**Title**
Identifying groomed jet splittings in heavy-ion collisions



**Journal**
Physical Review C, 102(4)

**ISSN**
2469-9985

**Authors**
Mulligan, James
Płoskoń, Mateusz

**Publication Date**
2020-10-01

**DOI**
10.1103/physrevc.102.044913

Peer reviewed

# Identifying groomed jet splittings in heavy-ion collisions

James Mulligan [1,2] and Mateusz Płoskoń [2]
[1]*Physics Department, University of California, Berkeley, California 94720, USA*
[2]*Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

(Received 1 July 2020; accepted 28 September 2020; published 28 October 2020)

Measurements of jet substructure in heavy-ion collisions may provide key insight into the nature of jet quenching in the quark-gluon plasma. Jet grooming techniques from high-energy physics have been applied to heavy-ion collisions in order to isolate theoretically controlled jet observables and explore possible modification to the hard substructure of jets. However, the grooming algorithms used have not been tailored to the unique considerations of heavy-ion collisions, in particular to the experimental challenge of reconstructing jets in the presence of a large underlying event. We report a set of simple studies illustrating the impact of the underlying event on identifying groomed jet splittings in heavy-ion collisions, and on associated groomed jet observables. We illustrate the importance of the selection of the grooming algorithm, as certain groomers are more robust against these effects, while others, including those commonly used in heavy-ion collisions, are susceptible to large background effects, which, when uncontrolled, can mimic a jet-quenching signal. These experimental considerations, along with appropriate theoretical motivation, provide input to the choice of grooming algorithms employed in heavy-ion collisions.

DOI: 10.1103/PhysRevC.102.044913

## I. INTRODUCTION

Jet grooming techniques were developed in the high-energy physics community to mitigate pileup contamination and improve the theoretical calculability of jet observables in $pp$ collisions. The Soft Drop algorithm, for example, reduces nonperturbative effects by selectively removing soft large-angle radiation, which allows for well-controlled comparisons of measurements with perturbative QCD (pQCD) calculations [1–3]. Grooming techniques have recently been applied to heavy-ion collisions in order to establish whether jet quenching in the quark-gluon plasma modifies the hard substructure of jets, such as the splitting function, and to elucidate whether jets lose energy coherently, as a single color charge, or incoherently, as multiple independent substructures [4–12]. Moreover, Monte Carlo (MC) event generators suggest that jet splittings identified by grooming algorithms are correlated to parton shower splittings, raising the possibility that identifying groomed jet splittings in heavy-ion collisions may provide a handle on the space-time evolution of jet propagation through the hot QCD medium.

Measurements of the Soft Drop groomed momentum fraction $z_g$ have been made in $pp$ and heavy-ion collisions at the Large Hadron Collider (LHC) and the Relativistic Heavy Ion Collider (RHIC) [13–17]. These measurements have opened a new avenue in heavy-ion jet physics. Measurements by the CMS and ALICE Collaborations show a modification of the $z_g$ distribution in Pb-Pb collisions relative to $pp$ collisions—however, the results have not been corrected for background effects. Local background fluctuations in a heavy-ion environment can result in an incorrect splitting (unrelated to the jet) being identified by the grooming algorithm. This problem is analogous to the well-known experimental problem of "combinatorial" jets in heavy-ion collisions, which is typically treated by either (1) reporting jet measurements in the background-free region of phase space, namely, at sufficiently large $p_T$ and/or small $R$, or (2) subtracting the combinatorial jet distribution on an ensemble basis. In the case of groomed jet observables, the scale at which background effects occur is set by the subleading prong of the groomed jet, rather than the jet $p_T$ and $R$. The presence of background contamination in groomed jet observables has been recognized to some extent since the first measurements in heavy-ion collisions; however, the magnitude of the effect has not been quantified, nor has its qualitative impact been understood. Since the reported distributions contain a significant number of "mistagged" splittings, it remains unclear how to interpret the observed modifications.

Since the characteristic scale of these effects is set by the subleading prong of the groomed jet, the impact of local background fluctuations on groomed jet observables depends on the grooming algorithm employed. In this article, we present a simple set of studies on the performance of various grooming algorithms with respect to background contamination effects in heavy-ion collisions in order to confront the experimental question: How are grooming algorithms affected by the presence of a heavy-ion background? We identify groomers that are relatively robust against background effects, as well

as those that are susceptible to contamination. Finally, we discuss implications on the interpretation of previous measurements.

## II. ANALYSIS SETUP

We reconstruct jets from charged particles in central rapidity generated by PYTHIA [18] for proton-proton collisions at $\sqrt{s} = 5$ TeV using the anti-$k_{\mathrm{T}}$ algorithm from the FASTJET [19] package with resolution parameter $R = 0.4$. Before the jet finding, we select particles with $p_T > 0.15$ GeV/$c$. This setup corresponds to typical experimental configurations at the LHC. To approximate the heavy-ion background, we use a thermal model consisting of $N$ particles drawn from a Gaussian with $\langle \frac{dN}{d\eta} \rangle \approx 1800$ and $p_{\mathrm{T}}$ sampled from a Gamma distribution: $f_\Gamma(p_{\mathrm{T}}; \alpha, \beta) \propto p_{\mathrm{T}}^{\alpha-1} e^{-p_{\mathrm{T}}/\beta}$ with $\alpha = 2$. We select $\beta = 0.5$ to roughly fit the width of the $R = 0.4$ $\delta p_{\mathrm{T}}$ distribution in 0–10% Pb-Pb data of $\sigma \approx 11$ GeV/$c$ [20]. We perform event-wide constituent subtraction on the combined event consisting of the charged particles from the PYTHIA event together with the thermal background particles, using $R_{\mathrm{max}} = 0.25$ [21]. We then cluster the subtracted particles into jets and match these "combined" jets to those jets found by clustering only the PYTHIA particles.

### A. Groomers

To study the performance of different grooming criteria, we use the Soft Drop algorithm [1] and the Dynamical Grooming algorithm [22,23] but also new rather simple groomers which we call max-$z$, max-$p_T^{\mathrm{soft}}$, max-$\kappa$, max-$k_T$, and min-$t_f$. These are all defined by reclustering the jet with the Cambridge-Aachen (CA) algorithm, where every step of the clustering history is defined by a radiator and the two prongs that it decays to. We denote the two prongs $a$ and $b$ such that $p_T^{\mathrm{radiator}} = p_T^a + p_T^b$, where $p_T^b < p_T^a$, and $R_g = [(y_a - y_b)^2 + (\varphi_a - \varphi_b)^2]^{1/2}$ is the angular separation between the two (used interchangeably with $\theta_{\mathrm{g}} \equiv R_g/R$) with $\varphi$ being the azimuthal angle and $y$ the rapidity of the prongs. Therefore, $k_T \equiv p_T^b R_g$, $z \equiv p_T^b / p_T^{\mathrm{radiator}}$, and $\kappa \equiv z R_g$. We briefly describe the algorithms that we use below:

i. Soft Drop with $\beta = 0$ with three values of the symmetry parameter $z_{\mathrm{cut}} = 0.1, 0.2, 0.3$.
ii. Dynamical Grooming with three values of the grooming parameter $a = 0.1, 1.0, 2.0$.
iii. max-$z$: For every jet that contains more than one particle, identify the splitting where $z$ is the largest from all the splittings in the primary Lund plane.
iv. max-$p_T^{\mathrm{soft}}$: For every jet that contains more than one particle, identify the splitting where the soft prong has the largest $p_{\mathrm{T}}$ from all of the softer prongs within any pair in the primary Lund plane.
v. max-$\kappa$: For every jet that contains more than one particle, identify the splitting where $\kappa$ is the largest from all splittings in the primary Lund plane.
vi. max-$k_T$: For every jet that contains more than one particle, identify the splitting where $k_T$ is the largest from all splittings in the primary Lund plane.
vii. min-$t_f$: For every jet that contains more than one particle, identify the splitting where $zR_g^2$ is the largest from all the splittings in the primary Lund plane (in relation to the estimate of the formation time for the pair $t_f \sim \frac{1}{zR_g^2}$).

For an overview of the phase space that each of the grooming algorithms selects, we plot the primary Lund plane density $\rho(\kappa, R_g) = \frac{1}{N_{jet}} \frac{d^2N}{d\ln(\kappa)/d\ln(1/R_g)}$ for identified splittings in Fig. 1 [24]. We note that several of these groomers are expected to select similar phase space: max-$z$, max-$p_T^{\mathrm{soft}}$, and Dynamical Grooming $a = 0.1$ select approximately on the longitudinal momentum of the splitting; max-$\kappa$, max-$k_T$, and Dynamical Grooming $a = 1.0$ select approximately on the transverse momentum of the splitting; min-$t_f$ and Dynamical Grooming $a = 2.0$ select approximately on the mass of the splitting.

### B. Prong matching

To study the impact of the heavy-ion background on the reconstruction of groomed splittings, we examine where >50% of the PYTHIA subleading prong (by $p_{\mathrm{T}}$) is reconstructed in the combined event. We consider only the case where both the PYTHIA jet and the combined jet pass the grooming condition. We categorize six possibilities—the PYTHIA subleading prong has one of the following characteristics:

(1) It is correctly reconstructed in the subleading prong of the combined jet.
(2) It is reconstructed in the leading prong of the combined jet, and the PYTHIA leading prong is reconstructed in the subleading prong of the combined jet. That is, both prongs are correctly identified, but they "swap" that which is leading and that which is subleading. In this case, $z_{\mathrm{g}}$ and $\theta_{\mathrm{g}}$ are invariant—although iterative observables are not.
(3) It is reconstructed in the leading prong of the combined event, and the PYTHIA leading prong is not reconstructed in the subleading prong of the combined event. This is the most common way that an incorrect splitting is reconstructed, typically by a background fluctuation at large angle passing the grooming condition. Due to angular clustering, this by definition results in the subleading prong being absorbed in the leading prong, as shown in Fig. 2.
(4) It is reconstructed in the groomed-away constituents of the combined jet.
(5) It is reconstructed nowhere in the combined jet, but rather its constituents are elsewhere in the combined event.
(6) It is not reconstructed in any of the above categories; for example, it may have 1/3 of its $p_{\mathrm{T}}$ split between three categories.

## III. PERFORMANCE OF GROOMERS

For each groomer, we plot the fraction of subleading prongs in the combined events that are correctly tagged in

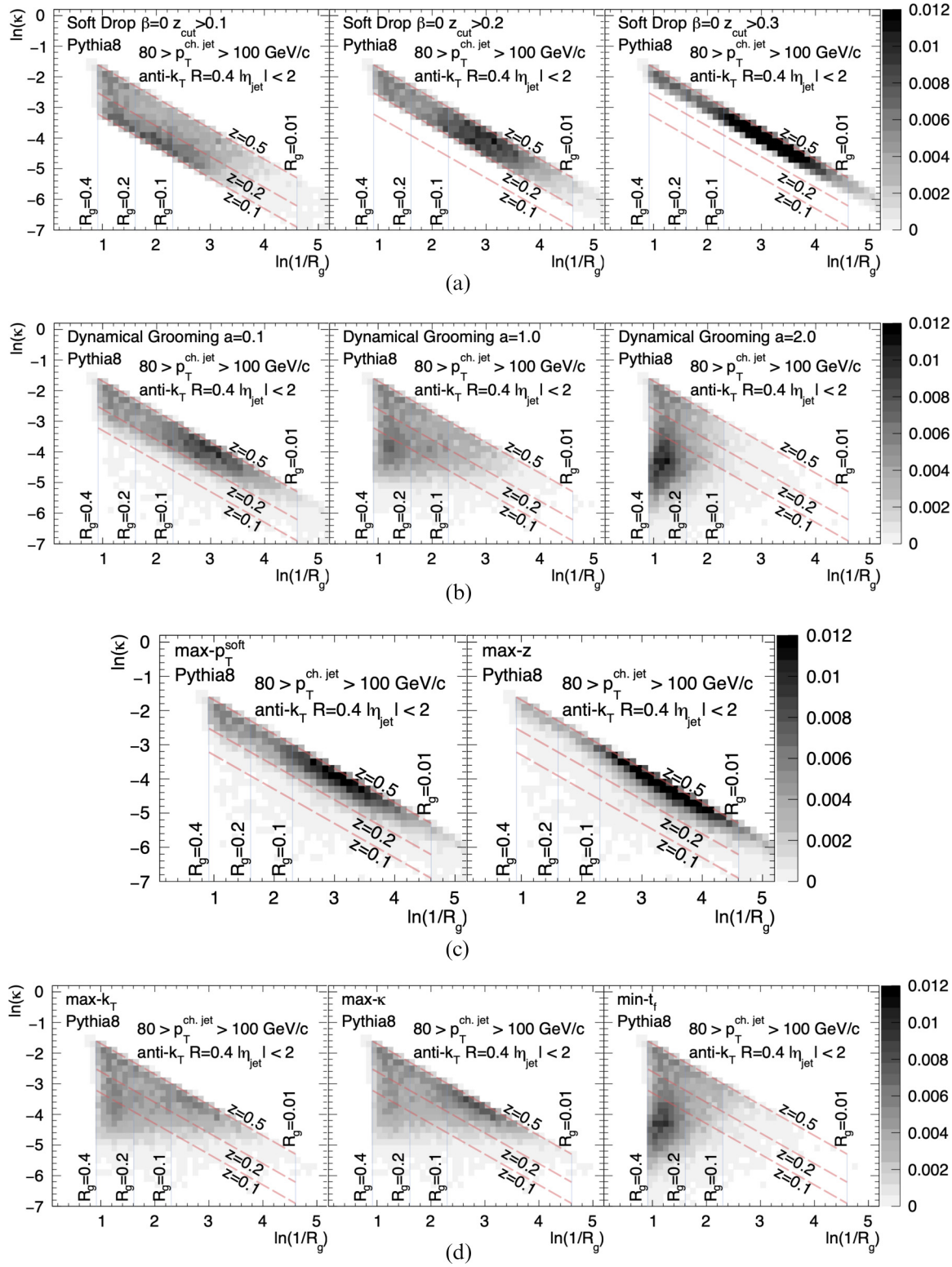


FIG. 1. Primary Lund plane density diagram of groomed splittings for various groomers. Events were generated using PYTHIA for proton-proton collisions at $\sqrt{s} = 5$ TeV. Jets reconstructed from charged particles at hadron-level: (a) Primary Lund plane obtained with Soft Drop grooming with $\beta = 0$ for different symmetry-cut $z_{\rm cut}$ parameters. (left) $z_{\rm cut} = 0.1$. (middle) $z_{\rm cut} = 0.2$. (right) $z_{\rm cut} = 0.3$. (b) Primary Lund plane obtained with Dynamical Grooming for different values of $a$. (left) $a = 0.1$. (middle) $a = 1.0$. (right) $a = 2.0$. (c) Primary Lund plane obtained with new groomers with the split selection depending on momentum of the prongs. (left) max-$p_{\rm T}^{\rm soft}$. (right) max-$z$. (d) Primary Lund plane obtained with new groomers with the split selection depending on momentum and the angle between the prongs. (left) max-$k_T$. (middle) max-$\kappa$. (right) min-$t_f$.

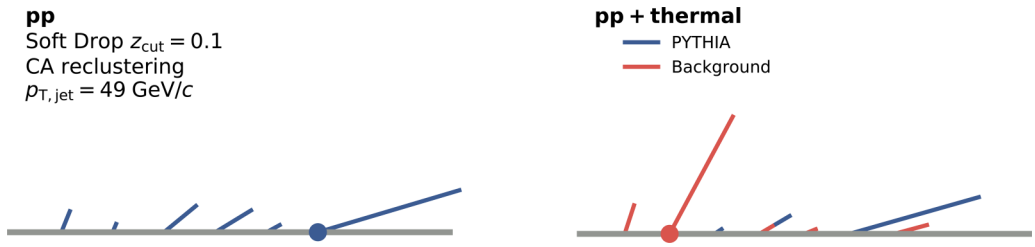


FIG. 2. Example of a PYTHIA jet (left) and the same jet embedded into the thermal background (right). In the case of thermal background, a background fluctuation at large angle passing the grooming condition results in the subleading prong being absorbed into the leading prong.

Fig. 3 as a function of jet $p_T$. Immediately, it is apparent that to increase the subleading prong purity one should (i) choose a suitable groomer, and/or (ii) measure high-$p_T$ jets. Groomers with an angular selection perform the worst, which is unsurprising given that combinatorial background preferentially occupies large-angle phase space, as compared with jets. Groomers which select on longitudinal momentum (Dynamical Grooming $a = 0.1$, max-$p_T^{soft}$, max-$z$) perform well, with Dynamical Grooming performing slightly worse, presumably due to its small angular component in the grooming condition. Soft Drop performs similarly to these for $z_{cut} = 0.2, 0.3$, where above $p_T = 70$ GeV/$c$ there appears to be an approximate saturation, in which case further increasing $z_{cut}$ does not increase the purity. Soft Drop with $z_{cut} = 0.1$, which is the most common configuration used in heavy-ion collisions, performs notably worse. This suggests that mistagged splittings arise from a characteristic longitudinal momentum scale above which background is suppressed due to uncorrelated background fluctuations on the geometric scale of a prong. These results were repeated by using Angantyr [25] to model the underlying heavy-ion event, and similar results were obtained with identical ordering of the groomers and purities within approximately 10% compared with the thermal background.

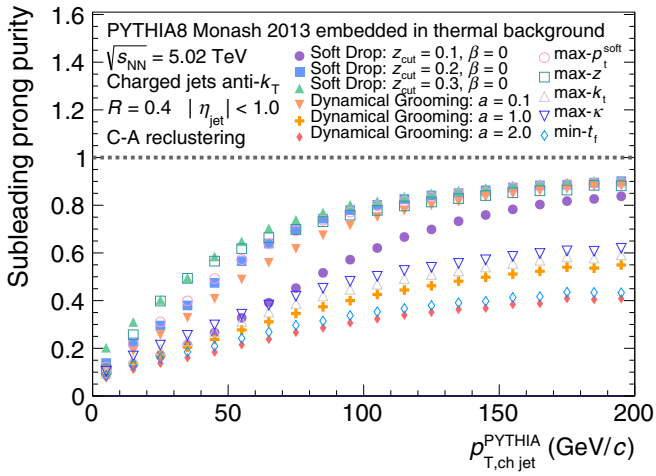


FIG. 3. Subleading prong purity as a function of $p_T$ for a variety of groomers. Note that the purity defined here includes only those cases where the PYTHIA subleading prong is correctly tagged as the subleading prong in the combined jet.

To determine the dependence of the mistagged fraction on the splitting observables, we decompose the distributions of $z_g$, $\theta_g$ according to where the PYTHIA subleading prong is reconstructed in the combined event, as described in Sec. II B. Figure 4 shows the $z_g$ (left) and $\theta_g$ (right) distributions when PYTHIA is embedded in the heavy-ion background. For smaller $z_{cut}$ and lower $p_T$ (top row), there is a large fraction of mistagged splittings, predominantly from the case where the subleading prong is mistagged in the leading prong (Fig. 2). The mistagged prongs are most prominent at small $z$ (where the true $z_g$ distribution is naturally peaked) and large $\theta$ (in the tail of the true $\theta_g$ distribution); however, they are not limited to these regions of phase space. We note that the correctly tagged distributions exhibit significant deviations from the true distributions, suggesting that there are strong correlations between the structure of the jet and its susceptibility to mistagging. By raising $z_{cut}$ (middle row) or increasing $p_T$ (bottom row), the mistagging rates are significantly reduced—suggesting that, at low-$p_T$, the Soft Drop groomer with $z_{cut} = 0.1$ is undesirable in heavy-ion collisions, and even with larger $z_{cut}$ or higher $p_T$ one should proceed with caution. The bottom panels of Fig. 4 show the fraction of subleading prongs in the embedded events that are correctly tagged, which is denoted as tagging purity [where we now include cases (1) and (2) from Sec. II B as correct identification]. We additionally plot the ratio of the embedded distribution to the true distribution, which shows significant deviations, typically larger for $\theta_g$ than $z_g$.

To investigate the robustness of the choice of grooming algorithm to these experimental background effects, we plot the two ratios from the bottom panels of Fig. 4 for a variety of groomers. In Fig. 5, we plot the subleading prong tagging purity. For $z_g$ (left), the purity is high at large $z_g$ but decreases substantially at small $z_g$. For $\theta_g$ (right), on the other hand, the purity is typically highest at low $\theta_g$, and decreases at large $\theta_g$. Groomers which select on the longitudinal hardness of the splitting (Soft Drop, Dynamical Grooming $a = 0.1$, max-$p_T^{soft}$, and max-$z$) perform the best; however, even in these cases the purity becomes low when the absolute scale of $z$ becomes small (Soft Drop $z_{cut} = 0.1$, and all others for $z_g$ small). Of the groomers considered here, Soft Drop is the only one with an absolute cutoff in the grooming condition, which constrains the observable to the high-purity region. This, in combination with the well-studied theoretical benefits of Soft Drop, suggests that Soft Drop with sufficiently large $z_{cut}$ is an appealing groomer for heavy-ion collisions. We note, however, that in this $p_T$ range, the purity remains significantly less than unity, which must be treated carefully. Nevertheless, by maximizing

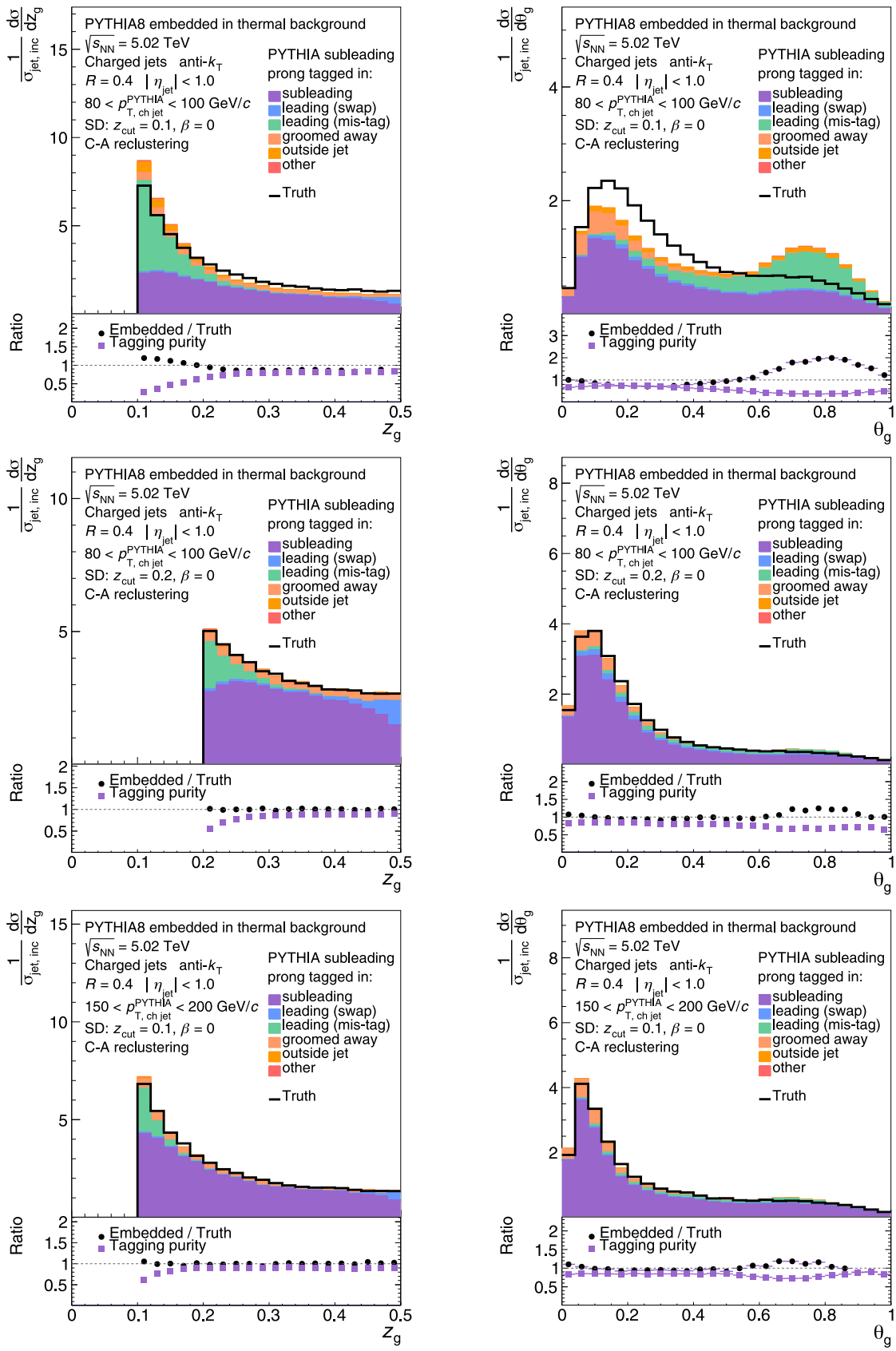


FIG. 4. Distributions of $z_g$ (left) and $\theta_g$ (right) when PYTHIA is embedded in the heavy-ion background, as well as from PYTHIA ("Truth"). The bottom panels show the purity and the ratio of the embedded distribution to the PYTHIA distribution. (top) Low $p_T$, $z_{cut} = 0.1$. (middle) Low $p_T$, $z_{cut} = 0.2$. (bottom) High $p_T$, $z_{cut} = 0.1$.

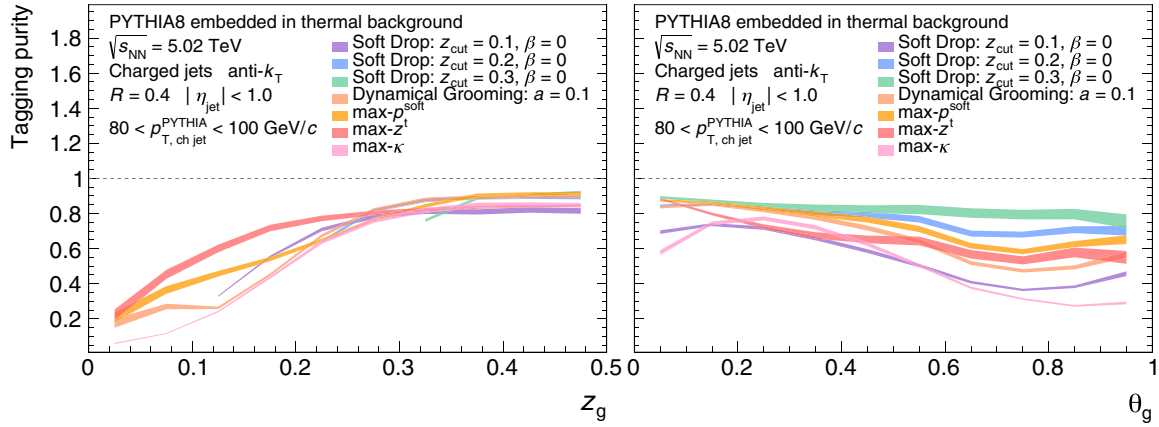


FIG. 5. Subleading prong purities as a function of $z_g$ (left) and $\theta_g$ (right) for a variety of groomers.

the purity, one can achieve improved experimental control, both by reducing the magnitude of corrections and modeling needed in the measurement, but also by enabling a stable unfolding procedure due to the rejection of large off-diagonal contamination of the response matrix, which is otherwise often unfeasible.

Figure 6 shows the ratio of the embedded $z_g$ and $\theta_g$ distributions to the PYTHIA distributions for a variety of groomers. This provides complementary information to the purity since it describes the impact not only of the fraction of mistagged splittings but also of how different the mistagged splittings are from the true splittings. Similar to the purity, the Soft Drop $z_{cut} = 0.1$ and max-$\kappa$ groomers perform poorly, whereas the other groomers perform relatively well. We see that this ratio is typically nearer to unity for $z_g$ compared with $\theta_g$, since for $z_g$ the mistagged splittings typically deplete and repopulate the low-$z$ region, whereas for $\theta_g$ the mistagged splittings are likely to populate large angles.

Finally, we note that the choice of reclustering algorithm can have a large impact on the splitting purity. To illustrate this, in Fig. 7 (left) we plot the mistagging distribution as a function of $\theta_g$ for Soft Drop $z_{cut} = 0.1$ with anti-$k_T$ (AKT) reclustering. Compared with CA reclustering (Fig. 4, top right), the purity is improved by approximately 20% at intermediate $p_T$, and large-angle mistagging is absent. This behavior can be understood since the anti-$k_T$ reclustering sequence is fundamentally different than that of CA. The anti-$k_T$ algorithm tends to cluster branches where at least one branch has large $p_T$, resulting in a clustering sequence dominated by the leading prong clustering together with surrounding individual particles—as compared with CA, which allows softer particles to cluster among themselves before combining with the leading prong. Accordingly, anti-$k_T$ reclustering has on average a larger number of primary splittings compared with CA reclustering. Since background prongs typically arise from local fluctuations of particle number at large angle, anti-$k_T$ reclustering leads to enhancement of the purity as large-angle particles are individually clustered and then groomed away. This is illustrated in Fig. 7 (right) for the same PYTHIA jet as in Fig. 2. However, it is important to note that anti-$k_T$ reclustering has certain theoretical drawbacks [26] and may therefore be undesirable. Nevertheless, due to the observed benefits with regard to background contamination, it may be worth further theoretical consideration.

## IV. RELEVANCE TO PREVIOUS MEASUREMENTS

In this section, we briefly outline the implications of our studies on the interpretation of published measurements of $z_g$ [15,16]. These measurements are reported without corrections for background effects or detector effects, but rather Pb-Pb data is compared with an embedded reference. In both

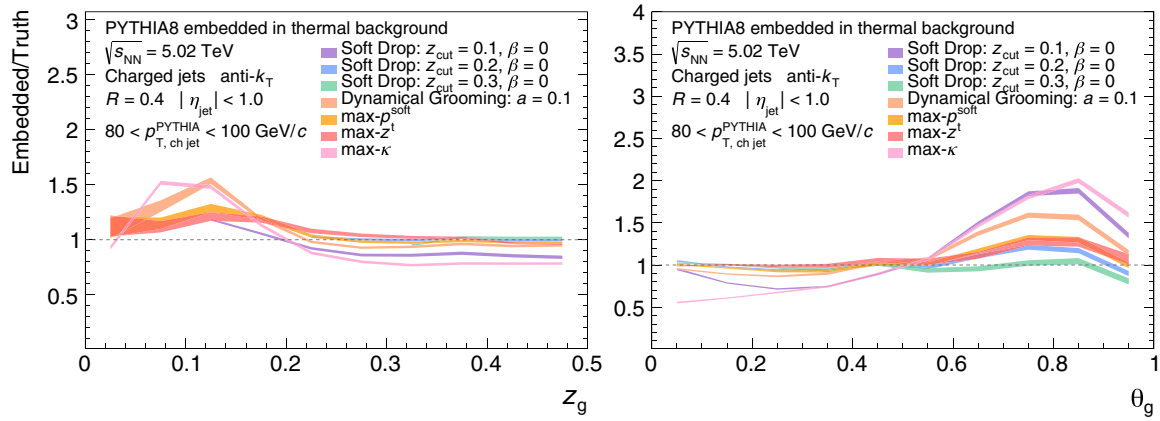


FIG. 6. Ratio of embedded $z_g$ (left) and $\theta_g$ (right) distributions to PYTHIA for a variety of groomers.

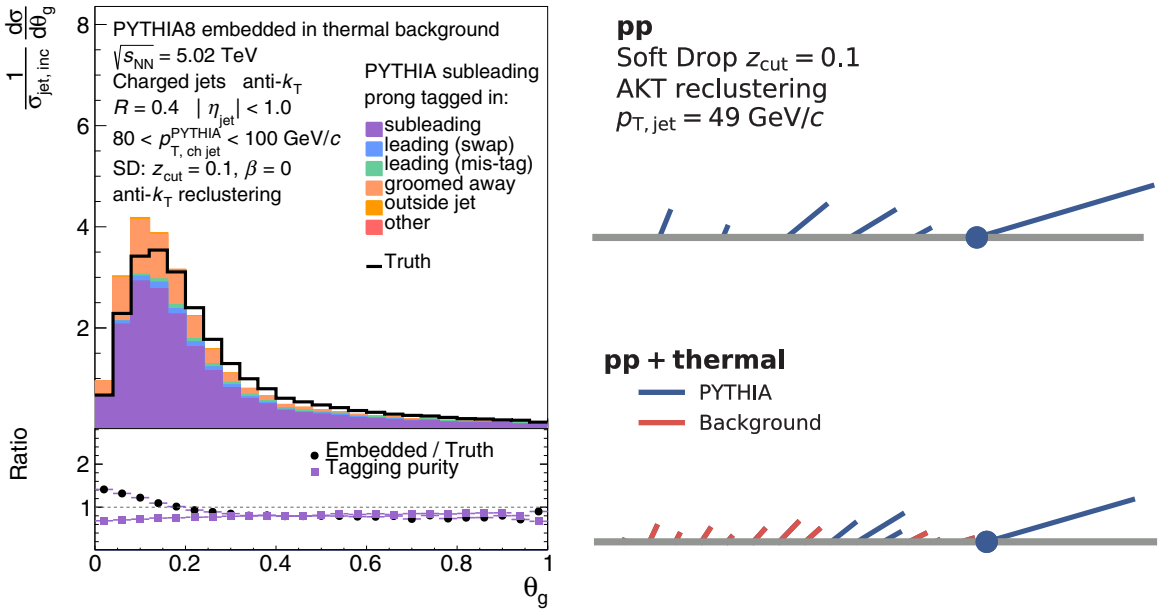


FIG. 7. (left) Distributions of $\theta_g$ when PYTHIA is embedded in the heavy-ion background, as well as from PYTHIA ("Truth"), for anti-$k_T$ reclustering. (right) Example of a PYTHIA jet (identical to Fig. 2) and the same jet embedded into thermal background for anti-$k_T$ reclustering. In the case of thermal background, anti-$k_T$ reclustering results in large-angle background particles being individually clustered to the leading branch, which results in them being groomed away.

Refs. [15,16], cuts on $R_g$ are employed, which are expected to induce suppression (or enhancement) of the remaining $z_g$ distribution in Pb-Pb relative to $pp$.[1] There are two relevant effects that the presence of mistagged splittings can have on such measurements.

First, mistagged splittings dilute quenching effects, which can change the shape of apparent modifications. When comparing Pb-Pb data to an embedded reference, mistagged subleading prongs are not expected to exhibit jet quenching since they arise from the combinatorial background. Since the tagging purity varies with $z_g$, this means that nontrivial changes to the shape of the Pb-Pb/$pp$ ratio can be induced. In particular, the tagging purity is low at small values of $z_g$ and high at large values of $z_g$. To illustrate the impact of this, consider a simple toy example for kinematics similar to the ALICE measurement with $\Delta R > 0.2$, as shown in Fig. 8 (left). Suppose that the true $R_{AA}$ induced by the $R_g$ cut is 0.5, independent of $z_g$. If we assume that mistagged splittings are unaffected by jet quenching, then the observed AA distribution will be given by

$$P_{AA}(z_g) = f_{\text{matched}} R_{AA} P_{pp}(z_g) + (1 - f_{\text{matched}}) P_{pp}(z_g),$$

where $f_{\text{matched}}$ is the tagging purity. Note that, as $f_{\text{matched}} \to 1$, $P_{AA}(z_g) \to R_{AA} P_{pp}(z_g)$, whereas if $f_{\text{matched}} \to 0$, $P_{AA}(z_g) \to P_{pp}(z_g)$. Since the tagging purity is low at small $z_g$ and high at large $z_g$, this generically causes the observed $R_{AA}$ to exhibit an apparent relative suppression of symmetric splittings—due entirely to background effects and unrelated to jet quenching. We note that the exact shape of the apparent relative suppression is model-dependent; there are many model-dependent choices one could make which we do not pursue further here;[2] however, the feature that the measured $R_{AA}$ will exhibit a spurious relative suppression emerges generically, independent of the details of jet quenching and depending only on the fact that the purity is low at small $z_g$ and high at large $z_g$. Based on these considerations it is difficult to conclude that symmetric splittings are more suppressed than asymmetric splittings using the ALICE measurement alone. The right panel of Fig. 8 shows a similar toy example corresponding approximately to CMS kinematics, which suggests that dilution effects are substantially smaller due to the higher purity at high $p_T$, but may still be significant. Note that if one fully corrects the distributions via unfolding instead of performing detector-level embedding comparisons, one eliminates the susceptibility to dilution effects, since the response matrix encodes appropriate corrections of any residual mistagged splittings to their true splittings.

[1]The measurements are normalized differently: In the case of the CMS Collaboration, any suppression due to the $R_g$ cut is self-normalized away, whereas in the case of the ALICE Collaboration, any suppression due to the $R_g$ cut persists in the $z_g$ distribution.

[2](A) The shape of the true $R_{AA}$ could be different—it could, for example, even show enhanced suppression of asymmetric splittings. (B) The mistagged splittings may exhibit a nontrivial correlation with $R_g$ and/or be affected by quenching. Consider the case of the true subleading prong being absorbed into the true leading prong, due to a large-angle local background fluctuation becoming the subleading prong. We then have $z_{\text{fake}} = p_{t,\text{bkgd}}/(p_{t,\text{bkgd}} + p_{t,\text{lead,true}} + p_{t,\text{sub,true}})$. In AA, the true prongs undergo energy loss, which may shift the $z_g$ distribution towards larger values relative to $pp$ embedded in a background. (C) The purity depends on both the model of the background and the jet.

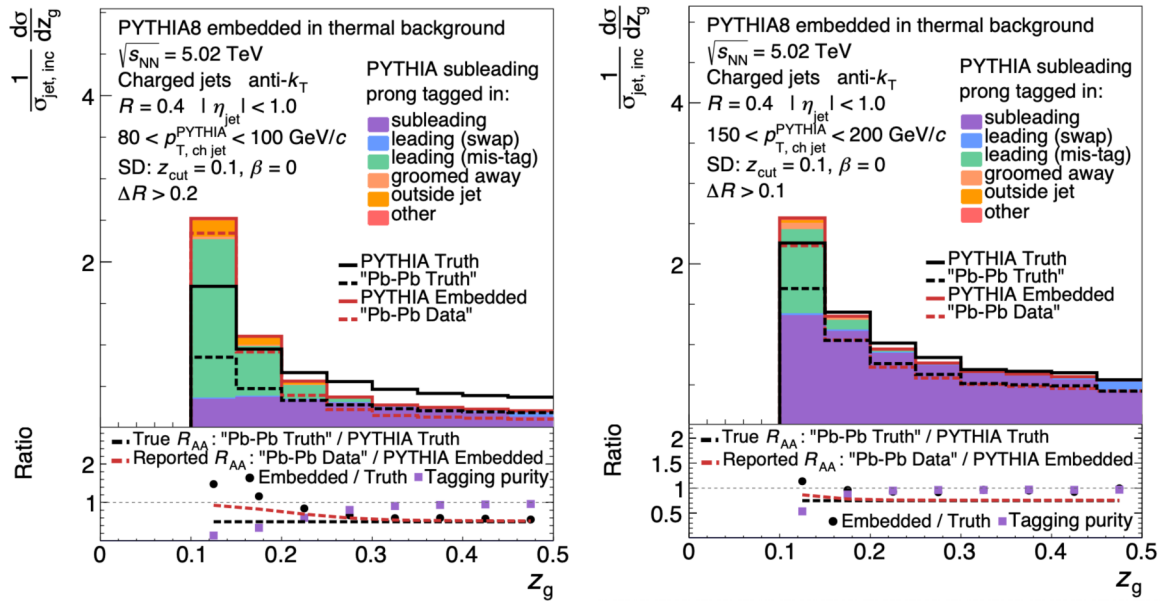


FIG. 8. Simple model showing that the presence of mistagged splittings can induce an artificial shape in the $z_g$ ratio, unrelated to jet quenching. Here, the normalization due to the $R_g$ selection (denoted by $\Delta R$) are taken in both numerator and denominator to be from the PYTHIA distribution in order to remove smearing effects (but keep the suppression quantified by $R_{AA}$). Note that the momentum scale here is taken from PYTHIA, whereas the experimental selection is a partially uncorrected Pb-Pb scale.

Second, the magnitude of MC-based corrections (relevant to Ref. [16]) grows as the number of mistagged splittings grows. In Fig. 8 (left), the ratio "Embedded/Truth" gives an estimate of the size of MC-based corrections one has to perform to compare Pb-Pb data to an embedded reference, and is on the order of 100%. Note that the shape of this correction is correlated with the experimentally observed modification. Moreover, the distributions are effectively self-normalized, aside from the suppression induced by the $R_g$ cut—meaning that small-$z_g$ modification necessarily causes large-$z_g$ modification.

## V. CONCLUSION

We performed a set of basic studies on the behavior of various jet grooming algorithms in the presence of the large combinatorial background characteristic of heavy-ion collisions. We found that such background and its region-to-region density fluctuations cause a significant number of splittings to be incorrectly identified as a genuine structure of signal jets. The robustness of groomers against this experimental challenge is an important criteria for their usage in jet substructure measurements in heavy-ion collisions. We quantified the performance of grooming algorithms by using the purity of the identified splittings differentially in both the jet momentum and individual substructure observables. Our studies show that subleading prongs are prone to misidentification (lost, replaced by a background flux of particles, and thus often merged into the leading prong) and that, in general, the contamination decreases (the groomer performance improves) with increased $p_T$ of the jets. We identified a set of grooming algorithms that perform relatively well; however, in our test setup, we found that groomers used in some of the existing heavy-ion measurements result in a significant contamination of the reported distributions with false splittings. Since these background-induced splits can generically mimic jet quenching effects, future measurements at the LHC and RHIC will need to leverage the grooming algorithms that maximize the purity of the genuine splittings. One of the important challenges will be to properly quantify the uncertainties in the reported quantities due to residual contamination effects.

The studies presented here ought to be extended to further explore the model dependence of the background and the impact of jet fragmentation on the performance of grooming algorithms. Moreover, similar purity studies should be extended to any observable where a substructure object must be tagged jet by jet, such as reclustered subjets. Alternate experimental approaches, such as ensemble-based background subtraction of mistagged splittings, should also be explored. The groomers that we have considered can be refined and expanded—a promising direction to explore may be to combine a high-purity groomer with an additional phase-space selection (e.g., $\kappa$, $t_f$). Investigation of alternate reclustering algorithms and iterative grooming techniques may also be promising, as ultimately one wants to optimize a combination of the reclustering algorithm and grooming condition to construct splittings that are both robust to mistagging and sensitive to relevant physics (and calculable). This, of course, calls for further theoretical guidance.

## ACKNOWLEDGMENTS

We thank Peter Jacobs, Marco van Leeuwen, Felix Ringer, and Nima Zardoshti for helpful discussions. This work was supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under the contract DE-AC02-05CH11231.

[1] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, Soft drop, J. High Energy Phys. 05 (2014) 146.

[2] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, Towards an understanding of jet substructure, J. High Energy Phys. 09 (2013) 029.

[3] A. J. Larkoski, S. Marzani, and J. Thaler, Sudakov safety in perturbative QCD, Phys. Rev. D **91**, 111501 (2015).

[4] Y.-T. Chien and I. Vitev, Probing the Hardest Branching Within Jets in Heavy-Ion Collisions, Phys. Rev. Lett. **119**, 112301 (2017).

[5] Y. Mehtar-Tani and K. Tywoniuk, Groomed jets in heavy-ion collisions: Sensitivity to medium-induced bremsstrahlung, J. High Energy Phys. 04 (2017) 125.

[6] N.-B. Chang, S. Cao, and G.-Y. Qin, Probing medium-induced jet splitting and energy loss in heavy-ion collisions, Phys. Lett. B **781**, 423 (2018).

[7] G. Milhano, U. A. Wiedemann, and K. C. Zapp, Sensitivity of jet substructure to jet-induced medium response, Phys. Lett. B **779**, 409 (2018).

[8] R. Kunnawalkam Elayavalli and K. C. Zapp, Medium response in jewel and its impact on jet shape observables in heavy ion collisions, J. High Energy Phys. 07 (2017) 141.

[9] P. Caucal, E. Iancu, and G. Soyez, Deciphering the $z_g$ distribution in ultrarelativistic heavy ion collisions, J. High Energy Phys. 10 (2019) 273.

[10] J. Casalderrey-Solana, G. Milhano, D. Pablos, and K. Rajagopal, Modification of jet substructure in heavy ion collisions as a probe of the resolution length of quark-gluon plasma, J. High Energy Phys. 01 (2020) 44.

[11] F. Ringer, B.-W. Xiao, and F. Yuan, Can we observe jet $P_T$-broadening in heavy-ion collisions at the LHC? Phys. Lett. B **808**, 135634 (2020).

[12] H. A. Andrews *et al.*, Novel tools and observables for jet physics in heavy-ion collisions, J. Phys. G **47**, 065102 (2020).

[13] G. Aad *et al.* (ATLAS Collaboration), Measurement of soft-drop jet observables in $pp$ collisions with the atlas detector at $\sqrt{s} = 13$ TeV, Phys. Rev. D **101**, 052007 (2020).

[14] J. Adam *et al.* (STAR Collaboration), Measurement of groomed jet substructure observables in $pp$ collisions at $\sqrt{s} = 200$ GeV with STAR, arXiv:2003.02114.

[15] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of the Splitting Function in $pp$ and Pb-Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV, Phys. Rev. Lett. **120**, 142302 (2018).

[16] S. Acharya *et al.* (ALICE Collaboration), Exploration of jet substructure using iterative declustering in $pp$ and Pb-Pb collisions at LHC energies, Phys. Lett. B **802**, 135227 (2020).

[17] K. Kauder, Measurement of the shared momentum fraction $z_g$ using jet reconstruction in $p + p$ and Au + Au collisions with STAR, Nucl. Phys. A **967**, 516 (2017).

[18] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, An introduction to PYTHIA 8.2, Comput. Phys. Commun. **191**, 159 (2015).

[19] M. Cacciari, G. Salam, and G. Soyez, The anti-$k_T$ jet cluster algorithm, J. High Energy Phys. 04 (2008) 063.

[20] B. Abelev *et al.* (ALICE Collaboration), Measurement of event background fluctuations for charged particle jet reconstruction in Pb-Pb collisions at $\sqrt{s} = 2.76$ TeV, J. High Energy Phys. 03 (2012) 053.

[21] P. Berta, L. Masetti, D. W. Miller, and M. Spousta, Pileup and underlying event mitigation with iterative constituent subtraction, J. High Energy Phys. 08 (2019) 175.

[22] Y. Mehtar-Tani, A. Soto-Ontoso, and K. Tywoniuk, Dynamical grooming of QCD jets, Phys. Rev. D **101**, 034004 (2020).

[23] Y. Mehtar-Tani, A. Soto-Ontoso, and K. Tywoniuk, Tagging boosted hadronic objects with dynamical grooming, arXiv:2005.07584.

[24] F. A. Dreyer, G. P. Salam, and G. Soyez, The Lund jet plane, J. High Energy Phys. 12 (2018) 064.

[25] C. Bierlich, G. Gustafson, L. Lönnblad, and H. Shah, The Angantyr model for heavy-ion collisions in PYTHIA8, J. High Energy Phys. 10 (2018) 134.

[26] C. Frye, A. J. Larkoski, M. D. Schwartz, and K. Yan, Factorization for groomed jet substructure beyond the next-to-leading logarithm, J. High Energy Phys. 07 (2016) 64.