Extracting jet evolution via substructure measurements in pp collisions

J. Bielcikova,[1,2] H. Caines,[3] R. Kunnawalkam Elayavalli,[6] J. Putschke,[4] and M. Robotkova[1,2]
(STAR Collaboration)
[1] *Nuclear Physics Institute, Czech Academy of Sciences, 250 68 Prague, Czech Republic*
[2] *Czech Technical University in Prague, FNSPE, Prague, 115 19, Czech Republic*
[3] *Yale University, New Haven, Connecticut 06520*
[4] *Wayne State University, Detroit, Michigan 48201*
[5] *Vanderbilt University, Nashville, TN 37235*
[6] *Vanderbilt University, Nashville, Tennessee 37235*
(Dated: January 13, 2025 version 0)

Jets have long served as an experimental proxy for hard scattered quarks and gluons in high-energy particle collisions. Clustering techniques involved in jet finding allow for a systematic study of the internal structure of jets accessible at RHIC energies. We present multi-dimensional measurements of a varying suite of SoftDrop groomed jet substructure observables in *pp* collisions at $\sqrt{s} = 200$ GeV at STAR. The correlation between the splitting fraction $z_{\rm g}$ versus the groomed jet radius $R_{\rm g}$ at the first split highlight an inherent variance in jet shower topologies. For the first time, we present the z_g and R_g at the first, second and third identified SoftDrop splits along the harder branch as we travel along the jet shower for varying jet and initiator prong momenta. We observe a consistent trend of narrowing (angle) and hardening (energy fraction) of the splittings in the jet clustering tree which highlights enhanced sensitivity to non-perturbative corrections and restrictions in phase-space or virtuality for later splittings.

Introduction Quantum Chromo-Dynamics (QCD) is the established theory describing the interactions and dynamics of quarks and gluons, collectively referred to as partons. A fundamental feature of QCD is the evolution of its interaction strength as a function of energy scale or distance measure. The strong coupling constant α_s that serves as the interaction strength of QCD, has a characteristic exponential increase at low energies or large distance scales that makes the calculations diverge. This breakdown of perturbative expansion in QCD calculations results in the unique feature of quarks and gluons where they hadronize into color neutral particles. Jets originated as the first experimental evidence of quarks and gluons gathered from collimated sprays of hadrons in annihilation experiments of electrons and positrons [1–3]. These jets were understood as having arisen from the couples processes of parton shower and fragmentation/hadronization where the early time dynamics is described via perturbative QCD (pQCD) calculations, and the later times are fundamentally non-perturbative (npQCD) with the formation of hadrons. This is the reason why jets are often described as multi-scale objects where each jet necessarily traverses both the pQCD and npQCD regimes on its way from the hard scattering to the detector where it is observed. In the last two decades, significant progress has been made in our understanding of QCD at higher orders (NLO, NNLO $\cdots$) and varying length scales (NLL resummations []) due to the large volume of jet data from relativistic hadron–hadron colliders [4].

Jets are composite objects containing rich substructure information that can be exploited via jet finding algorithms [5]. Recent effort in the high-energy physics community has been in the area of developing novel experimental algorithms that translate a jet clustering tree to a theoretically motivated description of a parton shower [6–9]. These algorithms typically employ an iterative clustering procedure that generates a tree-like structure, which upon inversion, provides access to substructure at different steps along the cluster tree. The most common toolkit for such measurements is SoftDrop (SD) [8], which grooms away soft radiation at the edge of the jet cone, removing extreme asymmetrical splittings from the clustering trees expected to have large contribution from npQCD and are not associated with the original partonic jet. The SD algorithm employs a Cambridge/Aachen re-clustering of jet constituents [10, 11] and imposes a criterion at each step as one walks backwards in the de-clustered tree,

$$z_g = \frac{\min(p_{\rm T,1}, p_{\rm T,2})}{p_{\rm T,1} + p_{\rm T,2}} > z_{\rm cut} \left(\frac{R_{\rm g}}{R_{\rm jet}}\right)^{\beta} ; \quad R_{\rm g} = \Delta R(1,2), \tag{1}$$

where $1, 2$ are the two prongs at the current stage of declustering, $p_{\rm T}$ is the transverse momentum of the respective prong, $R_{\rm jet}$ is the jet resolution parameter and ΔR is the radial distance in the rapidity () and azimuthal angle (ϕ) plane. The free parameters in Eq. (1) are $z_{\rm cut}$, a momentum fraction threshold, and β, the angular exponent which are typically set to 0.1 and 0, respectively [12]. These parameter values make SoftDrop observables calculable in a Sudakov-safe manner, and at the infinite jet momentum limit they converge to the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) splitting functions [13–15].

Measurements from the Relativistic Heavy-Ion Collider (RHIC) and from the Large Hadron Collider (LHC) have shown that jet substructure observables, when calculated via SD grooming technique [16–25], allows one to immediately compare pQCD calculations at the first hard splitting without the need for large hadronization corrections. Since the jet clustering tree extends beyond the first split, one can iteratively apply the SD procedure on the hardest (highest $p_{\rm T}$) surviving branch and measure the jet substructure at each split along the declustered tree [26]. Such measurements would enable, for the first time, a time-differential study of the parton shower and evolution of both the momentum ($z_{\rm g}$) and angular scales ($R_{\rm g}$) within a jet. The 2-D representation of the momentum fraction and angular separation of these integrated splittings along all branches is known as the Lund Plane (LP) as has been measured in several different collaborations. The advantage of the LP is that it groups splitings of similar category such as perturbative, large angle or non-perturbative and soft in specific kinematic regions. In doing so, one necessarily integrates over the order of the splits which could carry important information regarding when specific changes occur to the splitting tree from a time aspect which is what we focus on in this current *letter*.

STAR recently measured the SoftDrop groomed shared momentum fraction ($z_{\rm g}$) and groomed jet radius ($R_{\rm g}$) at the first surviving split for jets of varying transverse momenta and jet radii [21]. These double differential measurements demonstrated a significant variation in $R_{\rm g}$ with increasing jet $p_{\rm T}$, reflecting momentum dependent narrowing of jet substructure, whereas $z_{\rm g}$ was found to vary only slowly and had a relatively constant shape for jet $p_{\rm T}$ > 30 GeV/c. In this Letter and a companion article [27], we present for the first time 3D measurements of SoftDrop groomed jet substructure observables and reconstruct a collection of observables corresponding to $z_{\rm g}^{n}$ and $R_{\rm g}^{n}$ at a given split n. We limit our measurement to the first three surviving splits within each jet and present the results fully corrected in 3D corresponding to the jet or initiator $p_{\rm T}$, $z_{\rm g}/R_{\rm g}$, and the split number n for jets of varying $p_{\rm T,jet}$ and for splits of varying initiator $p_{\rm T}$. This set of measurements serve as the first ever differential study of the self-similarity of the QCD splitting functions throughout the splitting tree.

Data set The data used in this analysis were collected by the STAR detector in pp collisions at $\sqrt{s} = 200$ GeV in 2012. Jets are clustered from two primary detectors contributing the charged and neutral energy compositions. Charged particle tracks and their momentum are reconstructed from hits in the Time Projection Chamber (TPC) [] while the transverse energy (E_T) of neutral hadrons is included by measuring the energy deposited in the Barrel Electromagnetic Calorimeter (BEMC) [], which has a tower size of 0.05 × 0.05 in azimuth ϕ and pseudorapidity η. To avoid double-counting, the energy deposited by charged hadrons in the BEMC is accounted for by full hadronic correction, in which the transverse momentum of any charged-particle track that extrapolates to a tower is subtracted from the transverse energy of that tower. Tower energies are set to zero if they would otherwise become negative via this correction. Both the TPC and the BEMC uniformly cover the full azimuth and a pseudorapidity range of $\eta < 1$.

Events were selected by an online jet patch trigger in the BEMC, which required an uncorrected sum patch ADC value above a certain threshold, corresponding to $\sum E_T > 7.3$ GeV, in one of 18 partially overlapping 1.0×1.0 in (η, ϕ) groupings of towers. Events are restricted to have a primary vertex position along the beam axis of $v_z \leq 30$ cm. All charged-particle track and tower selections are consistent with previous publications with this dataset from STAR and available in [21].

Analysis methods At the first split, the observables are represented in a three-dimensional space defined by the distributions of $z_{\rm g}$ vs. $R_{\rm g}$ vs. jet $p_{\rm T}$. These distributions are unfolded using the Iterative Bayesian unfolding method [28]. The detector response is estimated via PYTHIA 6 (Perugia 2012 tune [29] and further tuned to STAR data [30]) events passed through a GEANT3 simulation of the STAR detector. These simulated events are embedded into zero-bias *pp* data and the resulting events are analyzed in a similar fashion to the real data. Jet matching is performed by requiring the angular distance between jets to satisfy $\Delta R < 0.4$.

Since the splits are identified at the detector level, detector effects on the jet clustering tree could destroy the split hierarchy, i.e. splits at the particle level can be lost or mis-categorized in the detector-jet clustering tree, along with the addition of fake splits arising from particles of uncorrelated sources, such as interactions with detector material. To correct the split hierarchy, we introduce an additional matching requirement of the splits based on the initiator prong at the particle and detector-level via $\Delta R(\text{initiator}_{\rm det,part}) < 0.1$ to build a hierarchy matrix with particle-level splits on the $x-$axis and detector-level splits on the $y-$axis. The hiererchy matrix of the splits is an established procedures in similar measurements of the LP across various systems.

The systematic uncertainties follow the same procedure outlined in [21], and are broadly grouped into two categories: detector performance and analysis procedure. The former sources of uncertainties constitute variations of the tracking efficiency by ±4% and tower energy scale by ±3.8%. The systematic uncertainty due to the analysis procedure includes hadronic correction, i.e. correcting 100% to 50% of the matched track's momentum from a tower's energy to avoid double counting of energy depositions. Uncertainty due to the unfolding procedure is taken as the maximal envelope of variations in the iteration parameter and shape uncertainties arising from the prior (varied by the differences to PYTHIA 8 [31] and HERWIG 7 [32]). Lastly, the split matching criterion is varied by ± 0.025 and the consequent variation to the

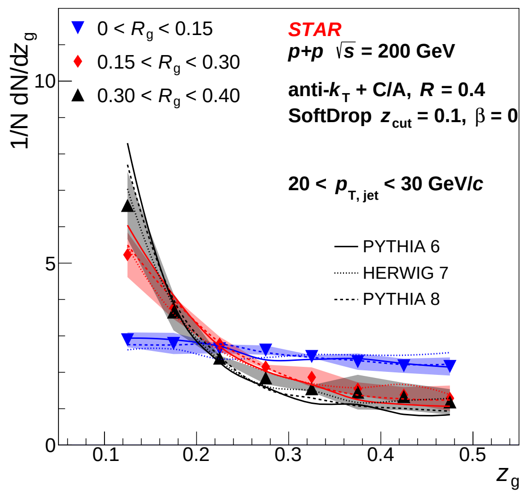

FIG. 1. Fully corrected $z_{\rm g}$ distributions for three bins (see legend) for jets with transverse momentum $p_{\rm T,jet} = 20$–30 GeV$/c$ and $R = 0.4$ in pp collisions at $\sqrt{s} = 200$ GeV. The data are also compared with Monte Carlo generators PYTHIA and HERWIG (see legend).

fully corrected result is taken as a shape uncertainty.

Results - Correlations at the first split

In Figure 1, the fully corrected iterative SoftDrop $z_{\rm g}$ distributions at the first split along the parton shower are displayed for jets with transverse momentum $p_{\rm T,jet} = 20$–30 GeV$/c$ reconstructed with the resolution parameter $R = 0.4$. The distributions are shown separately for three distinct $R_{\rm g}$ bins, and the bands surrounding the data points represent systematic uncertainties. The data reveal a strong dependence of $z_{\rm g}$ on the $R_{\rm g}$ value. For small $R_{\rm g}$ values ($R_{\rm g} = 0 - 0.15$), the corresponding $z_{\rm g}$ distribution is essentially flat implying equal probability for selection of hard or soft splittings. Consequently, larger $R_{\rm g}$values $R_{\rm g} = 0.15 - 0.3$ and $R_{\rm g} = 0.3 - 0.4$ the $z_{\rm g}$ distributions gradually regains its pQCD inspired steeply falling shape with enhanced probabilities at small $z_{\rm g}$ values, suggesting a preference for softer wide-angle splittings as the first emissions along the jet shower.

These distributions are also compared with leading-order Monte-Carlo generators with different implementations of parton shower and hadronization mechanisms. The models considered include PYTHIA 6 with the STAR Perugia tune, PYTHIA 8 with the Monash tune based on LHC data, and HERWIG 7 with a modified UE-EE-4-CTEQ6L1 tune with the reference energy for underlying event estimation set to RHIC. While PYTHIA utilizes either kT or pT ordering, HERWIG employs an angular-ordered parton shower. For hadronization, PYTHIA utilizes the Lund string model, while HERWIG employs the cluster model. All three MC models capture the overall trend of the data well indicating a consistent trend of harder splits arrising from narrower emissions.

Results - Splittings along the jet shower

In Figure 2, we report the fully corrected iterative SoftDrop $z_{\rm g}$ (top) and $R_{\rm g}$ (bottom) distributions for the first, second and third splits along the jet clustering tree. As before the distributions are reported for jets with transverse momentum $p_{\rm T,jet} = 20$–30 GeV$/c$ (left) and 30–50 GeV$/c$ (right) reconstructed with the resolution parameter $R = 0.4$ and the systematic uncertainties are represented by bands around the respective data points. We observe a significant modification of the shape of $z_{\rm g}$ and $R_{\rm g}$ distributions as we travel along the jet shower from the first to the third split due to a constriction of the available phase space for radiations. While at the first split, the $z_{\rm g}$distribution is increasing steep at low $z_{\rm g}$ values, at later splits it starts to flatten. The $R_{\rm g}$ consequently shows that with increasing number of the split along the parton shower, we observe narrower distributions with their peak position shifts toward smaller $R_{\rm g}$ values. Such an evolution can be connected to the jet's virtuality and its subsequent evolution from hard scattering scale (Q^2) to the hadronization scale ($\Lambda_{\rm QCD}$).

Comparison of the data with leading order MC event generators again demonstrates an overall qualitative agreement with the data albeit slight differences at the first split exist which are reduced for the second and third splits.

This measurement serves as evidence for a significant correlation between the shape of the splitting fractions and the opening angles within jets or the split number with a consistent quantitative picture emerging of jet structure where later splits are narrow and harder in energy while early splits are wider and softer.

Conclusions We have presented for the first time 3D corrected SoftDrop groomed studies of $z_{\rm g}$ vs. $R_{\rm g}$ distributions for jets produced in pp collisions at 200 GeV at RHIC at the first split, and the distributions of $z_{\rm g}$ and $R_{\rm g}$ for the first, second, and third splits, respectively. Notably, we observe a striking resemblance between the trends of the $z_{\rm g}$ distribution at the first split with small $R_{\rm g}$ and the $z_{\rm g}$ distribution at the third split which is consistent with angular ordering. Flattening of the $z_{\rm g}$ distribution is also indicative of enhanced correction to pQCD style description of vacuum splits. Armed with this knowledge, one can select specific topologies of jets with predominantly earlier or later splits and facilitate a multi-prong comparison of data with varying MC generators with different perturbative (parton showers) and non-perturbative (hadronization, multi-parton interactions) implementations to highlight the transition between the two regions of the jet shower. This technique opens up the exciting possibility of space-time tomography in AA collisions and enables differential measurements of jet energy loss for specific substructure.

We thank the RHIC Operations Group and RCF at BNL, the NERSC Center at LBNL, and the Open Science Grid consortium for providing resources and support. This work was supported in part by the Office of Nuclear Physics within the U.S. DOE Office of Science, the

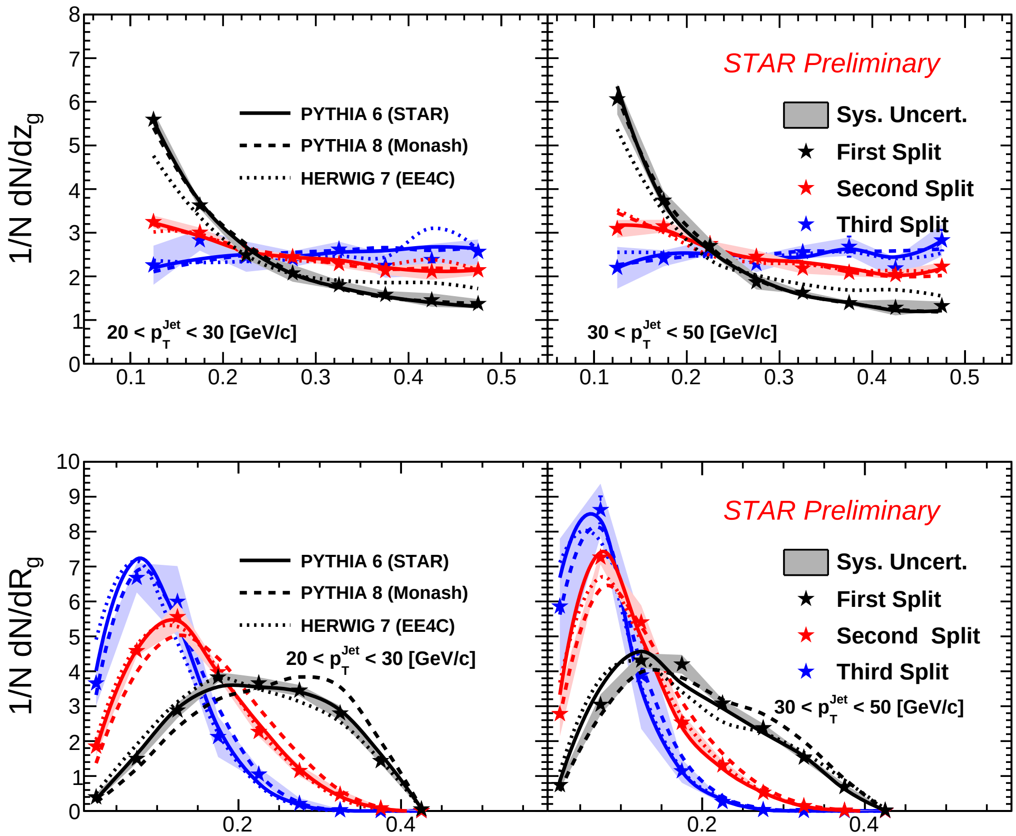

FIG. 2. Fully unfolded z_g (top) and R_g (bottom) distributions for different splits (see legend) of jets with transverse momentum $p_{T,jet} = 20$–30 GeV/c (left) and 30–50 GeV/c (right) for $R = 0.4$ in pp collisions at $\sqrt{s} = 200$ GeV. The data are also compared with Monte Carlo generators PYTHIA and HERWIG (see legend).

U.S. National Science Foundation, National Natural Science Foundation of China, Chinese Academy of Science, the Ministry of Science and Technology of China and the Chinese Ministry of Education, the Higher Education Sprout Project by Ministry of Education at NCKU, the National Research Foundation of Korea, Czech Science Foundation and Ministry of Education, Youth and Sports of the Czech Republic, Hungarian National Research, Development and Innovation Office, New National Excellency Programme of the Hungarian Ministry of Human Capacities, Department of Atomic Energy and Department of Science and Technology of the Government of India, the National Science Centre and WUT ID-UB of Poland, the Ministry of Science, Education and Sports of the Republic of Croatia, German Bundesministerium für Bildung, Wissenschaft, Forschung and Technologie (BMBF), Helmholtz Association, Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) and Agencia Nacional de Investigación y Desarrollo (ANID) of Chile.

[1] R. Brandelik *et al.* (TASSO), Phys. Lett. B **86**, 243 (1979).
[2] E. M. Riordan, Science **256**, 1287 (1992).
[3] J. Ellis, Int. J. Mod. Phys. A **29**, 1430072 (2014), arXiv:1409.4232 [hep-ph].
[4] A. Ali and G. Kramer, Eur. Phys. J. H **36**, 245 (2011), arXiv:1012.2288 [hep-ph].
[5] S. Marzani, G. Soyez, and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, Vol. 958 (Springer, 2019) arXiv:1901.10342 [hep-ph].
[6] M. Dasgupta and G. P. Salam, Phys. Lett. B **512**, 323 (2001), arXiv:hep-ph/0104277.
[7] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, JHEP **09**, 029 (2013), arXiv:1307.0007 [hep-ph].
[8] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, JHEP **05**, 146 (2014), arXiv:1402.2657 [hep-ph].
[9] Y. Mehtar-Tani, A. Soto-Ontoso, and K. Tywoniuk, Phys. Rev. D **101**, 034004 (2020), arXiv:1911.00375 [hep-ph].
[10] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, JHEP **08**, 001 (1997), arXiv:hep-ph/9707323.
[11] M. Wobisch and T. Wengler, in *Workshop on Monte*

Carlo Generators for HERA Physics (Plenary Starting Meeting) (1998) pp. 270–279, arXiv:hep-ph/9907280.
[12] A. J. Larkoski, S. Marzani, and J. Thaler, Phys. Rev. **D91**, 111501 (2015), arXiv:1502.01719 [hep-ph].
[13] V. N. Gribov and L. N. Lipatov, Phys. Lett. B **37**, 78 (1971).
[14] Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).
[15] G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).
[16] A. M. Sirunyan *et al.* (CMS), Phys. Rev. Lett. **120**, 142302 (2018), arXiv:1708.09429 [nucl-ex].
[17] A. M. Sirunyan *et al.* (CMS), JHEP **10**, 161 (2018), arXiv:1805.05145 [hep-ex].
[18] G. Aad *et al.* (ATLAS), Phys. Rev. D **101**, 052007 (2020), arXiv:1912.09837 [hep-ex].
[19] S. Acharya *et al.* (ALICE), Phys. Lett. B **802**, 135227 (2020), arXiv:1905.02512 [nucl-ex].
[20] G. Aad *et al.* (ATLAS), Phys. Rev. Lett. **124**, 222002 (2020), arXiv:2004.03540 [hep-ex].
[21] J. Adam *et al.* (STAR), Phys. Lett. B **811**, 135846 (2020), arXiv:2003.02114 [hep-ex].
[22] M. Abdallah *et al.* (STAR), Phys. Rev. D **104**, 052007 (2021), arXiv:2103.13286 [hep-ex].
[23] S. Acharya *et al.* (ALICE), JHEP **05**, 061 (2022), arXiv:2107.11303 [nucl-ex].
[24] S. Acharya *et al.* (ALICE), JHEP **05**, 244 (2023), arXiv:2204.10246 [nucl-ex].
[25] S. Acharya *et al.* (ALICE), JHEP **07**, 201 (2023), arXiv:2211.08928 [nucl-ex].
[26] F. A. Dreyer, L. Necib, G. Soyez, and J. Thaler, JHEP **06**, 093 (2018), arXiv:1804.03657 [hep-ph].
[27] STAR, Phys. Rev. D.
[28] G. D'Agostini, Nucl. Instrum. Meth. **A362**, 487 (1995).
[29] T. Sjostrand, S. Mrenna, and P. Z. Skands, JHEP **05**, 026 (2006), arXiv:hep-ph/0603175 [hep-ph].
[30] J. Adam *et al.* (STAR), Phys. Rev. D **98**, 032011 (2018), arXiv:1805.09742 [hep-ex].
[31] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, Comput. Phys. Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph].
[32] J. Bellm *et al.*, Eur. Phys. J. **C76**, 196 (2016), arXiv:1512.01178 [hep-ph].