ describes the change in the subjet location relative to the collinear subjet. The combination of the two gives rise to the linear structure $\Upsilon_{1\kappa}^{\circledcirc} = \Upsilon_{1,0\kappa}^{\circledcirc} + \beta\Upsilon_{1,1\kappa}^{\circledcirc}$ as shown in Eq. (2) and constitutes a nontrivial prediction. Finally, it is useful to factor out the parton-level groomed jet mass cross section from $\mathrm{d}\sigma^{\infty,\circledcirc}$:

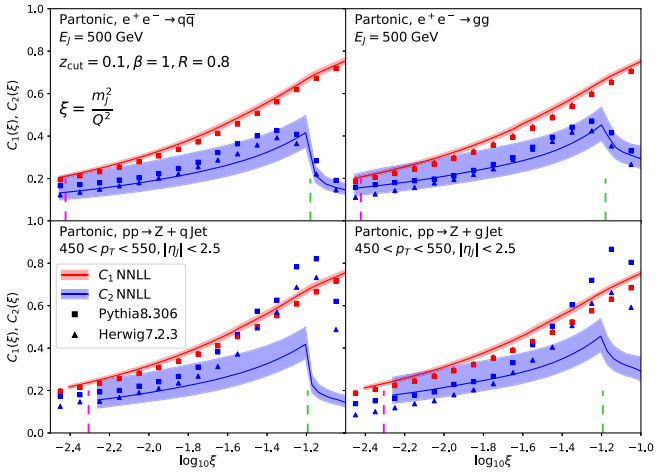

FIG. 2. Weighted cross sections for hadronization corrections normalized to parton level jet mass spectrum as defined in Eq. (6) for $z_{\mathrm{cut}} = 0.1$ and $\beta = 1$.

$$\frac{\mathrm{d}\hat{\sigma}_\kappa}{\mathrm{d}m_J^2} C_1^\kappa(m_J^2) \equiv \frac{\mathrm{d}\hat{\sigma}_\kappa^{\infty}}{\mathrm{d}m_J^2}, \qquad \frac{\mathrm{d}\hat{\sigma}_\kappa}{\mathrm{d}m_J^2} C_2^\kappa(m_J^2) \equiv \frac{\mathrm{d}\hat{\sigma}_\kappa^{\circledcirc}}{\mathrm{d}m_J^2}. \quad (6)$$

This definition is advantageous as it enables us to combine an analytical calculation of the coefficients $C_{1,2}^\kappa(m_J^2)$ with the jet mass cross section $\mathrm{d}\hat{\sigma}_\kappa$ obtained from parton showering, which will be discussed below. In our study, we employ a recently improved calculation of $C_{1,2}^\kappa(m_J^2)$ at NNLL accuracy with correct treatment of the soft drop cusp at this order, as detailed in Ref. [91].

Figure 2 compares the NNLL computation of $C_{1,2}^\kappa$ to partonic PYTHIA and HERWIG. The VINCIA results at parton level are almost identical to PYTHIA. It is important to emphasize that the parton shower outcomes are leading logarithmic (LL) accurate; hence, we do not expect them to agree with NNLL predictions. Nevertheless, we observe a substantial agreement between NNLL $C_{1,2}^\kappa$ and Monte Carlo (MC) simulations within perturbative uncertainty for all four processes. Consequently, the NNLL-based NP parameter extraction should align with the intrinsic parameters from the hadronization model. This validates our use of the more precise NNLL calculations to calibrate hadronization parameters in our proof-of-principle study. As we will see later, the utilization of NNLL computation allows us to convert the perturbative uncertainty on these coefficients into the precision of NP parameters. We hope for a direct application of our approach to collider data for extracting NP parameters.

Next, we note that the small errors for C_1^κ in Fig. 2 are due to the cancellation of correlated uncertainties in the two factors in Eq. (6). For pp, the agreement for the boundary term is poor for jet masses close to the cusp, which is caused by the initial-state radiation (ISR) contribution. However, the NP corrections in the cusp region are relatively suppressed, and ISR NP corrections are also

expected to be smaller, as they involve subleading r_g^2 moments of the boundary cross section [91]. Therefore, these effects do not significantly impact the analysis presented below.

Calibrating hadronization models. With state-of-the-art NNLL perturbative results for $C_{1,2}^{\kappa}(m_J^2)$ available, we are able to precisely calibrate hadronization models. In addition, by incorporating NNLL perturbative uncertainty, we can significantly improve upon the proof-of-principle analysis of Ref. [90], which lacked uncertainty estimates due to the use of LL predictions. We simulate $e^+e^- \to gg$, $e^+e^- \to q\bar{q}$, $pp \to Z + q$ jet and $pp \to Z + g$ jet processes using parton showers from PYTHIA8.3 [92], VINCIA2.3 [97], and HERWIG7.2 [94] with their default hadronization models. We then reconstruct anti-k_T [98] jets with $R = 0.8$ using FASTJET [99] and analyze them using jet analysis software JETlib written by some of us [100]. We analyze dijet events in e^+e^- collisions with center-of-mass energies of $Q = 500$, 750, and 1000 GeV. In pp collisions, we use leading jets with p_T in the ranges [400, 600], [600, 800], and [800, 1000] GeV and consider soft drop parameters $z_{\rm cut} \in \{0.05, 0.1, 0.15, 0.2\}$ and $\beta \in \{0, 0.5, 1, 1.5, 2\}$. As the NP parameters are predicted to be independent of jet kinematics and grooming parameters, our analysis is performed over a wide range of these parameters.

We begin by explicitly defining the SDOE region where our analysis is carried out. We first define a dimensionless variable $\xi \equiv m_J^2/Q^2$, where

$$Q^{(pp)} \equiv p_T R, \qquad Q^{(ee)} \equiv 2E_J. \tag{7}$$

In terms of ξ, the SDOE region is then defined as $\xi \in [\xi_{\rm SDOE}, \xi_0']$, where

$$\xi_{\rm SDOE} \equiv \xi_0 \left(\frac{\rho \Lambda_{\rm QCD}}{Q \xi_0}\right)^{\frac{2+\beta}{1+\beta}}, \qquad \xi_0' \equiv \frac{\xi_0}{(1+\zeta^2)^{\frac{2+\beta}{2}}}. \tag{8}$$

Here, ξ_0 is the location of the soft drop cusp [77,91],

$$\xi_0^{(pp)} = z_{\rm cut}\left(\frac{R}{R_0}\right)^{\beta}, \qquad \xi_0^{(ee)} = z_{\rm cut}\left(\sqrt{2}\frac{\tan\frac{R}{2}}{\sin\frac{R_0}{2}}\right)^{\beta}, \tag{9}$$

while ζ is defined by

$$\zeta^{(pp)} \equiv \frac{R}{2\cosh\eta_J}, \qquad \zeta^{(ee)} \equiv \tan\frac{R}{2}, \tag{10}$$

such that ξ_0' in Eq. (8) is the soft-wide angle transition point of the NNLL calculation. We set $\Lambda_{\rm QCD} \to 1$ GeV, the typical scale of transition from parton showers to hadronization. The parameter ρ in Eq. (8) determines the onset of the SDOE region, and we set $\rho = 4.5$. In principle, any choice satisfying $\rho \gg 1$ is acceptable. We explore other choices of ρ in the supplemental material [101].

Finally, we perform a least-squares fit for the NP parameters by defining our χ^2 statistic as

$$\chi^2 \equiv \sum_i \frac{[(\vec{\sigma}_{\kappa,\rm had}^{\rm MC})_i - (\vec{\sigma}_{\kappa,\rm part+NP}(\Omega_{1\kappa}^{\infty}, \ldots))_i]^2}{(\Delta\vec{\sigma})_i^2}. \tag{11}$$

Here, we define $\vec{\sigma}_X$ as a vector consisting of cross section values for $n_{\rm bins} = 10$ bins within the fit range and all permutations of p_T (or E_J), $z_{\rm cut}$, and β values considered above. We use $\vec{\sigma}_{\kappa,\rm had}^{\rm MC}$ to denote the hadron-level MC groomed jet mass cross section and define $\vec{\sigma}_{\kappa,\rm part+NP}$ by incorporating the NP constants $\Omega_{1\kappa}^{\infty}$, $\Upsilon_{1,0\kappa}^{\circledcirc}$, $\Upsilon_{1,1\kappa}^{\circledcirc}$, and NNLL computation of $C_{1,2}^{\kappa}$ in Eq. (6) to the parton-level MC spectrum $\mathrm{d}\hat{\sigma}_{\kappa}^{\rm MC}$, following Eq. (2). The uncertainty in the denominator is defined as

$$(\Delta\vec{\sigma})_i^2 \equiv (0.05(\vec{\sigma}_{\rm part\times C_1})_i)^2 + (0.25(\vec{\sigma}_{\rm part\times C_2})_i)^2, \tag{12}$$

where, guided by the size of perturbative uncertainties in Fig. 2, we have assigned 5% and 25% uncertainty, respectively, to the weighted cross sections for shift and boundary corrections. The NP constants $\Omega_{1\kappa}^{\infty}$, $\Upsilon_{1,0\kappa}^{\circledcirc}$, and $\Upsilon_{1,1\kappa}^{\circledcirc}$ are then varied to minimize this χ^2 statistic. An example of the fit for mass distribution is shown in Fig. 1.

The fit results for the NP constants with scale variations of $C_{1,2}^{\kappa}$ for PYTHIA are presented in Table I. As expected, the parameters are found to be $\lesssim 1$ GeV. Quark jets within the two quark processes show similar parameter values within uncertainties. Even when NP parameters for quark jets are simultaneously fit for in e^+e^- and pp processes, an excellent χ^2 value of 0.840 per degree of freedom (dof) is obtained. This is not surprising, as soft drop isolates the jet from surrounding radiation. To investigate this further, correlations between $\Omega_{1\kappa}^{\infty}$ and $\Upsilon_{1,0\kappa}^{\circledcirc}$ for the four processes are shown in Fig. 3, where each ellipse represents a 1σ deviation. Perturbative uncertainties are accounted for by repeating the fit with varying $C_{1,2}^{\kappa}$ up and down within the uncertainty band shown in Fig. 2. Excellent agreement within uncertainties is observed between the NP parameters for quark jets in pp and e^+e^- collisions in PYTHIA simulations, and a moderate agreement is observed for VINCIA and HERWIG. In contrast, while HERWIG exhibits similar levels of agreement for gluon jets and quark jets at both colliders, PYTHIA and VINCIA show significant disagreement. This indicates that, contrary to the expectation for groomed jets, hadronization modeling of gluon jets in isolation in e^+e^- collisions in PYTHIA and VINCIA differs significantly from jets in hadron colliders. This deviation goes beyond perturbative uncertainty. The differing results between PYTHIA and VINCIA suggest an interplay of parton showers with hadronization models. We anticipate that this discovery will encourage further investigation into improving the modeling of gluon jets in event generators. In the supplemental material [101], we show correlations for other combinations and fit results for HERWIG and VINCIA.

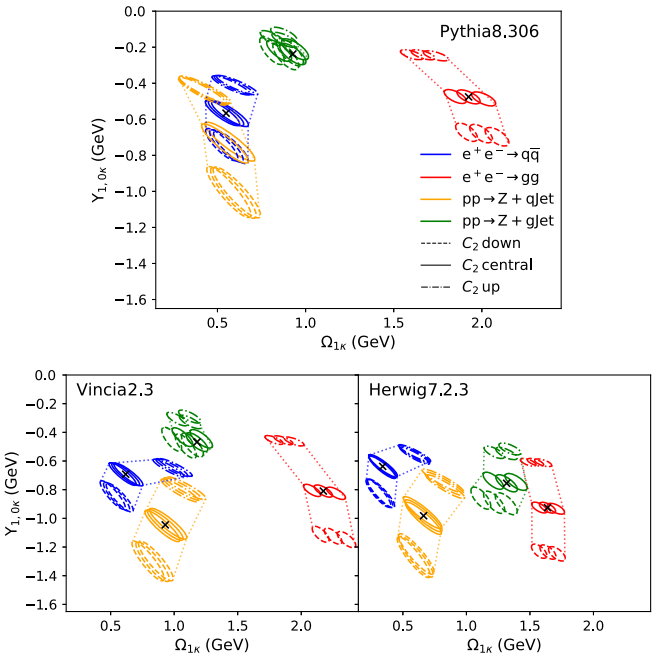

FIG. 3. Testing jet flavor universality of soft drop NP parameters in PYTHIA8.3 (top), VINCIA2.3 (bottom, left), and HERWIG7.2 (bottom, right).

We proceed to test the grooming parameters independence of the NP constants. To this end, we adopt the same methodology as in Ref. [78] and test this behavior by comparing the fit results for individual $z_{\rm cut}$ and β values with the global fit. In Fig. 4, we demonstrate the linear β dependence of the boundary correction by fitting for a single parameter $\Upsilon^{\circledcirc}_{1\kappa}(\beta)$ for each value of β. We fix $\Omega^{\infty}_{1\kappa}$ to its global-fit value in this case due to degeneracy in the NP parameters. The error bars account for perturbative uncertainty in $C^{\kappa}_{1,2}$ by refitting with minimum and maximum variations. We find that all three simulations perform well in each of the four cases. We also conducted a similar analysis to test the $z_{\rm cut}$ independence of NP parameters and found that the three event generators pass the test for both quark and gluon jets in e^+e^- collisions but exhibit a linear trend in $z_{\rm cut}$ for both flavors in pp collisions. More details can be found in the supplemental material [101]. However, the larger χ^2 values for gluon jets, as seen in Table I for PYTHIA (also true for HERWIG and VINCIA), suggest that the modeling of hadronization in gluon jets is less consistent with our field theory predictions. Finally, our analysis of the $e^+e^- \to q\bar{q}$ process using NNLL predictions of $C^{\kappa}_{1,2}$ demonstrates significant improvement in the universality behavior with respect to the grooming parameters, compared to Ref. [78] where LL predictions were used.[1] In conclusion, our universality tests of the NP parameters generally display expected behaviors in all the cases considered but also reveal some tension with hadronization models, pointing to interesting avenues for further improvement.[2] Our findings motivate the use of real-world collider data for further analyses and underscore the need for continued development of hadronization models and their interface with parton showers in next-generation event-generators.

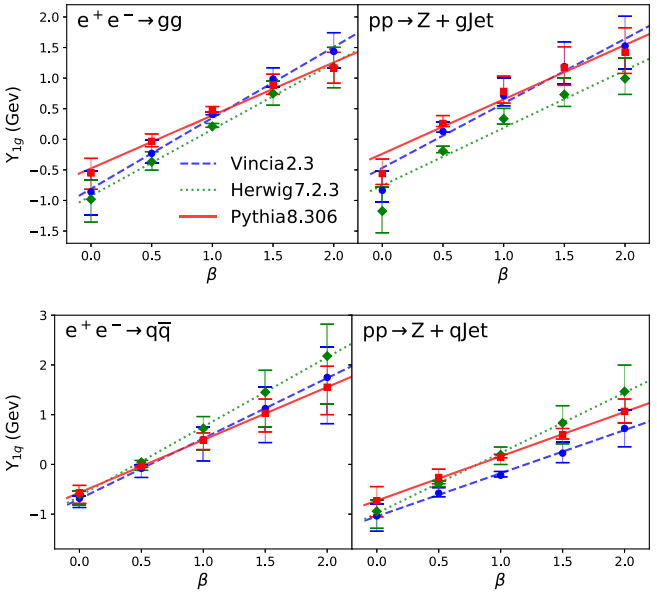

FIG. 4. Test for linear β dependence of boundary corrections ($\Upsilon^{\circledcirc}_{1\kappa} = \Upsilon^{\circledcirc}_{1,0\kappa} + \beta\Upsilon^{\circledcirc}_{1,1\kappa}$) in gluon (top) and quark (bottom) jets for e^+e^- (left) and pp (right) collisions.

TABLE I. Fit results for NP constants in PYTHIA8.3 for quark and gluon jets in e^+e^- and pp collisions.

Quark jets	Ω^{∞}_{1q} (GeV)	$\Upsilon^{\circledcirc}_{1,0q}$ (GeV)	$\Upsilon^{\circledcirc}_{1,1q}$ (GeV)	$\chi^2_{\rm min}$/dof
$e^+e^- \to q\bar{q}$	$0.55^{+0.06}_{-0.03}$	$-0.57^{+0.16}_{-0.19}$	$1.06^{+0.31}_{-0.35}$	$0.77^{+0.03}_{-0.00}$
$pp \to Z + q$	$0.56^{+0.05}_{-0.14}$	$-0.73^{+0.29}_{-0.28}$	$0.89^{+0.27}_{-0.25}$	$0.65^{+0.01}_{-0.02}$
Gluon jets	Ω^{∞}_{1g} (GeV)	$\Upsilon^{\circledcirc}_{1,0g}$ (GeV)	$\Upsilon^{\circledcirc}_{1,1g}$ (GeV)	$\chi^2_{\rm min}$/dof
$e^+e^- \to gg$	$1.92^{+0.16}_{-0.32}$	$-0.48^{+0.23}_{-0.22}$	$0.87^{+0.25}_{-0.25}$	$3.13^{+0.05}_{-0.20}$
$pp \to Z + g$	$0.93^{+0.01}_{-0.12}$	$-0.24^{+0.11}_{-0.01}$	$0.89^{+0.20}_{-0.23}$	$1.34^{+0.05}_{-0.10}$

Conclusion. In this paper, we have presented a comprehensive framework for analyzing nonperturbative corrections in soft drop jet mass by combining earlier work on nonperturbative factorization with high precision calculations of multidifferential soft drop cross sections. Our analysis of hadronization models shows that the

[1]Note that our numerical results for $e^+e^- \to q\bar{q}$ also differ from those in Ref. [78] due to different prescription for error in Eq. (12) and newer versions of MC.

[2]For example, the analysis at LL in Ref. [78] already revealed problems in the HERWIG8.2 hadronization model, which resulted in its improvement in version 8.3.

nonperturbative parameters display the universal behaviors predicted by field theory, thereby demonstrating the reliability of our framework. Our work has immediate implications for precision phenomenology involving soft drop jet mass. For example, in Ref. [77], our results are used to evaluate the impact of the NP corrections on the sensitivity and ultimate precision achievable on α_s at the LHC using SD jet mass. The results show that when the nonperturbative parameters in Eq. (2) are left unconstrained hadronization effects in the $\beta = 1$ case are 3% (8%) for quark (gluon) jets, which are of the same size as the NNLL perturbative uncertainty. Our analysis suggests that with high-precision calculations for the soft drop jet mass and the boundary correction (C_2^κ in Fig. 2) it will be feasible to significantly constrain one or more of the NP constants and hence improve the ultimate precision achievable on α_s determination at the LHC. In summary, our work has provided a crucial understanding of hadronization corrections that are necessary for precision measurements with soft drop jet mass. Our framework serves as a benchmark tool for improving hadronization modeling in MC event generators and motivates analyses with real-world collider data.

Finally, while our analysis has focused on groomed observables, we believe that our approach presented here for analyzing hadronization corrections has broader applicability. It can be effectively extended to other jet substructure observables, such as energy-energy correlators [102–104] and event shapes in a multijet region [105]. The SDOE region offers a rich interplay between perturbative and nonperturbative dynamics, primarily due to the presence of perturbative emissions. Thus, our work not only provides a comprehensive understanding of hadronization corrections for groomed observables but also opens up new possibilities for studying other jet substructure observables where the leading-order correction exhibits similar characteristics.

Acknowledgments. We would like to thank Mrinal Dasgupta and Michael Seymour for helpful discussions. We are grateful to Simon Plätzer for many discussions and support with analysis with HERWIG. We thank Holmfridur Hannesdottir, Johannes Michel, and Iain Stewart for numerous discussions and feedback on the manuscript. We provide a numerical implementation of the NNLL calculation in C++ building on core classes of SCETlib [106], which will be made available as a part of the SCETlib::SD module [107]. We thank Johannes Michel for support with above-mentioned implementation in SCETlib. K. L. was supported by the LDRD program of LBNL and the U.S. DOE under Contract No. DE-SC0011090. A. P. was a member of the Lancaster-Manchester-Sheffield Consortium for Fundamental Physics, which is supported by the UK Science and Technology Facilities Council (STFC) under Grant No. ST/T001038/1. A. F. also gratefully acknowledges support from the above-mentioned grant.

[1] A. J. Larkoski, I. Moult, and B. Nachman, Phys. Rep. **841**, 1 (2020).
[2] R. Kogler *et al.*, Rev. Mod. Phys. **91**, 045003 (2019).
[3] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Phys. Rev. D **81**, 094023 (2010).
[4] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. C **75**, 59 (2015).
[5] D. Krohn, J. Thaler, and L.-T. Wang, J. High Energy Phys. 02 (2010) 084.
[6] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, J. High Energy Phys. 09 (2013) 029.
[7] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, J. High Energy Phys. 05 (2014) 146.
[8] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008).
[9] C. Frye, A. J. Larkoski, J. Thaler, and K. Zhou, J. High Energy Phys. 09 (2017) 083.
[10] F. A. Dreyer, L. Necib, G. Soyez, and J. Thaler, J. High Energy Phys. 06 (2018) 093.
[11] A. J. Larkoski, I. Moult, and D. Neill, J. High Energy Phys. 02 (2018) 144.
[12] A. J. Larkoski, I. Moult, and D. Neill, arXiv:1708.06760.
[13] J. Baron, S. Marzani, and V. Theeuwes, J. High Energy Phys. 08 (2018) 105.
[14] Z.-B. Kang, K. Lee, X. Liu, and F. Ringer, Phys. Lett. B **793**, 41 (2019).
[15] Y. Makris and V. Vaidya, J. High Energy Phys. 10 (2018) 019.
[16] A. Kardos, G. Somogyi, and Z. Trócsányi, Phys. Lett. B **786**, 313 (2018).
[17] D. Napoletano and G. Soyez, J. High Energy Phys. 12 (2018) 031.
[18] C. Lee, P. Shrivastava, and V. Vaidya, J. High Energy Phys. 09 (2019) 045.
[19] D. Gutierrez-Reyes, Y. Makris, V. Vaidya, I. Scimemi, and L. Zoppi, J. High Energy Phys. 08 (2019) 161.
[20] A. Kardos, A. Larkoski, and Z. Trócsányi, Acta Phys. Pol. B **50**, 1891 (2019).
[21] Y. Mehtar-Tani, A. Soto-Ontoso, and K. Tywoniuk, Phys. Rev. D **101**, 034004 (2020).
[22] A. Kardos, A. J. Larkoski, and Z. Trócsányi, Phys. Rev. D **101**, 114034 (2020).
[23] A. Kardos, A. J. Larkoski, and Z. Trócsányi, Phys. Lett. B **809**, 135704 (2020).

[24] A. J. Larkoski, J. High Energy Phys. 09 (2020) 072.
[25] A. Lifson, G. P. Salam, and G. Soyez, J. High Energy Phys. 10 (2020) 170.
[26] P. Caucal, A. Soto-Ontoso, and A. Takacs, Phys. Rev. D **105**, 114046 (2022).
[27] Y.-T. Chien and I. Vitev, Phys. Rev. Lett. **119**, 112301 (2017).
[28] G. Milhano, U. A. Wiedemann, and K. C. Zapp, Phys. Lett. B **779**, 409 (2018).
[29] N.-B. Chang, S. Cao, and G.-Y. Qin, Phys. Lett. B **781**, 423 (2018).
[30] H. T. Li and I. Vitev, Phys. Lett. B **793**, 259 (2019).
[31] Y.-T. Chien and R. Kunnawalkam Elayavalli, arXiv:1803.03589.
[32] C. Sirimanna, S. Cao, and A. Majumder, *Proc. Sci.*, HardProbes2018 (2019) 053 [arXiv:1901.03635].
[33] P. Caucal, E. Iancu, and G. Soyez, J. High Energy Phys. 10 (2019) 273.
[34] F. Ringer, B.-W. Xiao, and F. Yuan, Phys. Lett. B **808**, 135634 (2020).
[35] P. Cal, K. Lee, F. Ringer, and W. J. Waalewijn, Phys. Lett. B **833**, 137390 (2022).
[36] A. Larkoski, S. Marzani, J. Thaler, A. Tripathee, and W. Xue, Phys. Rev. Lett. **119**, 132003 (2017).
[37] Y.-T. Chien and I. W. Stewart, J. High Energy Phys. 06 (2020) 064.
[38] P. Cal, K. Lee, F. Ringer, and W. J. Waalewijn, J. High Energy Phys. 11 (2020) 012.
[39] I. W. Stewart and X. Yao, J. High Energy Phys. 09 (2022) 120.
[40] R. Kunnawalkam Elayavalli and K. C. Zapp, J. High Energy Phys. 07 (2017) 141.
[41] H. A. Andrews *et al.*, J. Phys. G **47**, 065102 (2020).
[42] Y. Mehtar-Tani and K. Tywoniuk, J. High Energy Phys. 04 (2017) 125.
[43] J. Casalderrey-Solana, G. Milhano, D. Pablos, and K. Rajagopal, J. High Energy Phys. 01 (2020) 044.
[44] J. Brewer, Q. Brodsky, and K. Rajagopal, J. High Energy Phys. 02 (2022) 175.
[45] M. Aaboud *et al.* (ATLAS Collaboration), Phys. Rev. Lett. **121**, 092001 (2018).
[46] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **120**, 142302 (2018).
[47] K. Kauder (STAR Collaboration), Nucl. Part. Phys. Proc. **289–290**, 137 (2017).
[48] A. M. Sirunyan *et al.* (CMS Collaboration), J. High Energy Phys. 11 (2018) 113.
[49] S. Acharya *et al.* (ALICE Collaboration), Phys. Lett. B **802**, 135227 (2020).
[50] ATLAS Collaboration, Report No. ATLAS-CONF-2019-035, 2019.
[51] G. Aad *et al.* (ATLAS Collaboration), arXiv:1912.09837.
[52] J. Adam *et al.* (STAR Collaboration), Phys. Lett. B **811**, 135846 (2020).
[53] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. Lett. **124**, 222002 (2020).
[54] A. M. Sirunyan *et al.* (CMS Collaboration), J. High Energy Phys. 10 (2018) 161.
[55] M. Dasgupta, A. Fregoso, S. Marzani, and A. Powling, Eur. Phys. J. C **73**, 2623 (2013).
[56] A. J. Larkoski, J. Thaler, and W. J. Waalewijn, J. High Energy Phys. 11 (2014) 129.
[57] M. Dasgupta, A. Powling, and A. Siodmok, J. High Energy Phys. 08 (2015) 079.
[58] S. Marzani, L. Schunk, and G. Soyez, J. High Energy Phys. 07 (2017) 132.
[59] J. Chay and C. Kim, J. Korean Phys. Soc. **74**, 439 (2019).
[60] Z.-B. Kang, K. Lee, X. Liu, and F. Ringer, J. High Energy Phys. 10 (2018) 137.
[61] S. Marzani, L. Schunk, and G. Soyez, Eur. Phys. J. C **78**, 96 (2018).
[62] S. Acharya *et al.* (ALICE Collaboration), Phys. Lett. B **776**, 249 (2018).
[63] M. Aaboud *et al.* (ATLAS Collaboration), Phys. Rev. Lett. **121**, 092001 (2018).
[64] ATLAS Collaboration, Report No. ATLAS-CONF-2018-014, 2018.
[65] G. Aad *et al.* (ATLAS Collaboration), Phys. Lett. B **812**, 135991 (2021).
[66] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. D **101**, 052007 (2020).
[67] CMS Collaboration, Report No. CMS-PAS-SMP-16-010, 2017.
[68] A. M. Sirunyan *et al.* (CMS Collaboration), J. High Energy Phys. 11 (2018) 113.
[69] A. M. Sirunyan *et al.* (CMS Collaboration), J. High Energy Phys. 10 (2018) 161.
[70] S. Acharya *et al.* (ALICE Collaboration), J. High Energy Phys. 05 (2022) 061.
[71] M. Abdallah *et al.* (STAR Collaboration), Phys. Rev. D **104**, 052007 (2021).
[72] H. Klest, First measurement of groomed events shapes in ep dis using archived H1 data, REF (2022).
[73] A. H. Hoang, S. Mantry, A. Pathak, and I. W. Stewart, Phys. Rev. D **100**, 074021 (2019).
[74] B. Bachu, A. H. Hoang, V. Mateu, A. Pathak, and I. W. Stewart, Phys. Rev. D **104**, 014026 (2021).
[75] ATLAS Collaboration, Report No. ATL-PHYS-PUB-2021-034, 2021.
[76] S. Marzani, D. Reichelt, S. Schumann, G. Soyez, and V. Theeuwes, J. High Energy Phys. 11 (2019) 179.
[77] H. S. Hannesdottir, A. Pathak, M. D. Schwartz, and I. W. Stewart, J. High Energy Phys. 04 (2023) 087.
[78] A. H. Hoang, S. Mantry, A. Pathak, and I. W. Stewart, J. High Energy Phys. 12 (2019) 002.
[79] C. Lee and G. F. Sterman, eConf C **0601121**, A001 (2006).
[80] G. P. Korchemsky and G. F. Sterman, Nucl. Phys. **B437**, 415 (1995).
[81] G. P. Korchemsky, G. Oderda, and G. F. Sterman, AIP Conf. Proc. **407**, 988 (1997).
[82] G. P. Korchemsky and G. F. Sterman, Nucl. Phys. **B555**, 335 (1999).
[83] A. V. Belitsky, G. P. Korchemsky, and G. F. Sterman, Phys. Lett. B **515**, 297 (2001).
[84] M. Dasgupta, L. Magnea, and G. P. Salam, J. High Energy Phys. 02 (2008) 055.
[85] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, Phys. Rev. Lett. **114**, 092001 (2015).
[86] C. W. Bauer, S. Fleming, and M. E. Luke, Phys. Rev. D **63**, 014006 (2000).

[87] C. W. Bauer, S. Fleming, D. Pirjol, and I. W. Stewart, Phys. Rev. D **63**, 114020 (2001).
[88] C. W. Bauer and I. W. Stewart, Phys. Lett. B **516**, 134 (2001).
[89] C. W. Bauer, D. Pirjol, and I. W. Stewart, Phys. Rev. D **65**, 054022 (2002).
[90] A. Pathak, I. W. Stewart, V. Vaidya, and L. Zoppi, J. High Energy Phys. 04 (2021) 032.
[91] A. Pathak, J. High Energy Phys. 08 (2023) 054.
[92] T. Sjostrand, S. Mrenna, and P. Z. Skands, Comput. Phys. Commun. **178**, 852 (2008).
[93] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, J. High Energy Phys. 02 (2009) 007.
[94] M. Bahr *et al.*, Eur. Phys. J. C **58**, 639 (2008).
[95] S. Mantry, J. K. L. Michel, A. Pathak, and I. W. Stewart, Hadron-level predictions for boosted top quark jets with light grooming at NNLL (to be published).
[96] Z.-B. Kang, K. Lee, X. Liu, D. Neill, and F. Ringer, J. High Energy Phys. 02 (2020) 054.
[97] N. Fischer, S. Prestel, M. Ritzmann, and P. Skands, Eur. Phys. J. C **76**, 589 (2016).
[98] M. Cacciari, G. P. Salam, and G. Soyez, J. High Energy Phys. 04 (2008) 063.
[99] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012).
[100] A. E. Ferdinand, A. Pathak *et al.*, JETlib: A C++ Package for Efficient Analyses with Monte Carlo Event Generators (2019).
[101] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevD.108.L111501 for further details of the Monte Carlo hadronization models calibration analysis.
[102] P. T. Komiske, I. Moult, J. Thaler, and H. X. Zhu, Phys. Rev. Lett. **130**, 051901 (2023).
[103] J. Holguin, I. Moult, A. Pathak, and M. Procura, Phys. Rev. D **107**, 114002 (2023).
[104] K. Lee, B. Meçaj, and I. Moult, arXiv:2205.03414.
[105] A. Bhattacharya, M. D. Schwartz, and X. Zhang, Phys. Rev. D **106**, 074011 (2022).
[106] M. A. Ebert, J. K. L. Michel, F. J. Tackmann *et al.*, DESY-17-099 (2018). webpage: http://scetlib.desy.de.
[107] A. Pathak, SCETlib::SD: A C++ library for soft drop resummed observables in SCETlib (2022).