Identifying heavy particle jets with a new nonlinear jet observable

Ankita Budhraja[*] and Ambar Jain[†]

Indian Institute of Science Education and Research, Bhopal MP 462066, India

We introduce a new jet observable *zest* defined on exclusively constructed jets and study its potential to discriminate jets originated from Standard Model heavy particles like W, Z bosons and top quark from gluon initiated jets. Zest exhibits properties such as boost invariance, stability against global color exchange among partons, and inclusion or exclusion of a few soft particles in the jet. We also observe that for gluon jets, zest distribution is mostly insensitive to the jet mass. These properties make zest a suitable candidate for vetoing gluon jets at the colliders. Zest, when used in conjunction with other substructure observables that are uncorrelated to it, can further improve gluon jet veto. We generalize zest and show that in one limit it is synonymous to particle multiplicity and in the other limit, it projects only the leading particle. Optimization on the parameter of generalized zest further improves the discrimination ability of the observable. We find that for the top quark-initiated jets, the discrimination provided by generalized zest is in close comparison with a class of machine learning-based top taggers. We propose that studying other nonlinear infrared and collinear unsafe observables may help in unveiling the hidden physics of machine learning-based observables.

I. INTRODUCTION

New physics searches are often accomplished by looking for excess in cross sections pertaining to specific decay channels of new physics particles. Constraining the measured cross section to specific decay channels can often reduce the background significantly enough to identify the excess with high confidence level. Thus, it is usually cleaner to search for new physics particles in the leptonic or photonic channels as they offer very small background at the Large Hadron Collider (LHC). However, the cleaner channels often have extremely small cross section that it makes the searches for new physics difficult or requires a large amount of data to get a significant sample of the signal events. Additionally, leptonic searches can be difficult if the physics or decay channels of the new particle are not known. Further difficulties may arise if there is a missing transverse energy in a leptonic channel. Similar difficulties also arise for the Standard Model (SM) heavy particles, namely, W, Z bosons, Higgs boson and top quarks, except that the physics of these particles is well known. These heavy particles dominantly decay to light quarks and gluons (collectively partons) which

then shower and fragment into jets of hadrons. Since jets are invariably produced at the LHC due to hard scattering of the constituent partons, heavy SM particles are often faked by light quark/gluon radiations and radiations coming from other sources, such as initial state radiation, underlying event [1,2] and pile-up [3]. At present collider energies, these heavy SM particles are often produced with a large Lorentz boost factor and their hadronic decay products are realized as a single collimated "fat jet" [4–7]. In such cases, the mass of the heavy particle should be reflected in the mass of the fat jet. One may expect to disentangle the signal from the large background of light parton jets, using the jet mass cut,[1] but as it happens to be the case, gluon jet cross section has a significant long tail even after accounting for underlying event [1,2] and pile-up [3], which can be dealt with effectively using the jet grooming techniques, such as soft drop [8–10], pruning [11,12], trimming [13], and mass drop/filtering [14–16]. Further discussion on underlying event and pile-up is beyond the scope of this work. As far as this paper is concerned, we will assume that we start with a well-groomed jet. Since a simple mass cut is not effective in identifying the particle originating the jet in spite of grooming, thus a great deal of effort has been made in jet identification by constructing observables that can reject the parton jets and reduce the background [4–7,17,18]. The emphasis of such observables is to improve the signal rate to mistag rate ratio. This will also be the spirit of our work.

[*]ankitab@iiserb.ac.in
[†]ambarj@iiserb.ac.in

---

[1]In this work, we mainly focus on gluon jets as the background since their cross section is much larger.

The SM heavy particles provide a perfect playground for improving on such jet identification strategies as their decay channels are well understood. The purpose of our work is to improve upon these strategies by studying a new nonlinear jet observable *zest* dependent only on the transverse momentum distribution of particles in the jet, similar to transverse-zeal introduced in the context of jet quenching studies [19].

The remainder of the paper is organized as follows. In Sec. II, we define zest and discuss its properties. In Sec. III, we discuss the details of our simulations. In Sec. IV, we outline the features exhibited by zest distribution of gluon-initiated jets that make it suitable for vetoing the gluon jets. We provide the principal results of our zest-based analysis in Sec. V. In Sec. VI, we introduce another substructure observable *bib* which also depends only on the transverse momentum but linearly and is largely uncorrelated to zest. We contrast the two observables and perform a bivariate analysis to improve the discrimination ability in Sec. VII. In Sec. VIII, we generalize zest by introducing a real parameter p that can be tuned to modify the contribution of the soft particles in the jet. The parameter p can be optimized to provide improved discrimination for the heavy SM particle jets from the gluon background, as discussed in Sec. VIII A and Sec. VIII B. We show that for the top quark-initiated jets, the discrimination provided by generalized zest is in close comparison to recent machine learning (ML) approaches. We propose that studying such infrared and collinear (IRC) unsafe observables may help to uncover the hidden physics behind ML approaches. Since, zest is collinear unsafe and is computable for hadronic final states only, we study its dependence on different hadronization models in Sec. IX. We conclude in Sec. X.

II. ZEST OF A JET

Given a well-groomed jet composed of hadrons, reconstructed through a suitable jet algorithm like the anti-k_t clustering algorithm [20], *zest* of the jet is defined as

$$\zeta = \frac{-1}{\log(\sum_{i \in \text{Jet}} e^{-P_T/|\mathbf{p}_{Ti}|})}, \qquad (1)$$

where $P_T = \sum_{i \in \text{Jet}} |\mathbf{p}_{Ti}|$ and $\mathbf{p}_{Ti}$ is the transverse momentum of the ith particle in the jet with respect to the jet axis. Note that zest is composed purely out of the transverse momenta of the final state particles. For two leading particles in the jet carrying equal transverse momentum of p_T, i.e., $P_T = 2p_T$, we get

$$\zeta = \frac{-1}{\log(2) - 2} \approx 0.765. \qquad (2)$$

This value remains roughly the same even if the two particles do not carry equal fractions of P_T, i.e., for a generic break up of $P_T = xP_T + (1-x)P_T$ with $0 < x < 1$. This is confirmed also by looking at Fig. 1 where we have plotted the normalized distribution of 10^5
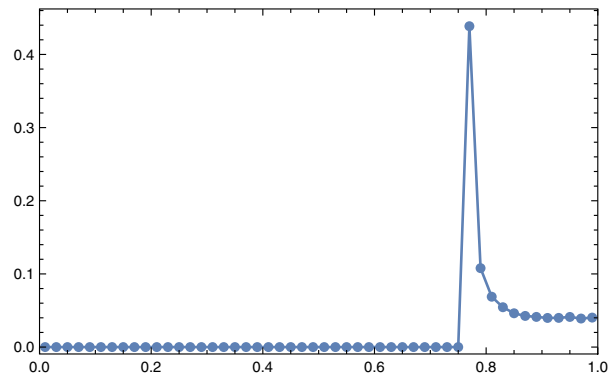

FIG. 1. Zest distribution for two leading particles in a jet with a fixed P_T. We see a dominant two-particle peak at a zest value of 0.76.

randomly generated pairs of particles with total $P_T = 1$. From the plot, we see that the two particle distribution has a sharp peak at about ~0.77 and a flat tail up to $\zeta = 1$ which happens when one of the particles has a small contribution to P_T. A similar analysis done for only three particles in a jet gives a slightly broader multiplicity peak at $\zeta \sim 0.53$. Therefore, we see that the zest distribution captures multiplicity peaks whenever there are a few energetic particles in the jet. These multiplicity peaks shift toward smaller and smaller value as the number of particles contributing to the jet increases. In the special case of a jet containing n particles each carrying equal $|\mathbf{p}_{Ti}|$, ζ reduces to $\frac{1}{n - \log n}$. Thus we expect zest to be strongly correlated to *multiplicity of a jet*. Later we will generalize zest such that multiplicity will emerge as a limiting case. For a jet containing a single particle, Eq. (1) leads to indeterminate result, thus for consistency, we define $\zeta = 1$.[2]

Zest has certain interesting properties which makes it a useful observable to study jet substructure: (1) It is invariant under boosts made along the jet axis as it is composed entirely of transverse momentum components of the jet constituents, (2) It is sensitive to transverse momentum distribution of energetic particles while it is largely insensitive to the soft particles in the jet since their contribution to the observable is exponentially suppressed, and (3) Zest is also mostly insensitive to the global color flow of partons. In other words, the zest distribution of the gluon-initiated jets (or any other colored particle) is stable against the change of color flow direction of the colored particles forming the jet. A colored parton that evolves into a color neutral hadron jet recoils against another colored parton. The color neutral final state is obtained due to the exchange of several soft partons. Changing the direction of the recoiling colored partner effects only the exchange of a few soft particles between the partners. Since zest is stable

---

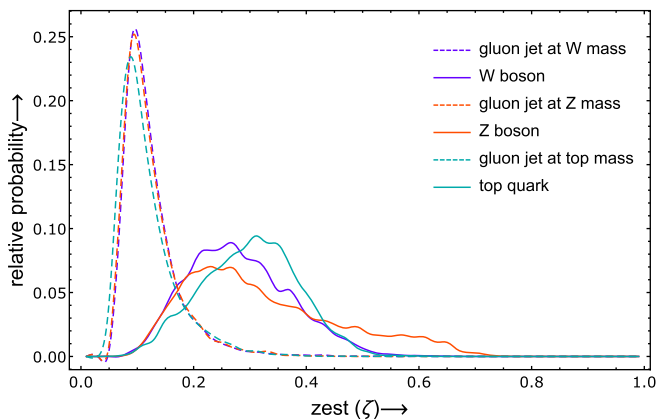[2]We thank the referee for pointing the subtlety associated with the single particle jet.

FIG. 2. Zest distribution for W, Z boson and top quark initiated jets along with gluon jets of corresponding masses.

against a few changes in the soft sector, it also becomes stable to the color recoil.

While zest is a nonlinear and collinear unsafe jet observable due to which it cannot be calculated in perturbation theory, it can be computed using Monte Carlo based event generators [21,22]. Although collinear unsafe, zest may provide a new perspective into the jet substructure that may not be accessible through IRC safe jet observables. In particular, the jet observables based on machine learning techniques, while not calculable through perturbation theory, provide the best discrimination ability [23–29]. The physical features extracted by such observables that enable the discrimination remains unknown. We anticipate that studying zest and other IRC unsafe observables may shed some light on the black box offered by neural-network-based observables. It goes without saying that zest is hadronization model dependent, however we expect that its ability to discriminate between the originating particles will be independent of hadronization models used in a Monte Carlo simulator, as is demonstrated later in Sec. IX.

In Eq. (1), we have defined zest using transverse momenta of particles with respect to the jet axis. It is possible to define alternate observables of similar type by replacing p_{Ti} with other kinematic quantities such as energy of the particle or longitudinal momentum of the particle. The reasons for not considering these alternates in this paper are (1) Zest definition made with transverse momenta is invariant under boosts along the jet axis thus zest can only mildly dependent on the energy of the jet which would not hold for the alternates, and (2) Although the results for alternate possibilities are not presented in this paper, we have verified that zest serves as a significantly better candidate for vetoing the gluon-initiated jets compared to its alternates.

III. SIMULATION DETAILS

The details of our Monte Carlo simulations are presented below.

(1) All jets are simulated using PYTHIA 8 [21] by inserting W and Z bosons with energy 500 GeV along the central axis (y-axis) and allowing them to decay to hadronic modes only.[3]

(2) We collect all the particles in the forward hemisphere and call it a jet. Apart from hemisphere jets, sometimes we also construct anti-k_t jets with different jet radii, $R = \{0.5, 0.7, 0.9, 1.2\}$, using the FASTJET package [30] in PYTHIA.

(3) For simulating a color-singlet top quark jet, we perform an e^+e^- annihilation and allow the intermediate Z boson to decay to a top-antitop pair at 500 GeV each, traveling back-to-back. We further constrain the top decay channels to hadrons only. Jets are constructed by dividing the event into two hemispheres using the thrust axis.

(4) To shower a color singlet gluon-initiated jet, we insert an energetic gluon of energy 500 GeV and offshellness same as the heavy particle mass in the event record along the y-axis, while a soft onshell gluon of energy 1 MeV with opposite color is inserted in the opposite direction to the $+y$-axis. The direction of the color conserving gluon partner of the hard gluon does not effect the analysis, as zest is stable to global color flow of the partons (see Fig. 4). For convenience, we choose the direction of the recoiling gluon opposite to that of hard gluon.

(5) The energetic gluon jet is identified by constructing the thrust axis and then taking all particles in the forward hemisphere.

(6) The resulting gluon jets are accepted if their masses are within 10 GeV of the corresponding heavy particle mass. We observe that the selected gluons jets have energy ranging between 485 GeV to 500 GeV.

(7) In order to test the sensitivity of our observable to nonperturbative physics, we also generate jets in two distinct ways: (a) by varying the default parameters of the hadronization model implemented in PYTHIA, thereby producing a new set of final state events, and (b) by utilizing a different model for nonperturbative hadronization effects by using the HERWIG [22] event generator. We study this nonperturbative physics dependence for Z boson-initiated jet along with corresponding gluon background in Sec. IX.

(8) The HERWIG event is sampled using the process $e^+e^- \to Z \to$ jets at a center of mass energy equal to the mass of the Z boson, i.e., $E_{\mathrm{cm}} = m_Z$. We then

---

[3]For the study presented here, we have turned off all the strong, weak and electromagnetic decays of the primary hadrons formed from showering and subsequent hadronization. Since zest is a hadronic observable and is correlated to multiplicity, allowing further decays of hadrons will change the definition of the observable, thus for simplicity, we do not study the hadron decays and detector analysis in this paper.
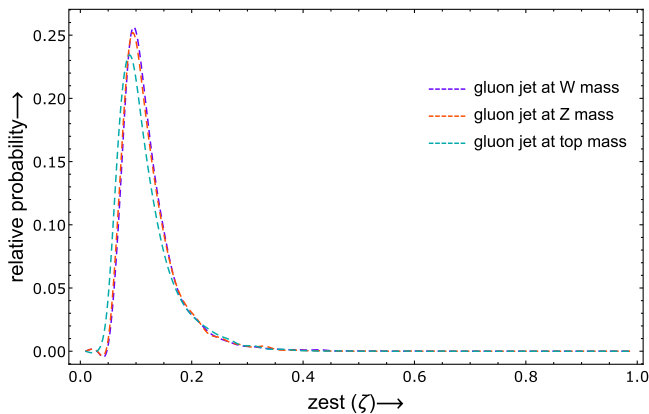
FIG. 3. Zest distribution of gluon-initiated jets with a jet mass around W boson mass, Z boson mass and top quark mass.

boost all the final state particles along a direction, which we take as the +y-axis. The resulting boosted jet has a mass about the Z boson and energy around 500 GeV.

(9) In this work, we will not study Higgs-initiated jets because a Higgs boson dominantly decays to a $b\bar{b}$ pair which receives a large background from $Z \rightarrow b\bar{b}$ and $g \rightarrow b\bar{b}$, in which case the analysis proposed here does not offer a substantially good discrimination ability.

We have avoided generating these jets from proton-proton collisions, so that we can study zest in a more controlled environment, since some of the properties of zest are strikingly different from other well-studied IRC safe observables. For instance, in Sec. IV, we study the zest distribution of gluon-initiated jets by changing the energy and direction of the color-complementary gluon, which will not be possible in a proton-proton environment. In addition, the event simulated from proton-proton collisions will necessarily require a suitable jet algorithm and jet grooming technique to provide a well-defined jet of hadrons, a complication that we want to avoid in this first study. Nevertheless, we will show in Secs. IV and V that zest is stable against the change in jet radius and inclusion/exclusion of a few soft particles to the jet. Thus, for simplicity, the complications of the complete proton-proton collisions are kept out of this paper.

IV. ZEST DISTRIBUTION OF GLUON-INITIATED JETS

As gluon jets, on an average, consist of a larger number of particles, the zest distribution of gluon-initiated jets is expected to peak at smaller values of the observable. We validate this in Fig. 3 by plotting zest distribution for jets initiated by offshell gluons of masses around W, Z boson and top quark masses. We see that the zest distribution for gluon jets peaks at small $\zeta \sim 0.1$, as expected. Interestingly, we also see that these zest distributions are nearly independent of the jet mass. To investigate this, we have shown the ratio of the mean of n maximum p_{Ti} to P_T for $n = 1, 2, 5$ and 10 for various jet masses in Table I. We find that these ratios are nearly independent of jet mass, which explains why the zest distribution for gluons is insensitive to the jet mass.

We also verify color flow independence of zest, discussed in the previous section, by looking at the zest-distribution of gluon jets recoiling against complementary color gluon at various angles. The result is shown in Figs. 4(a) and 4(b). In Fig. 4 we have considered jets of jet energy about 300 GeV, for two different cases: (i) a hard gluon of energy 300 GeV recoiling a soft gluon of energy 1 MeV at 180°, and (ii) a hard gluon of energy 300 GeV recoiling a hard gluon of energy 200 GeV at 180°. From the plot in Fig. 4(a), we note that the choice of the recoiling partner energy does not strongly effect the zest distribution. Furthermore, from Fig. 4(b), we observe the color stability of gluon-initiated jets against the direction of its recoiling color neutralising partner. Since, the energy as well as the direction of the recoiling colored particle does not largely effect the analysis, for the rest of the paper, we will take the recoiling gluon as soft and back-to-back to the primary hard gluon.

As zest is largely insensitive to soft particles in the jet, its distribution curve is not effected by including or excluding a few soft particles in the jet. We verify this in Fig. 5 where we have plotted the zest distribution curves for gluon-initiated jets of mass around the top quark mass with following three variations: (i) including all the particles in the jet, (ii) excluding all the soft particles less than 3 GeV from the jet (iii) including few extra randomly generated soft particles less than 3 GeV in the jet. We see that the zest distribution shifts only very slightly.

The stability of the gluon zest distribution against jet mass variation, color flow direction, and inclusion/exclusion of

TABLE I. The entries in the table give the ratio of mean of n maximum $|p_{Ti}|$ to P_T of the jet, that is $\frac{\langle max_n|p_{Ti}|\rangle}{P_T}$. The first set of numbers specify the mean value and the second is the standard deviation calculated over a set of 10000 jets with the specified jet mass.

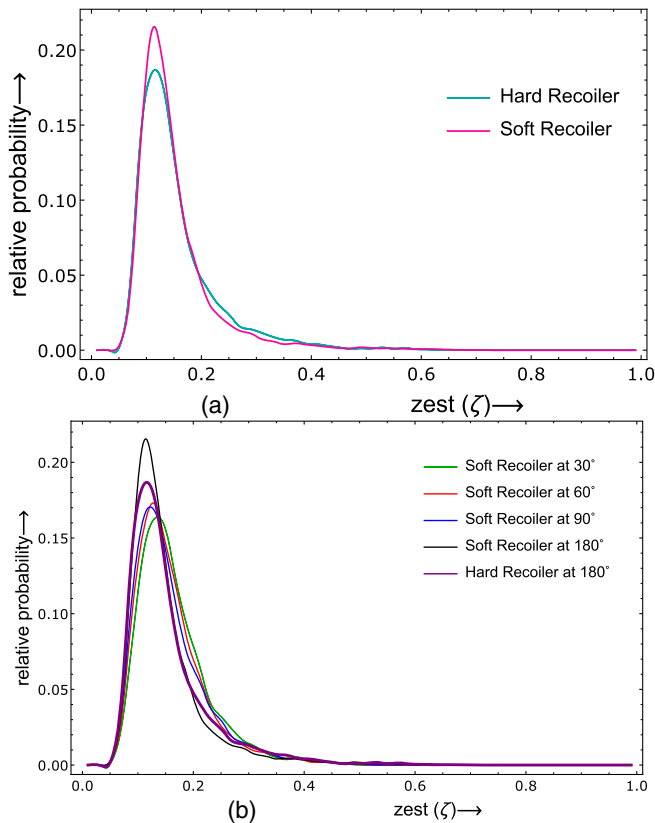| m_J	n			
	1	2	5	10
Gluons around W mass	0.121 ± 0.049	0.105 ± 0.035	0.081 ± 0.018	0.0617 ± 0.0091
Gluons around Z mass	0.120 ± 0.048	0.104 ± 0.035	0.079 ± 0.018	0.0603 ± 0.0094
Gluons around top mass	0.116 ± 0.051	0.099 ± 0.037	0.074 ± 0.019	0.0552 ± 0.0102

FIG. 4. (a) Zest distribution of gluon-initiated jets with a hard gluon of energy about 300 GeV when the recoiling gluon is soft (blue) and hard (pink). (b) Zest distribution of gluon-initiated jets of energy about 300 GeV when the soft recoiler is taken at 30° (green), 60° (red), 90° (light blue), 180° (black) vs a hard recoiler at 180° (purple).

few soft particles makes it a potentially useful observable for vetoing gluon jets at the LHC. For instance, a zest cut at about $\zeta \sim 0.3$ is expected to largely cut out the gluon background to heavy SM particles and hopefully also for heavier new physics resonances.
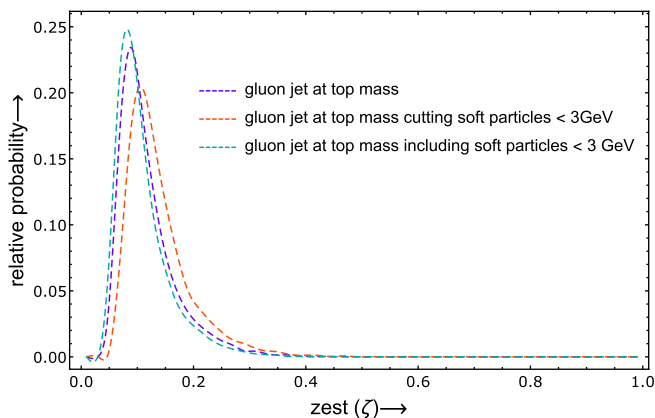

FIG. 5. Zest distribution of gluon-initiated jets around top quark mass by including or excluding a few soft particles < 3 GeV to the jet.

V. ZEST AS FILTER

As noted earlier, the zest distribution of gluon-initiated jets peaks at small values of the observable in comparison to their heavy particle counterparts. This accounts for an appreciable distinguishing criterion offered by zest in characterization of heavy particle jets from the background gluon jets. We analyze this discriminating ability provided by zest by looking at the ζ-distribution curves for W, Z boson and top quark-initiated jets along with the corresponding gluon-initiated jets with a mass about the heavy particle masses. This comparison is shown in Fig. 2. From the plot, we see that the zest distribution of the gluon-initiated jets peaks at relatively smaller values of ζ compared to their heavy particle counterparts. In addition, the zest distribution of gluon-initiated jets is found to be narrow, independent of jet energy and has a little overlap with a similar distribution of most other heavy particle jets. Moreover, we find that for the top quark-initiated jet, the zest distribution peaks at a slightly higher value than W or Z boson-initiated jets.

The tagging efficiency of a zest-based binary classifier can be further evaluated by comparing the performance curves of these jets. In Fig. 6, we present the relative operating characteristic (ROC) curves for the zest-based filter. We note that zest provides good signal statistics along with a high background rejection rate. For example, we note that for all the heavy particle jets considered, nearly 80 to 90% of the signal stays in the accepted sample after applying a zest cut to remove 90% of the gluons.

In experiments, jets are often constructed by a suitable jet algorithm employing a cut on the jet cone radius R. At large jet radius, the wide angle soft particles contribute to the jet, while as the jet radius decreases, the observable receives contribution from collinear particles in the jet and a few soft particles. Since zest is not significantly effected by the inclusion or exclusion of a few soft particles, we anticipate the zest distribution to be stable against change in the jet cone radius. This is confirmed by Fig. 7, where ζ-distribution
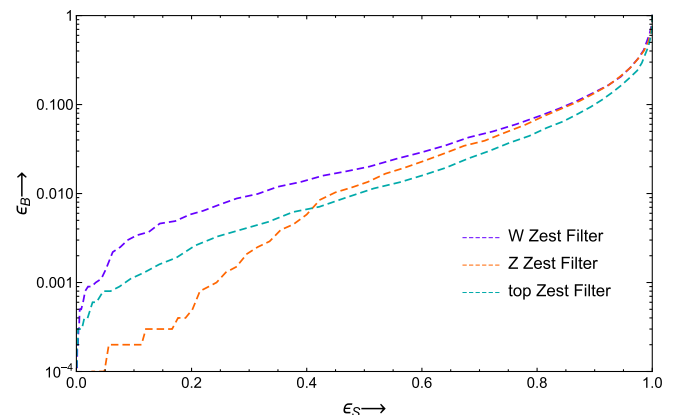

FIG. 6. ROC curves for W, Z boson and top quark-initiated jets using *zest* as the binary classifier.
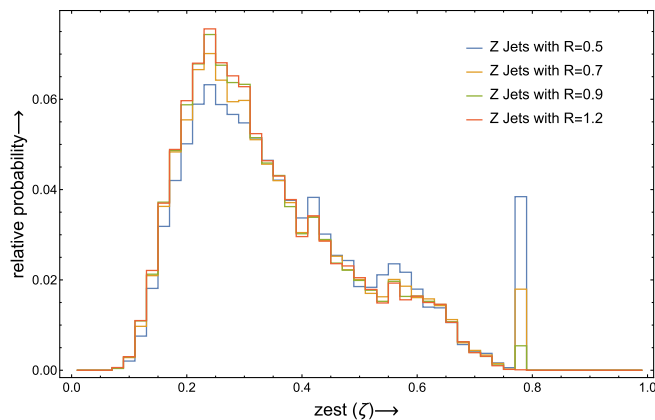
FIG. 7. Zest distribution for Z boson-initiated jets with different jet radii. Here jets are constructed using the anti-kT clustering algorithm for different R values.

curves with different jet radii, namely $R = \{0.5, 0.7, 1.0, 1.2\}$, are presented for Z-boson initiated jets. From the plot, we also note that the two particle peak at $\zeta \approx 0.77$ gets enhanced when the jet radius is sufficiently small, aligned with our expectation.

VI. BOOST-INVARIANT BROADENING AS FILTER

To contrast zest with an observable linearly dependent on transverse momentum $|\mathbf{p}_T|$ of the particles, we introduce a simple observable *boost-invariant broadening* or *bib* (for brevity) which is defined by the equation,

$$bib = \frac{1}{m_J} \sum_{i \epsilon Jet} |\mathbf{p}_{Ti}| = \frac{P_T}{m_J}. \qquad (3)$$

Here as well, all transverse momenta are measured with respect to the jet axis and P_T is the same quantity as defined in the context of zest. It is interesting to note that *bib* is similar to the event shape observable *broadening* [31], however the use of m_J instead of E_J renders it invariant under boosts made along the jet axis. Zest and *bib* have boost-invariance along the jet axis and jet energy independence in common with each other. Although there can be other observables which can be chosen for a bivariate analysis, *bib* provides a simple contrast to zest: (i) *bib* is a linear function of the transverse momentum of the particles while zest is not, (ii) *bib* is calculable using standard perturbative techniques,[4] while zest is collinear unsafe and relies on a suitable Monte Carlo generator that encodes a

---

[4]As *bib* is defined by the ratio of two IRC safe observables, it is essentially Sudakov safe [32,33] which implies that even though the observable does not has a valid fixed order expansion in α_s, it can still be calculated using all-order resummation. This ensures that the singular region of the phase space for the observable is exponentially suppressed.
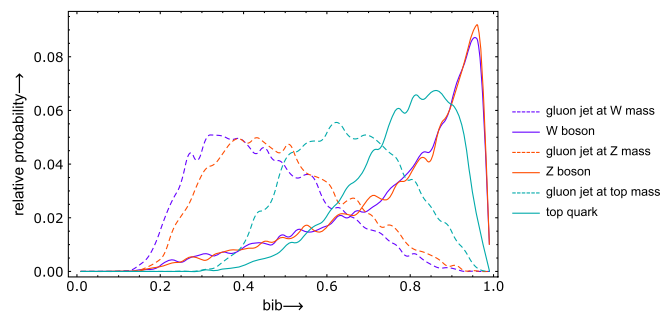

FIG. 8. *bib*-distribution of heavy particle initiated jets and gluon jets forming the background.

definite model to capture the hadronization effects for its computation, and (iii) *bib* is sensitive to the global color flow of partons while zest is insensitive to it. The results presented in this section and for the bivariate analysis in the next section, are strictly for the case in which the soft gluon is back-to-back with the primary hard gluon.

The physics of the primary splitting of heavy particles leading to partonic decays is different from that of branching of quarks and gluons. Parton splittings dictated by quantum chromodynamics (QCD) are dominated by soft and collinear radiation having small transverse momenta, thus we expect gluon initiated jets to have a peak at a relatively smaller value in *bib*-distribution compared to jets initiated by heavy particles at the same jet mass. The comparison of *bib*-distributions for heavy particles and corresponding gluon jets is shown in Fig. 8. We also present the ROC curves for *bib* when used as a binary classifier in Fig. 9. From the plot, we find that *bib* acts as an appreciable classifier for W, Z boson-initiated jets but does not offer good discrimination for heavier top quark-initiated jets. This is in contrast with the zest ROC curves of Fig. 6. Overall, zest provided better classifier performance compared to *bib*. Nonetheless, *bib* also contrasts zest in the stability of gluon distributions with respect to the jet mass. We have also verified that the two observables do not have a
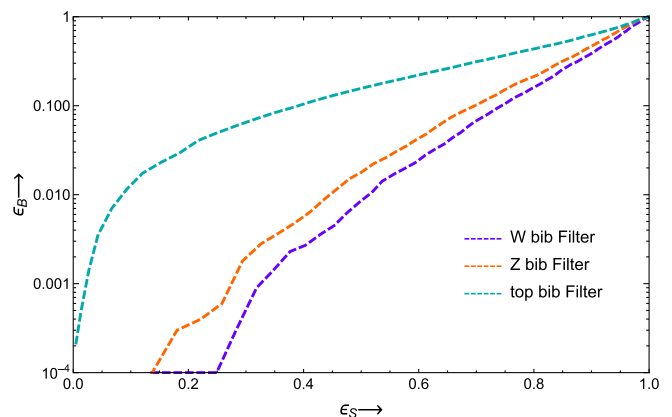

FIG. 9. ROC curves for heavy particle jets with *bib* as the classifier.

significant linear correlation. Thus, we expect that the two observables in conjunction can be used to improve the signal to background ratio. We present this bivariate analysis in the next section.

VII. BIVARIATE ANALYSIS

Discrimination for the originating particle can be further improved with a multivariate analysis. Recently, multivariate approaches are becoming extremely popular as they are expected to provide a more detailed characterization of the QCD radiation pattern within a jet which can be exploited to further enhance the new physics searches at the LHC [17,18]. For the bivariate analysis proposed here, we generate scatter plots in the *bib*-zest plane for gluon initiated jets and heavy particle jets, as shown in Figs. 10(a)–(c). We note that, in general, gluon jets occupy the extreme left corner of the *bib*-zest plane, while heavy particle jets are mostly concentrated toward the upper right region in the plot, thus providing statistical discrimination between the two types of jets.

The zest-*bib* distributions shown in Fig. 10 are reasonably well-approximated by an asymmetric two-dimensional Gaussian distribution. We fit both the heavy particle and gluon jet distributions with an asymmetric Gaussian distribution. On the fitted Gaussian distributions, we draw contours of different heights. An example for the 90%
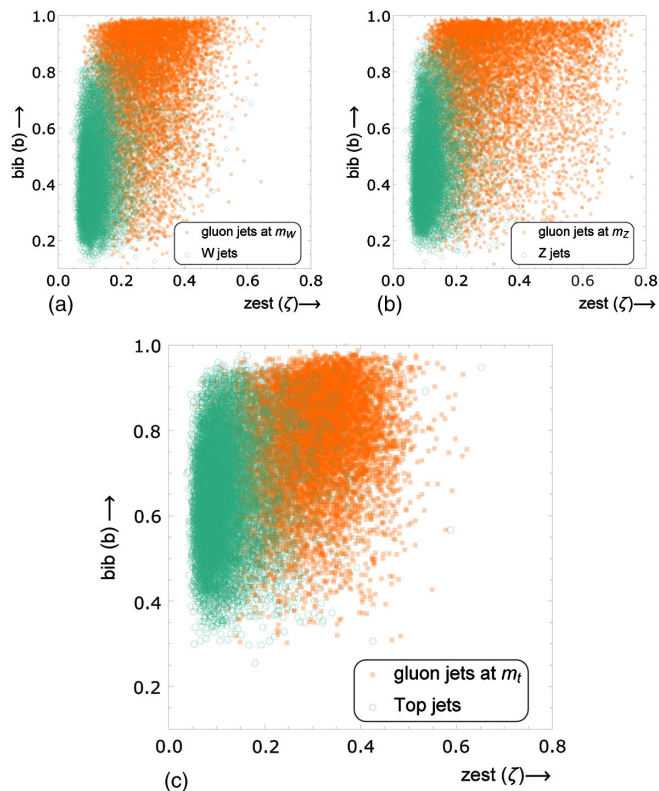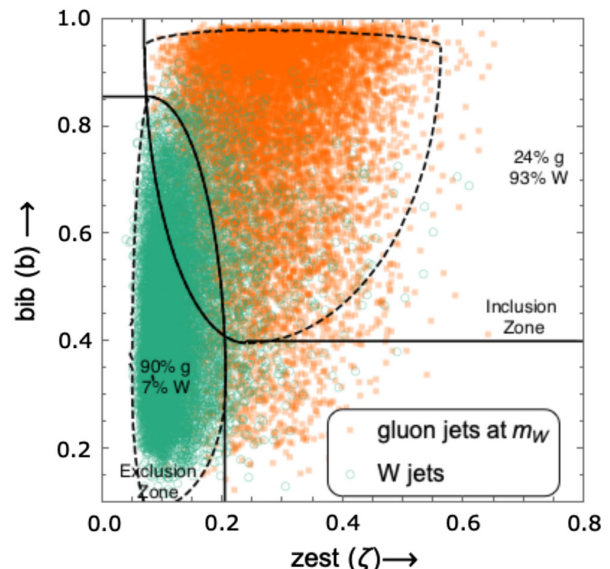

FIG. 11. Scatter plot in *bib*-zest plane for *W*-boson jet and the corresponding gluon jet, with the 90% Gaussian contour and inclusion and exclusion-zone cuts specified on it.

contour, illustrated with a dashed line, is shown in Fig. 11 for the *W*-boson initiated jets along with the corresponding gluon background. By 90% contour, we refer to the closed curve that includes 90% of the jets inside it. The tangents are drawn parallel to *bib*-axis and zest-axis on the gluon distribution contour to intercept with the plot frame at the bottom and to the left, respectively. The solid line represents an open curve made by combining the tangents and right-side of the contour meeting the tangents. The background region is defined as the area to the left of the solid curve. Similar exercise is also done on the heavy particle initiated jet distribution. The signal region is defined as the region to the right of the solid curve on the heavy particle distribution. We define two new cuts (1) the "inclusion-zone" cut and (2) the "exclusion-zone" cut. The inclusion-zone cut is the solid curve on the signal while the exclusion-zone cut is the solid curve on the background. We note that for the case of a *W* boson-initiated jet with 90% of the signal as the desired value, an exclusion-zone cut provides better statistics over an inclusion-zone cut.

The relative performance of these bivariate cuts in comparison to zest alone is studied through ROC curves in Fig. 12. We present three types of ROC curves for each type of heavy particle jet (a) zest based cuts only (represented by dashed lines), (b) cuts based on inclusion-zone obtained by including the percentage of events contained within the primary contour (represented by dotted lines), and (c) cuts based on exclusion-zone obtained by excluding the percentage of events contained within the primary contour (represented by solid lines). From the plot, we note that for a high signal rate, exclusion-zone statistics provides slightly better discrimination than only zest-based cuts for the *W*, *Z* bosons, however inclusion/exclusion-zone

FIG. 10. (a)–(c) Scatter plots in *bib*-zest plane for heavy particle and corresponding gluon jets.
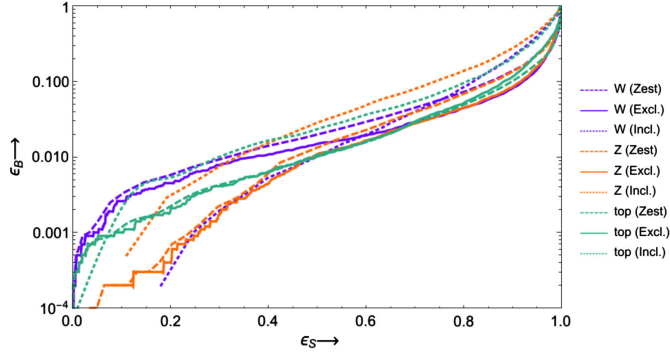
FIG. 12. ROC curves for heavy particle jets with zest, inclusion and exclusion—filter based binary classifiers.

provide no significant improvement in comparison to a zest only filter for the top quark-initiated jet. On the other hand, inclusion-zone statistics is more efficient when high gluon rejection is required and small signal rate is acceptable.

VIII. GENERALIZED ZEST (p-ZEST)

As pointed out earlier in Sec. II, in the special case of n-equal transverse momenta particles in the jet, zest is a function of *particle multiplicity* in a jet. In this section, we aim to generalize the observable such that the particle multiplicity emerges as a limiting case of generalized zest. We define the generalization of zest (or, "p-zest" for brevity) as follows:

$$\zeta_p = -\frac{1}{\log(\sum_i e^{-P_T^{(p)}/|\mathbf{p}_{Ti}|^p})}. \tag{4}$$

Here the sum is over all the final state particles in the jet, as earlier, and $P_T^{(p)}$ is defined as

$$P_T^{(p)} = \sum_i |\mathbf{p}_{Ti}|^p, \tag{5}$$

where $p > 0$. The ability to tune the parameter p may also provide a new way of looking at the jet substructure by allowing us to vary the net contribution of the soft sector.

For $p > 1$, the contribution of the low-pT particles to the determination of $P_T^{(p)}$ will reduce. Thus, mostly high-pT particles will contribute to p-zest. In the extreme limit of $p \to \infty$, only the parton with the largest p$_T$ contributes and ζ_∞ approaches 1. Therefore, this limit offers no useful jet discrimination ability. In contrast, in the limit $p \to 0$, all particles contribute equally giving

$$\zeta_0 = \frac{1}{n - \log n}, \tag{6}$$

where n is the number of particles in the jet. For a large enough n such that $n \gg \log n$, the above expression
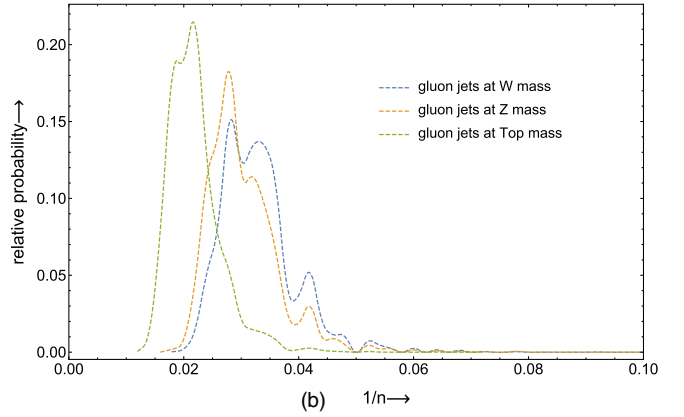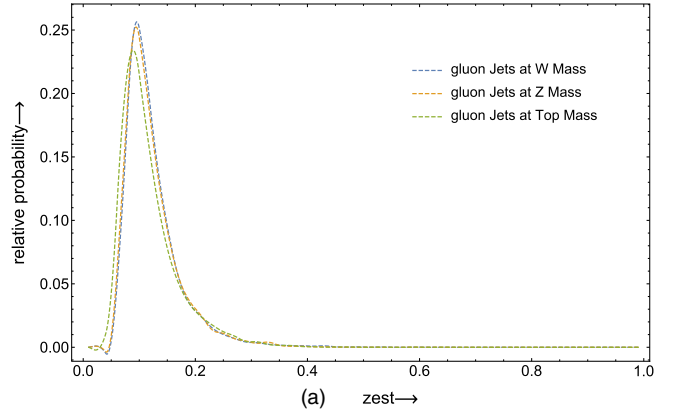

FIG. 13. (a) Zest distribution of gluon jets at various heavy particle mass windows, and (b) $1/n$ distribution of gluon jets at various heavy particle mass windows.

reduces simply to the inverse of particle multiplicity. Thus, we note that closer the value of p is to 0, the stronger is the correlation of p-zest and multiplicity. As p is increased, the correlation becomes smaller and vanishes for the extreme limit of $p \to \infty$. Though the observables zest (with $p = 1$) and multiplicity are correlated, zest exhibits interesting properties such as stability against change in the jet mass and stability against change in the jet radius for gluon-initiated jets, which are absent in the $1/n$ distribution. This is confirmed by the plots in Figs. 13 and 14. In Figs. 13(a) and 13(b), the zest and the $1/n$ distribution curves for gluon-initiated jets with masses about heavy particle mass, are presented. From the plots, we see that the $1/n$ distribution shifts to the left as the mass of the jet is increased while the zest distribution remains more or less unchanged.[5] Similarly, the stability of the zest distribution

---

[5]Note that this is not to say that $1/n$ does not offer gluon jet vetoing ability, instead it may very well provide the ability to veto gluon jets like zest. From the trend observed in Fig. 13(b), we expect the heavier jets to shift further to the left and hence a cut at around 0.04 may cut out the gluon jets largely. However, as we will show in Sec. VIII B, for the heavy top quark-initiated jets, zest offers an appreciable improvement over multiplicity.
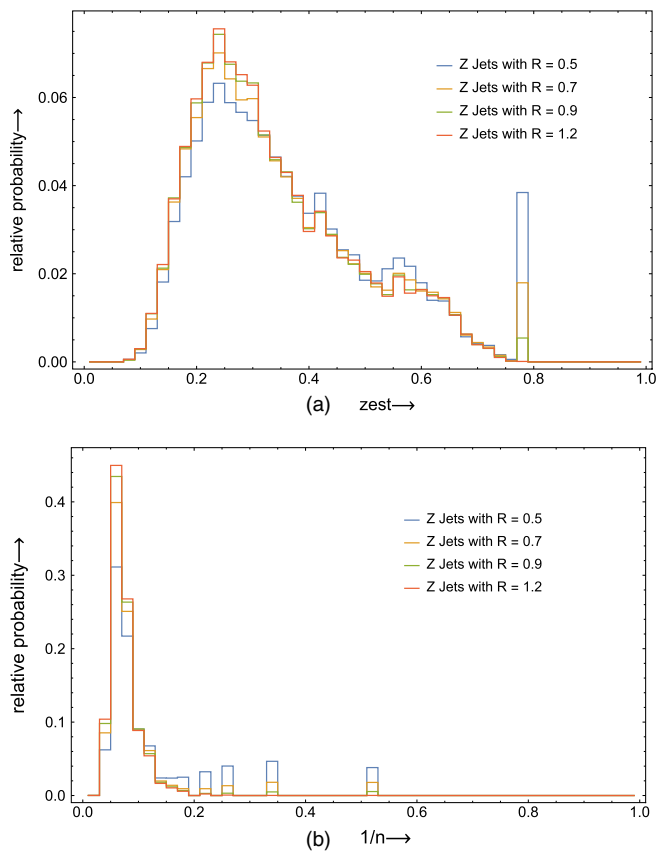
FIG. 14. (a) Zest distribution of Z boson-initiated jets for different R values, and (b) $1/n$ distribution of Z boson-initiated jets for varying jet radii.

FIG. 15. ROC curves for Z boson-initiated jets with various choices of p compared against the inverse of multiplicity as a filter.

A. Z boson-initiated jet

For the Z boson-initiated jet, we generate the ROC curves for $p = \{0.3, 0.5, 1, 1.5, 2, 3\}$ and particle multiplicity, as shown in Fig. 15. The solid cyan line with open squares on it illustrates the multiplicity distribution while the other solid lines with the corresponding legends represent the p-zest distributions. From the plots, we find that multiplicity provides a better discrimination in comparison to zest ($p = 1$) while p-zest with $p = 0.3$–0.5 provides the optimal discrimination. This is also confirmed by looking at Fig. 16 where with 80% of signal efficiency as the accepted value, we plot the curve for the background jet rejection offered by p-zest with varying values of parameter p. From the plot, we find that $p = 0.3$–0.5 provides the best discriminating ability for a Z-boson initiated jet, although it is not significantly better than just multiplicity.

B. Top quark-initiated jet

The ROC curves for top-initiated jets with p-zest and $1/n$ as the observables are presented in Fig. 17. Interestingly, here we find that p-zest for $p = 0.3$–1.0 performs significantly better than the $1/n$ distribution. The same is confirmed further by the background rejection

curves to varying the jet radius is shown in Fig. 14(a) for a Z boson-initiated jet of jet energy 300 GeV while the corresponding $1/n$ curves are shown in Fig. 14(b).[6] From the plots, we clearly see that the zest distribution curves are mostly stable to the change in jet radius while for the $1/n$ distribution, the curve for $R = 0.5$ appears to be shifted away from other R values. This is expected as zest is mostly unaffected by the inclusion or exclusion of a few soft particles while the multiplicity curves are somewhat sensitive to it.

Noting these differences, we investigate the discrimination ability offered by p-zest over particle multiplicity in a jet and zest ($p = 1$). It will be interesting to find p that provides the best discrimination. For this, we consider two type of heavy particle jets: (a) a Z boson-initiated jet, and (b) a top quark-initiated jet. The detailed analysis of these results is presented in the following subsections.

---

[6]Here we have used jets of initial energy 300 GeV to demonstrate this difference. This is because a Z boson at 500 GeV is extremely collimated and $R = 0.5$ is sufficient to capture all the radiations originating from this extremely boosted Z boson.
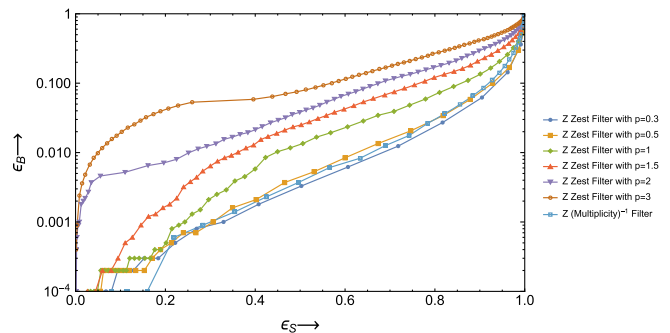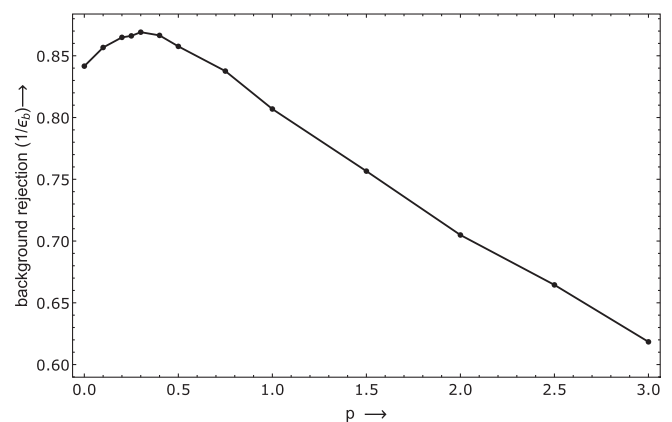

FIG. 16. Discrimination power offered by p-zest at fixed signal efficiency of 80% for Z-boson initiated jets.
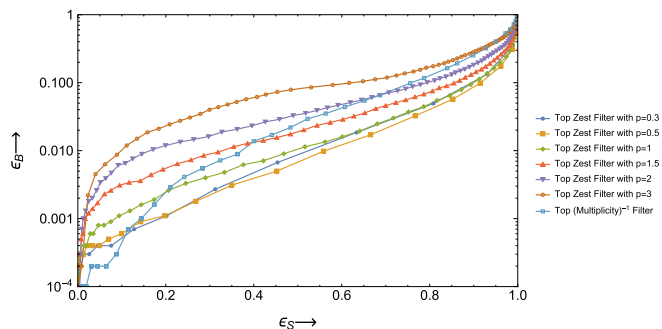
FIG. 17. ROC curves for top quark-initiated jets with various choices of p compared against the inverse of multiplicity as a filter.
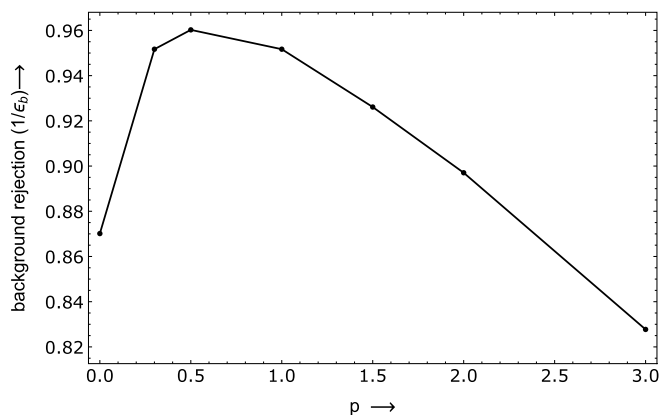

FIG. 18. Discrimination power offered by p-zest at fixed signal efficiency of 80% for top quark initiated jets.

efficiency curve with respect to p at fixed signal efficiency of 80%, as shown in Fig. 18.

From this analysis, we find that p-zest allows us to find the p value that optimizes the discrimination. We note that a p value of about 0.3 to 0.5 is optimal for the heavy particles studied here.

It is interesting to note that the discrimination ability offered by our preliminary p-zest study for the top quark-initiated jets is comparable with that of a class of ML-based top taggers [23]. Although our study does not include proton-proton collisions and detector effects considered in [23], we are not claiming a strict comparison of our results with ML-based techniques. While these ML algorithms offer the highest discrimination ability for heavy particle jets, the particular features that offer this improvement remain largely unknown. p-Zest may offer some insight into the physics of ML-based taggers. Therefore, we propose that studying IRC unsafe observables based on physical principles may provide a window into the physics hidden in ML based techniques.

IX. HADRONIZATION MODEL DEPENDENCE

As pointed out earlier, p-zest is collinear unsafe and cannot be calculated by standard perturbative techniques,

therefore we rely on Monte Carlo event generators that incorporate a suitable hadronization model for its computation. In this section, we study the effect of varying hadronization models on zest. This will allow us to understand how the observed value of zest is sensitive to the modeling of nonperturbative effects. This check is important as the observable can only be computed for hadronic final states.

To study the hadronization model dependence of zest, we vary the nonperturbative physics modeling by: (a) changing the hadronization parameters in PYTHIA, and (b) changing the hadronization model by using the HERWIG event generator to simulate the event. This will modify the spectrum of the final state primary hadrons, therefore allowing us to study the hadronization model dependence for zest.

In Fig. 19(a), we present the zest-distribution curves for a Z boson-initiated jet with different clustering mechanisms for implementing the long distance physics. The PYTHIA 8 event generator that we use for simulating the events, incorporates the Lund string model of hadronization [34] to cluster final state partons into hadrons. This model is built upon a "string" analogy, i.e., as the separation between the two partons increases, the potential energy stored in the string rises
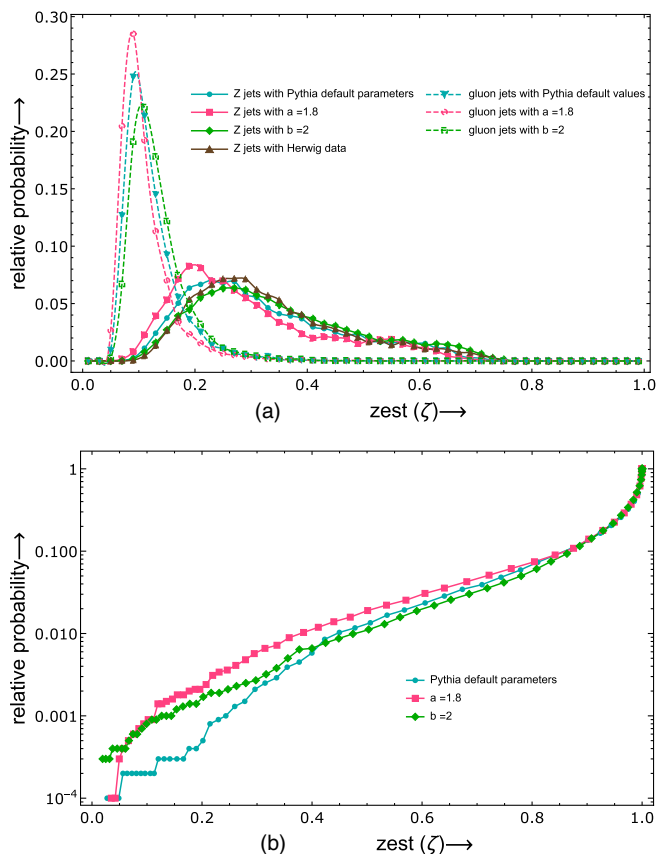

(a)

(b)

FIG. 19. (a) Zest distribution curves for Z boson-initiated jets with various choices of modeling the nonperturbative physics, and (b) ROC curves for Z boson-initiated jets by varying the hadronization parameters in PYTHIA.

linearly. At some point, the potential energy stored is so large that the string breaks forming a $q\bar{q}$ pair leading to individual hadrons, with the "fragmentation function" given by,

$$f(z) \propto z^{-1}(1-z)^a \exp(-bm_\perp^2/z), \qquad (7)$$

where z is the energy fraction carried by the hadron and $m_\perp$ is its transverse mass while a, b are some adjustable parameters. Changing a and/or b in the above equation, changes the way fragmentation happens, hence modifying the final state spectrum.[7] The zest-distribution shifts as shown in Fig. 19(a). Moreover, we observe a similar shift for the gluon zest distribution (background) and the performance curves change only very slightly, as shown in Fig. 19(b). For HERWIG, we cannot simulate single offshell gluons, thus we do not have corresponding curve for the gluon zest and ROC. However, from the heavy particle distribution for the HERWIG simulated event shown by the solid brown curve (with filled triangles) in Fig. 19(a), we see that the zest distribution lies well within the extreme bounds of hadronization parameters used in PYTHIA, shown by the solid magenta (filled squares) and green (filled diamonds) curves in Fig. 19(a). Thus, from our analysis we conclude that although zest is not calculable perturbatively, but zest based discrimination is stable against different hadronization models.

Since zest is collinear unsafe, its discrimination ability may depend upon the resolution of the detector. We have also verified that while the zest curves shift if we coarse-grain the angular resolution between the particles for both the gluon-initiated and heavy-particle-initiated jets, they

shift systematically such that the discrimination ability of the observable remains unaffected.

X. CONCLUSION

We have presented a new jet substructure observable, *zest*, and discussed its potential for discriminating standard model heavy particles forming jets from the QCD background of gluon-initiated jets. We have shown that the zest distribution of gluon-initiated jets is stable against the change in jet mass, change of global color flow of the partons and inclusion or exclusion of a few soft particles into/from the jet. These properties make it a suitable observable for vetoing the large gluon background at the colliders. Though zest is a nonlinear and collinear unsafe observable, we have demonstrated that zest-based discrimination is largely insensitive to different hadronization models. We have shown that zest can be generalized through a real parameter, p, which can be optimized to further enhance the discrimination ability. A p value between 0.3 to 0.5 is found to provide the optimal discrimination ability for all the heavy particle jets. We have also shown that for a top quark-initiated jet, p-zest provides a significant improvement over particle multiplicity in a jet. The discrimination provided by p-zest (with $p = 0.3$–0.5) for the top quark-initiated jets approaches ML-based results, and we propose that looking at other IRC unsafe jet observables may help to uncover the physics hidden through such ML-based techniques.

---

[7]The default values for the parameters a and b used in PYTHIA 8 are 0.3 and 0.8 respectively. Here we will vary these values to their extremum limits by setting a to 1.8 and b to 2.

[1] G. G. Barnafoldi, A. G. Agocs, and P. Levai, Underlying event studies for LHC energies, AIP Conf. Proc. **1348,** 124 (2011).

[2] M. Cacciari, G. P. Salam, and S. Sapeta, On the characterisation of the underlying event, J. High Energy Phys. 04 (2010) 065.

[3] D. Krohn, M. D. Schwartz, M. Low, and L.-T. Wang, Jet Cleansing: Separating data from secondary collision induced radiation at high luminosity, Phys. Rev. D **90,** 065020 (2014).

[4] D. Adams *et al.*, Towards an understanding of the correlations in jet substructure, Eur. Phys. J. C **75,** 409 (2015).

[5] A. Altheimer *et al.*, Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 2012, Eur. Phys. J. C **74,** 2792 (2014).

[6] A. Abdesselam *et al.*, Boosted objects: A probe of beyond the standard model physics, Eur. Phys. J. C **71,** 1661 (2011).

[7] A. Altheimer *et al.*, Jet substructure at the tevatron and LHC: New results, new tools, new benchmarks, J. Phys. G **39,** 063001 (2012).

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

[9] S. Marzani, L. Schunk, and G. Soyez, The jet mass distribution after Soft Drop, Eur. Phys. J. C **78,** 96 (2018).

[10] F. A. Dreyer, L. Necib, G. Soyez, and J. Thaler, Recursive soft drop, J. High Energy Phys. 06 (2018) 093.

[11] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Techniques for improved heavy particle searches with jet substructure, Phys. Rev. D **80,** 051501 (2009).

[12] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches, Phys. Rev. D **81**, 094023 (2010).

[13] D. Krohn, J. Thaler, and Lian-Tao Wang, Jet trimming, J. High Energy Phys. 02 (2010) 084.

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

[15] J. M. Butterworth, A. R. Davison, Mathieu Rubin, and Gavin P. Salam, Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider, Phys. Rev. Lett. **100**, 242001 (2008).

[16] S. Marzani, L. Schunk, and G. Soyez, A study of jet mass distributions with grooming, J. High Energy Phys. 07 (2017) 132.

[17] A. J. Larkoski, I. Moult, and B. Nachman, Jet substructure at the large hadron collider: A review of recent advances in theory and machine learning, Phys. Rep. **841**, 1 (2020).

[18] S. Marzani, G. Soyez, and M. Spannowsky, *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-object Phenomenology*, Lecture Notes in Physics Vol. 958 (Springer, 2019), ISBN 978-3-030-15709-8.

[19] R. Gavai, A. Jain, and R. Sharma, Unravelling medium effects in heavy ion collisions with zeal (2015).

[20] M. Cacciari, G. P. Salam, and G. Soyez, The anti-k_t jet clustering algorithm, J. High Energy Phys. 04 (2008) 063.

[21] T. Sjostrand, S. Mrenna, and P. Z. Skands, A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. **178**, 852 (2008).

[22] J. Bellm *et al.*, Herwig7.0/Herwig + +3.0 release note, Eur. Phys. J. C **76**, 196 (2016).

[23] A. Butter, T. Plehn, and R. Winterhalder, The machine learning landscape of top taggers, SciPost Phys. **7**, 014 (2019).

[24] O. Amram and C. M. Suarez, Tag N' train: A technique to train improved classifiers on unlabeled data (2020).

[25] L. Bradshaw, R. K. Mishra, A. Mitridate, and B. Ostdiek, Mass agnostic jet taggers, SciPost Phys. **8**, 011 (2020).

[26] J. Dolen, P. Harris, S. Marzani, S. Rappoccio, and N. Tran, Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure, J. High Energy Phys. 05 (2016) 156.

[27] Y.-C. Janice Chen, C.-W. Chiang, G.. Cottin, and D. Shih, Boosted W and Z tagging with jet charge and deep learning, Phys. Rev. D **101**, 053001 (2019).

[28] E. A. Moreno, O. Cerri, J. M. Duarte, H. B. Newman, T. Q. Nguyen, A. Periwal, M. Pierini, A. Serikova, M. Spiropulu, and J.-R. Vlimant, JEDI-net: A jet identification algorithm based on interaction networks, Eur. Phys. J. C **80**, 58 (2020).

[29] CMS Collaboration, Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment, Report No. CMS-PAS-JME-18-002 (2019).

[30] M. Cacciari, G. P. Salam, and G. Soyez, FastJet User manual, Eur. Phys. J. C **72**, 1896 (2012).

[31] Y. L. Dokshitzer, A. Lucenti, G. Marchesini, and G. P. Salam, On the QCD analysis of jet broadening, J. High Energy Phys. 01 (1998) 011.

[32] A. J. Larkoski and J. Thaler, Unsafe but calculable: Ratios of angularities in perturbative QCD, J. High Energy Phys. 09 (2013) 137.

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

[34] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, Parton fragmentation and string dynamics, Phys. Rep. **97**, 31 (1983).