Eur. Phys. J. C (2018) 78:334
https://doi.org/10.1140/epjc/s10052-018-5819-8

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

Fractal based observables to probe jet substructure of quarks and gluons

Joe Davighi[1], **Philip Harris**[2,a]

[1] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge, UK
[2] CERN, European Organization for Nuclear Research, Geneva, Switzerland

Received: 3 May 2017 / Accepted: 17 April 2018 / Published online: 25 April 2018

Abstract New jet observables are defined which characterize both fractal and scale-dependent contributions to the distribution of hadrons in a jet. These infrared safe observables, named Extended Fractal Observables (EFOs), have been applied to quark–gluon discrimination to demonstrate their potential utility. The EFOs are found to be individually discriminating and only weakly correlated to variables used in existing discriminators. Consequently, their inclusion improves discriminator performance, as here demonstrated with particle level simulation from the parton shower.

1 Introduction

A hadronic jet is produced from an initial parton via a sequence of perturbative QCD branching interactions (the parton shower), followed by the non-perturbative conversion of partons to the hadrons we observe in experiments (hadronization). A Markov chain description of the parton shower suggests the spatial distribution of partons will exhibit some fractal character [1–6], and this will be inherited by the final hadron distribution (invoking local parton-hadron duality [7]). However, true scale invariance of the hadron distribution within a jet is broken by the running of the branching probability, termination of the shower due to hadronization, and finite detector resolution. Here we define new observables to characterize jet branching structure, named Extended Fractal Observables (EFOs), which accommodate deviations from fractal structure through simple parametrizations. The idea is to apply box-counting techniques, used widely in the study of dynamical systems and scale invariant objects, to the substructure of QCD jets. Box counting has previously been employed in particle physics to calculate the fractal dimension of electromagnetic showers [8] for highly granular calorimetric reconstruction. Here, we extend the generality and information content of this technique in our characterization of QCD jets.

The motivation for this study is twofold. Firstly, we would like to characterize the spatial substructure of jets into a set of new observables. Secondly, we would like to demonstrate the use of such observables in the discrimination of quark and gluon jets. Quark and gluon discrimination has long been used as a tool to enhance the sensitivity of signatures with additional quarks [9–12]. In particular, weak boson fusion induced Higgs-production is enhanced due to the distinct signature of two additional hard quark jets in the gluon-dominated forward region of the detector [9,13–21]. Quark and gluon tagging are also expected to be useful for physics searches *beyond* the Standard Model, including the detection of supersymmetric particles [22,23]. Additionally, if well designed, these taggers can be further extended to the subjets of boosted boson signatures [24]. We demonstrate that modest improvements can be made to existing quark–gluon taggers by incorporating the new jet observables defined in this paper.

Finally, our construction of pixel-based jet observables resonates with the recent development of the jet image paradigm [25,26], in which the energy measured in each detector cell is interpreted as the intensity of a pixel in a 2D image. Within this approach, powerful machine-learning algorithms for classifying images have been brought to bear on a range of jet classification problems. This has included tagging boosted weak bosons [26,27], boosted top quarks [28], and heavy-flavors [29,30].

We define EFOs in the following section. In Sect. 3 we analyze the performance of these observables in quark–gluon discrimination, before concluding.

[a] e-mail: philip.coleman.harris@cern.ch

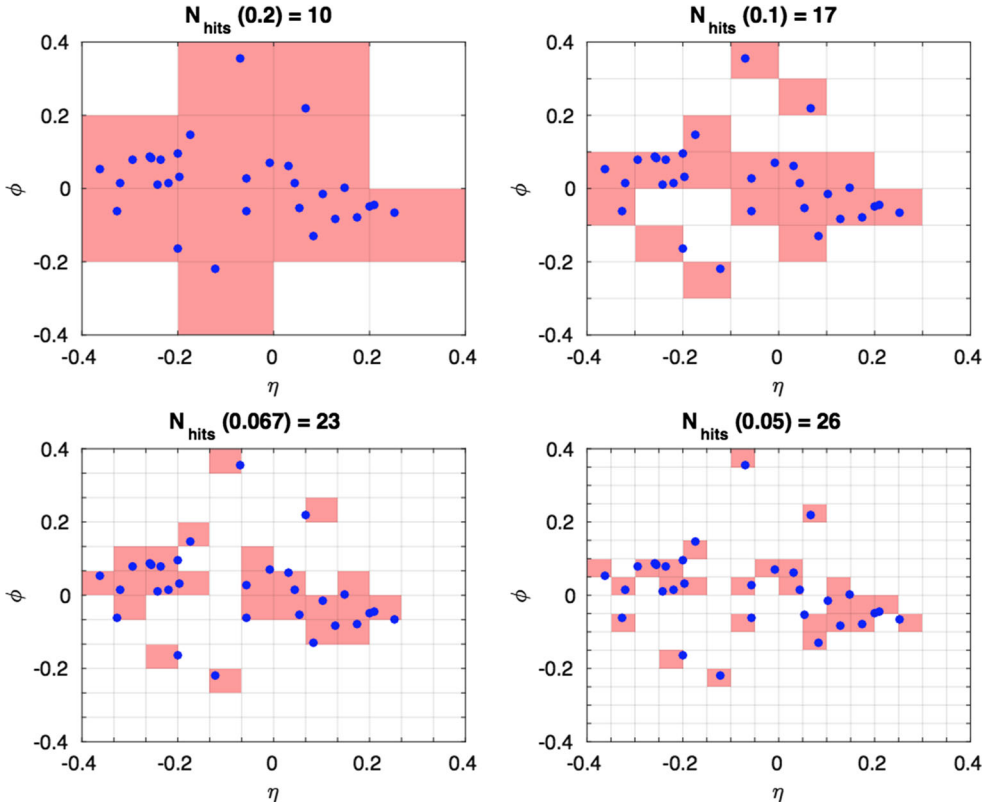


Fig. 1 An illustration of the iterated box-counting procedure used to calculate fractal-based quantities on a set of points. The filled blue circles are the (η, ϕ) angular coordinates of the hadrons within a particular sample jet (in particular, this jet has total $p_T = 157$ GeV, and 30 constituent hadrons). The box-counting is illustrated for four sample scales, corresponding to successively finer ϵ values of 0.2, 0.1, 0.067 and 0.05. The cells registering particle hits are highlighted with red shading

2 Extended fractal observables

The computation of the EFOs is performed on a jet by jet basis using a variation of the Minkowski-Bouligand (box-counting) dimension, as follows.

2.1 Variable definitions

To define our variables we implement a two-stage recipe: firstly, the jet cone is divided in the familiar (η, ϕ) angular coordinates into a square grid of cells, each cell having side-length ϵ. For a given scale ϵ, we count the number of cells $N_{hits}\,(\epsilon)$ which register particle hits with a total transverse momentum greater than some pixel-level soft cutoff, in this study chosen to be $p_T > 1.0$ GeV. This low energy cut represents a limiting threshold due to detector resolution. This counting is iterated over a range of scales, as is illustrated in Fig. 1. The second stage is to fit smooth functions to the variation of $y = \log N_{hits}\,(\epsilon)$ with $x = \log(1/\epsilon)$, and to extract the parameters of the fit as a set of (correlated) jet observables, which we call Extended Fractal Observables (EFOs). This is a generalization of the traditional box-counting method, in which only linear functions $y = mx + c$ are fitted, with the gradient m identified as the fractal dimension [8].

Indeed, in Fig. 2 there is no distinct region of linear scaling, as would be needed to extract a fractal dimension. Rather, $\log N_{hits}\,(\epsilon)$ levels off smoothly from large to small scales as saturation is approached, motivating a non-linear fit to extract whatever information this curve might encode about the jet. In particular, the hadronization region (i.e. at small ϵ) obviously carries non-perturbative information sensitive to the flavor of the jet. The observed curves are distinct between quarks, gluons and b-quarks, as summarized in Figs. 2 and 3. This scaling is a fundamental property of QCD resulting from the differences in the splitting of quarks and gluons. Further measurements of this scaling allows for an alternative approach to extract QCD properties such as the strong coupling constant [32,33].

The generic plateauing curves in Fig. 2 can be fitted by almost any non-linear function (given a suitably restricted range in x), so we studied fit functions with at most three parameters, for speed and robustness of fitting. Fits were carried out simply by a binned χ^2 minimization of the chosen function. Example fit functions included the following:

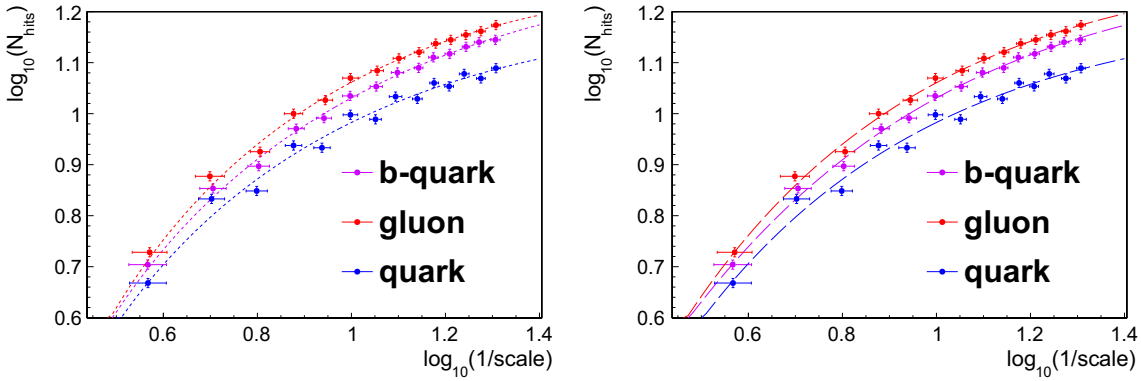


Fig. 2 Left: logarithmic fits to $\log N_{hits}\ (\epsilon)$ against $\log\,(1/\epsilon)$ for light quarks, bottom quarks, and gluons, of the form $y = p_0 + p_1 x + p_2 \log x$. The values of the fitted parameters $\{p_i\}$ define one possible set of Extended Fractal Observables. Right: fits to $\log(N_{hits})$ against $\log\,(1/\epsilon)$ using an asymptotically saturating fitting function, specifically $y = p_0 + p_1 \tanh(x - p_2)$

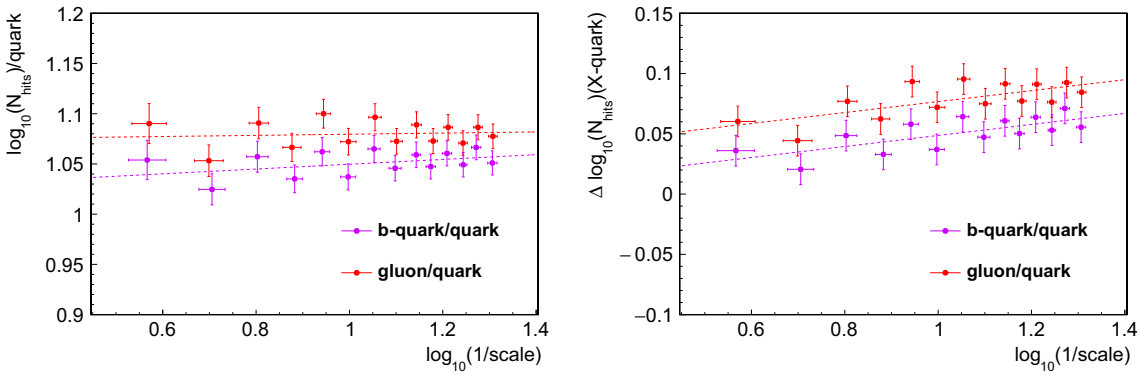


Fig. 3 Left: the *ratio* of $\log(N_{hits})$ with respect to the quark values, for b-quarks and gluons, as a function of $\log\,(1/\epsilon)$. A linear fit is added for comparison. Right: the *difference* of $\log(N_{hits})$ with respect to the quark values, for b-quarks and gluons. In the Modified Leading Logarithmic Approximation (MLLA), the differences in hadron multiplicity between quarks, b-quarks and gluons are predicted to be energy independent [31]. The small but non-zero slopes in this plot reflect the fact that box-counting at a given angular scale probes spatial information in addition to the rate of splitting at the corresponding energy scale

1. logarithmic fits of the form $y = p_0 + p_1 x + p_2 \log x$.
2. quadratic fits: $y = p_0 + p_1 x + p_2 x^2$.
3. hyperbolic tangent fits: $y = p_0 + p_1 \tanh(x - p_2)$.

The values of the best fit parameters $\{p_i\}$ for each fitting function constitute three possible sets of EFOs. For a polynomial in $x = \log(1/\epsilon)$, like the quadratic fit function, the fit reduces to a matrix inversion and thus has a well-defined convergence. The other two parametrizations are not polynomials, hence we perform a χ^2 minimization.

Functions which actually saturate, such as the hyperbolic tangent parametrization above, are more physically motivated because they can model the saturation itself (asymptoting to the jet multiplicity). However, for the range of box scales used in our study (of width $\epsilon \geq 0.05$, – see 2.2 below), and for all but the lowest p_T jets, the *non*-saturating fit functions also provide adequate models for the observed scaling. For the purpose of quark–gluon discrimination (see Sect. 3), the logarithmic fitting function was found to give the best discrimination performance of the three functions above (see Fig. 6 to compare the performance between the logarithmic and hyperbolic tangent fitting functions).

2.2 The range of box-counting scales

The range of angular scales ϵ has been chosen by paving the jet cone with a square grid of $N \times N$ cells, where the splitting scale N ranges in integer steps from 3 to 16. For each N, the angular scale is $\epsilon = 2R/N$, where R is the jet radius, in this study $R = 0.4$. The coarsest ϵ scale chosen, corresponding to $N = 3$, is essentially the coarsest scale carrying potentially discriminating information (for $N = 2$ the jet cone would be divided into four quarters, all of which will register a hit for realistic jet shapes). The finest ϵ scale chosen is $\epsilon_{min} = 0.8/16 = 0.05$, because this is approximately the angular detector resolution in both LHC experiments, CMS

and ATLAS [34,35]. For the $p_T \geq 100$ GeV jets studied here, the number of hits is just beginning to saturate at this scale (see Fig. 2), so we are probing into the hadronization region prior to the flat plateau.

Finally, we would like to highlight that these fractal-based observables are similar in spirit to calculating subjet rates of jets [15,36], given subjets clustered using the p_T-independent Cambridge-Aachen algorithm [37]. Both observables compute p_T-independent branching information on a succession of angular scales down to some threshold. And both observables perform what is essentially a further clustering on the substructure of the jet to extract this information pertaining to the branching history of the jet. In light of this, the EFO approach could be extended to utilize subjet counts (instead of hit grid cell counts) to assign scale-dependent multiplicities $N(\epsilon)$.

2.3 Infrared and collinear safety

Preserving infrared and collinear safety ensure calculability in perturbative QCD. An observable is infrared (collinear) safe if its value is unchanged by the emission of soft (comoving) particles. The EFOs, as defined in 2.1 with a pixel-level soft cutoff, are fully IRC safe.

Firstly, the box counting procedure is intrinsically collinear safe: if one particle splits into two particles with the same (η, ϕ) coordinates, we still count just one cell hit by both daughter particles, at any finite scale of probing. Hence collinear splittings will not affect the number of cells $N_{hits}(\epsilon)$ to register particle hits at any choice of scale. On the other hand, infrared safety of the EFOs can only be engineered by imposing some low momentum cutoff to cleanse the jet of its soft constituents. However, this soft cutoff must be implemented consistently with collinear safety. If we simply discarded all soft hadrons with, say $p_T < 1$ GeV, this would spoil collinear safety. To see this, consider the following pathological example: if a particle with $p_T = 1.5$ GeV splits into two comoving particles with $p_T = 0.8$ GeV and $p_T = 0.7$ GeV, then both would be discarded by a particle-level soft cut, and so $N_{hits}(\epsilon)$ would not be invariant under this collinear splitting.

This is remedied by defining a pixel-level (rather than particle-level) sort cutoff. That is, we only consider a cell to register a hit if it measures a total p_T greater than our soft cutoff of 1 GeV. This way, if the troublesome 1.5 GeV particle in the example above splits collinearly into any number of daughters, the pixel still measures a total p_T of 1.5 GeV, and so registers a hit regardless of these splittings. Thus, box-counting with a pixel-level soft cutoff is fully IRC safe. In addition, a pixel-level rather than particle-level cut is more naturally realized experimentally since a pixel hit is consistent with an LHC detector cell.

Numerically, the performance of a quark–gluon discriminant built using the EFOs was found to be essentially insensitive to varying the value of this p_T cut (over values between 0.1 GeV and 1.0 GeV), suggesting the variables are not strongly shaped by the IR emission, at least in simulations. In the following section, a p_T cut of 1 GeV is used throughout. Finally, we acknowledge that pixel-level cutoffs have been used previously in the context of jet images analyses (for example in [25]) to ensure IRC safety in the same context.

3 Performance in Quark–Gluon discrimination

We now investigate whether these observables might be a useful new tool in the important and challenging problem of distinguishing light quarks from gluon jets.

3.1 Event generation and setup

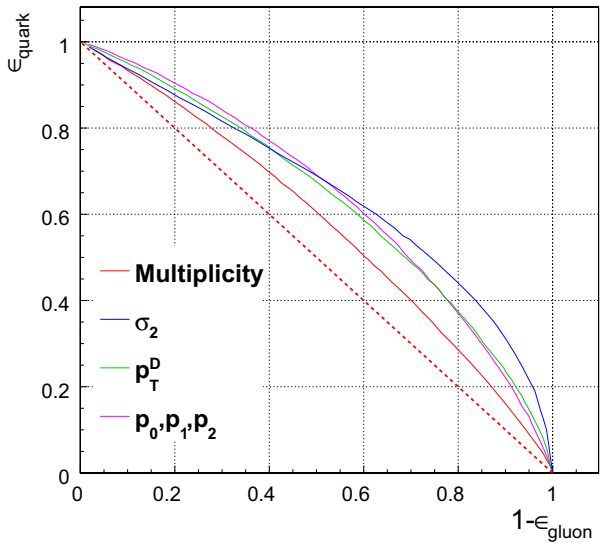
In this study, we use QCD dijet samples at a center-of-mass energy of 13 TeV. Because previous quark–gluon studies have revealed that discrimination performance varies a lot between the different generators [9–11,14,38],[1] we here produce and shower events (at leading order) using both Herwig++ (version 2.7.0 with tune UE-EE-5C) [39,40] and Pythia 8 (version 8.185 with tune CUETP8M1) [41], with order 150k events in each. Jets are clustered with the anti-k_T algorithm using the final state particles following showering and hadronization; a cone size of $R = 0.4$ and the FastJet code package [42] are used for the jet clustering. The EFOs (here computed using the logarithmic fitting function), along with a set of other established jet observables, have been computed for the highest p_T jet in each event. We define the flavor of that jet by matching to the highest-p_T parton within $R < 0.3$ of the jet axis, and classify the event as signal (background) if matched to a light quark (gluon).[2]

As a baseline for comparison, we shall consider the variables currently used by the Compact Muon Solenoid (CMS) quark–gluon tagger, which are [10]: (i) the total number of reconstructed particles in the jet (the multiplicity) [43]; (ii) the $p_T D$ variable ($C_1^{\beta=0}$) [44],

$$p_T D = \frac{\sqrt{\Sigma_i p_{T,i}^2}}{\Sigma_i p_{T,i}}, \tag{1}$$

[1] Herwig has been consistently seen to give the more conservative estimates of discrimination power, both with respect to Pythia and real LHC data.

[2] Note that b(bottom)-jets may be efficiently identified using a secondary vertex tagger, and separately vetoed.



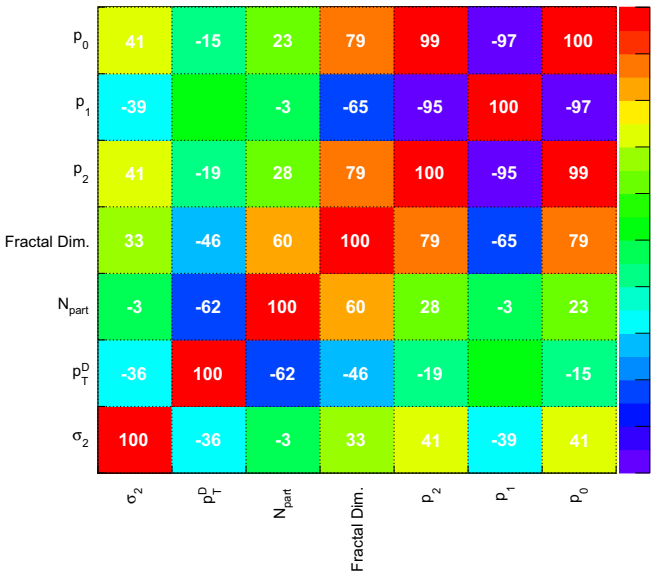


Fig. 4 Left: single variable performance ROC curves. The EFOs, minor axis, and $p_T D$ are significantly more discriminating than multiplicity. The EFOs are most discriminating for high signal efficiency ($\gtrsim$ 70%), below which jet minor axis becomes most discriminating. Right: linear correlation coefficients between pairs of variables, for quark jets (the values are similar for gluon jets). We see only weak correlations between the EFOs and the three existing QGD variables

where i sums over the constituents of the jet, which describes the distribution of transverse momentum between the particles in the jet; and (iii) σ_2, the (p_T-weighted) semi-minor axis of the jet in the (η, ϕ) plane [10], defined by

$$\sigma_2 = (\lambda_2 / \Sigma_i p_{T,i}^2)^{1/2}, \tag{2}$$

where λ_2 is the smaller eigenvalue of the 2×2 symmetric matrix with components $M_{11} = \Sigma_i p_{T,i}^2 \Delta\eta_i^2$, $M_{22} = \Sigma_i p_{T,i}^2 \Delta\phi_i^2$, and $M_{12} = -\Sigma_i p_{T,i}^2 \Delta\eta_i \Delta\phi_i$. Throughout this study, we build multi-variable quark–gluon discriminants using a boosted decision tree (BDT), implemented using the Toolkit for Multivariate Analysis (TMVA) via adaptive boosting. The p_T of the quark and gluon samples are reweighted to match the exact same kinematics in both cases, so as to avoid selection biases induced by kinematic differences in the simulation.

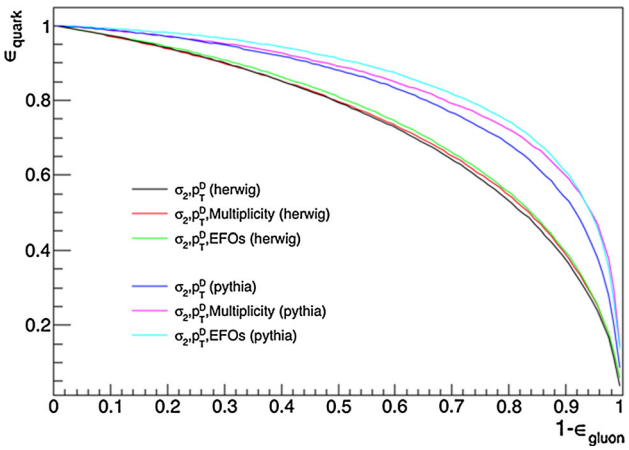


Fig. 5 ROC curves for BDT discriminators constructed from various combinations of observables, as indicated by the legend, for events showered using both Herwig and Pythia with jet $p_T \geq 100$ GeV. The discrimination is superior in Pythia. We see in both event generators that including the EFOs rather than multiplicity (which is used in the CMS tagger) yields a marginally better performance

3.2 Results

We first compare the discriminator performance of single variables and the correlations between them, before going on to compare multi-variable taggers built with and without inclusion of the new EFO observables.

We can measure discriminator performance by receiver operator characteristic (ROC) curves, which plot background rejection against signal efficiency. Roughly speaking, the more convex the curve, the better the performance. The left plot of Fig. 4, made using the Herwig samples, shows that the EFOs[3] are individually well-discriminating, particularly if we seek high signal efficiency. Their performance is significantly better than that of the jet multiplicity variable.

[3] We use a BDT discriminator built from the combination of the three EFOs, p_0, p_1 and p_2. While the combination of all three EFOs adds little discrimination beyond that of a single EFO due to their near-perfect correlation, the selection of any single p_i would be arbitrary for the sake of this comparison.

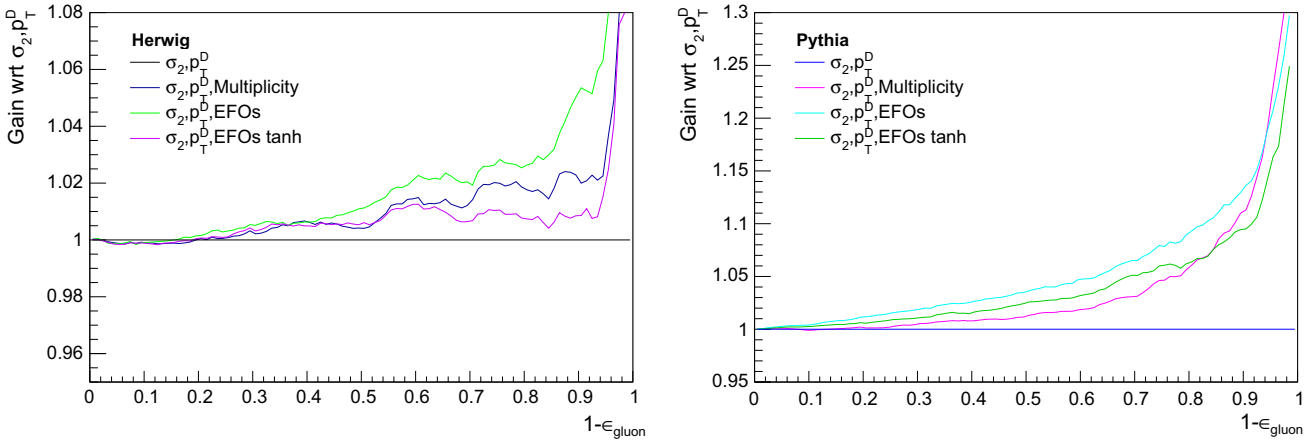


Fig. 6 Left: the relative gain for the three-variable taggers with respect to a baseline tagger using just $p_T D$ and σ_2, for the Herwig events (which yield more conservative discrimination estimates). The gain is also plotted for EFOs computed with the hyperbolic tangent fitting function specified in Sect. 2.1, for which the performance is worse. Right: for Pythia events. Note the wider range of the y-axis, to accommodate the larger gains found in Pythia

The right plot of Fig. 4 presents the linear correlation coefficients (calculated using the TMVA toolkit) between the EFOs and the existing CMS quark–gluon tagger variables: multiplicity, $p_T D$ and σ_2. We also include a computation of the fractal dimension, which has been calculated from a linear fit over a small range of box scales. Strong correlations are present amongst the EFOs, as is natural given they are parameters derived from the same fit. However, their correlations with the other variables are no greater than 43% (for either quarks or gluons).[4] Interestingly, the EFOs are most highly correlated with σ_2, not multiplicity as might have been expected. This evidence suggests the discrimination power of the EFOs is not simply a result of higher multiplicities in gluon jets, and therefore that the addition of these parameters to a quark–gluon discriminator might improve performance.

We find that replacing the multiplicity variable in the existing CMS quark–gluon tagger with the EFO variable yields a gain in discriminator performance, albeit only a modest one. This gain is seen using both Herwig and Pythia event generators (with the setup described above) in the ROCs presented in Fig. 5, which are for jets with $p_T \geq 100$ GeV. We see the performance in Pythia is significantly better than Herwig for each combination of variables, consistent with previous studies [9–11,14].

Moreover, the incremental gain upon replacing multiplicity with the EFOs is larger in Pythia than Herwig, so Herwig gives the more conservative estimate of the impact of including the EFOs. We see the gain in performance (relative to a baseline tagger using just $p_T D$ and σ_2) more clearly in Fig. 6, with the left panel for Herwig and the right for Pythia. The gain is at the level of 1– 2% in the more conservative

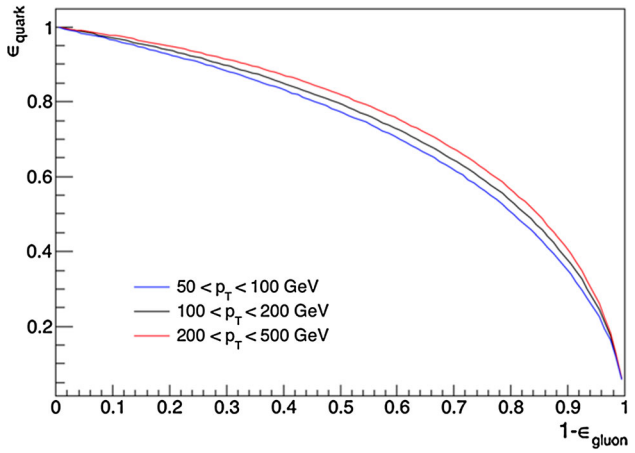


Fig. 7 Performance of a possible new quark–gluon tagger (using $p_T D$, σ_2, and the EFOs), in three p_T bins, for Herwig-generated dijet events. Quarks and gluons are found to be easier to distinguish using this tagger at higher p_T

Herwig setup, and slightly larger in Pythia (note the different scaling of the y-axis). To emphasize a previous point, these gains were found to be stable across different values of the soft p_T cut. Finally, we investigated how the performance varies with energy scale, by performing the analysis in p_T bins of 50–100, 100–200, and 200–500 GeV. Discrimination was found to increase with p_T in both Herwig and Pythia (see Fig. 7 for the Herwig results).

Combining all four variables (multiplicity, $p_T D$, σ_2 and the EFOs) was seen to give no further improvement. This suggests all the information from multiplicity is captured by the EFOs,[5] while the converse is not true. In summary,

[4] Note that the traditional fractal dimension is more strongly correlated to existing QGD variables, particularly multiplicity.

[5] This is unsurprising, because jet multiplicity is simply the asymptotic number of hits as we approach the saturation region.

we have presented evidence in this study that the Extended Fractal Observables provide an additional handle that captures the salient features of jet multiplicity, incorporates new information from showering and hadronization, and which is also better behaved under IRC emission (see 2.3).

4 Conclusions

In this study we defined new jet observables, the Extended Fractal Observables, by a generalization of the box-counting method used in the study of fractal systems. Defined with a pixel-level low momentum cutoff, these observables are infrared and collinear safe. We have then sought to apply the EFOs to improve quark–gluon discrimination. At the generator level, we find some modest improvement in discrimination by gluon rejection when we replace multiplicity with the EFOs in the existing CMS tagger, across both Herwig++ and Pythia 8. Extending the performance of these new variables to include detector effects can naturally be performed in the LHC environment with the CMS Particle Flow algorithm [45] in conjunction with the PUPPI algorithm [46] to reconstruct particle candidates in the presence of high pile-up.

5 Outlook

This method of studying jet substructure is a new approach. As such, there are many directions in which we would like to proceed, including:

1. Exploring particle hits in a 3-dimensional coordinate space spanned by η, ϕ and z^{-1}, where z is the fractional transverse momentum of the jet constituent.
2. Applying the EFOs beyond Quark–Gluon discrimination, for example to the identification of pile-up jets, or initial state radiation.
3. These box-counting methods extend very naturally from the substructure of a single jet to a whole-event analysis. Such a novel approach may provide new insight into searches for new physics topologies such as those in supersymmetry or top quark pair production [47].
4. Furthermore, box-counting analyses could provide a useful characterization of event shapes in heavy ion collisions, where studies of jet properties beyond jet reconstruction are traditionally difficult, but well motivated [48–50].
5. Finally, we would like to emphasize that the calculation of EFOs on quark and gluon jets probes parton shower scaling that results from the QCD color factor ratio. Calculating EFOs on cosmic ray air shower profiles [51] could therefore help discriminate QCD-induced air showers from more interesting signals; of particular interest, showers induced by electroweak sphalerons. Experimentally, the calculation of EFOs in this air shower context is conceptually appealing: the 1660 individual Cerenkov detectors (spread over 3000 km^2) of the Pierre Auger Observatory in Argentina [52] would naturally function as the finest-scale cells in our box-counting algorithm. These techniques could therefore be useful in probing physics at energies far beyond that of the LHC.

Acknowledgements JD's work has been supported by The Cambridge Trust, and by the STFC consolidated grant ST/L000385/1. We thank the CERN summer student program where this work was initiated. We also thank Andrew Larkoski for his insightful comments when performing these studies, and Bryan Webber for helpful discussions. Finally, we thank Eric Metodiev for helpful comments.

Funded by SCOAP[3].

References

1. G. Gustafson, A. Nilsson, Multifractal dimensions in QCD cascades. Z. Phys. C **52**, 533–542 (1991)
2. J.D. Bjorken, Fractal phase space as a diagnostic tool for high-energy multijet processes. Phys. Rev. D **45**, 4077–4087 (1992)
3. B. Andersson, P. Dahlkvist, G. Gustafson, An infrared stable multiplicity measure on QCD parton states. Phys. Lett. B **214**, 604–608 (1988)
4. A.J. Larkoski, QCD analysis of the scale-invariance of jets. Phys. Rev. D **86**, 054004 (2012). arXiv:1207.1437
5. M. Jankowiak, A.J. Larkoski, Angular scaling in jets. JHEP **04**, 039 (2012). arXiv:1201.2688
6. D.E. Soper, M. Spannowsky, Finding physics signals with shower deconstruction. Phys. Rev. D **84**, 074002 (2011). arXiv:1102.3480
7. Y.L. Dokshitzer, V.A. Khoze, S.I. Troyan, On the concept of local parton-hadron duality. J. Phys. G Nucl. Part. Phys. **17**(10), 1585 (1991)
8. M. Ruan, D. Jeans, V. Boudry, J.-C. Brient, H. Videau, Fractal dimension of particle showers measured in a highly granular calorimeter. Phys. Rev. Lett. **112**(1), 012001 (2014). arXiv:1312.7662
9. ATLAS Collaboration, G. Aad et al., Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Eur. Phys. J. C **74**(8), 3023 (2014), arXiv:1405.6583
10. CMS Collaboration Collaboration, Performance of quark/gluon discrimination in 8 TeV pp data, Tech. Rep. CMS-PAS-JME-13-002, CERN, Geneva, 2013
11. J.R. Andersen et al., Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report, in 9th Les Houches Workshop on Physics at TeV Colliders (PhysTeV 2015) Les Houches, France, June 1-19, 2015, 2016. arXiv:1605.04692
12. P. Gras, S. Hche, D. Kar, A. Larkoski, L. Lnnblad, S. Pltzer, A. Sidmok, P. Skands, G. Soyez, J. Thaler, Systematics of quark/gluon tagging. JHEP **07**, 091 (2017). arXiv:1704.03878
13. D. Ferreira de Lima, P. Petrov, D. Soper, M. Spannowsky, Quark-Gluon tagging with Shower Deconstruction: Unearthing dark mat-

ter and Higgs couplings. Phys. Rev. D **95**(3), 034001 (2017). arXiv:1607.06031
14. J. Gallicchio, M.D. Schwartz, Quark and gluon jet substructure. JHEP **04**, 090 (2013). arXiv:1211.7038
15. J. Gallicchio, M.D. Schwartz, Quark and gluon tagging at the LHC. Phys. Rev. Lett. **107**, 172001 (2011). arXiv:1106.3076
16. J. Gallicchio, M.D. Schwartz, Pure samples of quark and gluon jets at the LHC. JHEP **10**, 103 (2011). arXiv:1104.1175
17. DELPHI Collaboration, P. Abreu et al., The scale dependence of the hadron multiplicity in quark and gluon jets and a precise determination of C(A) / C(F). Phys. Lett. B **449**, 383–400 (1999), arXiv:hep-ex/9903073
18. CLEO Collaboration, R.A. Briere et al., Comparison of particle production in quark and gluon fragmentation at s**(1/2) 10-GeV. Phys. Rev. D **76**, 012005 (2007). arXiv:0704.2766
19. J. Pumplin, Quark-gluon jet differences at LEP. Phys. Rev. D **48**, 1112–1116 (1993). arXiv:hep-ph/9301215
20. M.H. Seymour, The subjet multiplicity in quark and gluon jets. Phys. Lett. B **378**, 279–286 (1996). arXiv:hep-ph/9603281
21. C. Kilic, S. Schumann, M. Son, Searching for multijet resonances at the LHC. JHEP **04**, 128 (2009). arXiv:0810.5542
22. B. Bhattacherjee, S. Mukhopadhyay, M.M. Nojiri, Y. Sakaki, B.R. Webber, Quark–gluon discrimination in the search for gluino pair production at the LHC. JHEP **01**, 044 (2017). arXiv:1609.08781
23. K. Joshi, A.D. Pilkington, M. Spannowsky, The dependency of boosted tagging algorithms on the event colour structure. Phys. Rev. D **86**, 114016 (2012). arXiv:1207.6066
24. CMS Collaboration Collaboration, V Tagging Observables and Correlations, Tech. Rep. CMS-PAS-JME-14-002, CERN, Geneva, (2014)
25. P.T. Komiske, E.M. Metodiev, M.D. Schwartz, Deep learning in color: towards automated quark/gluon jet discrimination. JHEP **01**, 110 (2017). arXiv:1612.01551
26. J. Cogan, M. Kagan, E. Strauss, A. Schwartzman, Jet-images: computer vision inspired techniques for jet tagging. JHEP **02**, 118 (2015). arXiv:1407.5675
27. L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, A. Schwartzman, Jet-images—deep learning edition. JHEP **07**, 069 (2016). arXiv:1511.05190
28. L.G. Almeida, M. Backović, M. Cliche, S.J. Lee, M. Perelstein, Playing tag with ANN: boosted top identification with pattern recognition. JHEP **07**, 086 (2015). arXiv:1501.05968
29. P. Baldi, K. Bauer, C. Eng, P. Sadowski, D. Whiteson, Jet substructure classification in high-energy physics with deep neural networks. Phys. Rev. D **93**(9), 094034 (2016). arXiv:1603.09349
30. D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, D. Whiteson, Jet flavor classification in high-energy physics with deep neural networks. Phys. Rev. D **94**(11), 112002 (2016). arXiv:1607.08633
31. Y.L. Dokshitzer, F. Fabbri, V.A. Khoze, W. Ochs, Multiplicity difference between heavy and light quark jets revisited. Eur. Phys. J. C **45**, 387–400 (2006). arXiv:hep-ph/0508074
32. P. Bolzoni, B.A. Kniehl, A.V. Kotikov, Average gluon and quark jet multiplicities at higher orders. Nucl. Phys. B **875**, 18–44 (2013). arXiv:1305.6017
33. P. Bolzoni, B.A. Kniehl, A.V. Kotikov, Gluon and quark jet multiplicities at $N^3LO+NNLL$. Phys. Rev. Lett. **109**, 242002 (2012). arXiv:1209.5914
34. ATLAS Collaboration, G. Aad et al., The ATLAS experiment at the CERN large Hadron Collider, *JINST***3** (2008) S08003
35. CMS Collaboration, S. Chatrchyan et al., The CMS experiment at the CERN LHC. *JINST***3**, S08004 (2008)
36. B. Bhattacherjee, S. Mukhopadhyay, M.M. Nojiri, Y. Sakaki, B.R. Webber, Associated jet and subjet rates in light-quark and gluon jet discrimination. JHEP **04**, 131 (2015). arXiv:1501.04794
37. Y.L. Dokshitzer, G.D. Leder, S. Moretti, B.R. Webber, Better jet clustering algorithms. JHEP **08**, 001 (1997). arXiv:hep-ph/9707323
38. A.J. Larkoski, J. Thaler, W.J. Waalewijn, Gaining (Mutual) information about quark/gluon discrimination. JHEP **11**, 129 (2014). arXiv:1408.3122
39. M. Bahr et al., Herwig++ physics and manual. Eur. Phys. J. C **58**, 639–707 (2008). arXiv:0803.0883
40. M.H. Seymour, A. Siodmok, Constraining MPI models using σ_{eff} and recent Tevatron and lhc underlying event data. JHEP **10**, 113 (2013). arXiv:1307.5015
41. CMS Collaboration, V. Khachatryan et al., Event generator tunes obtained from underlying event and multiparton scattering measurements. Eur. Phys. J. C **76**(3), 155 (2016). arXiv:1512.00815
42. M. Cacciari, G.P. Salam, G. Soyez, Fast jet user manual. Eur. Phys. J. C **72**, 1896 (2012). arXiv:1111.6097
43. OPAL Collaboration, G. Alexander et al., A Comparison of b and (uds) quark jets to gluon jets. Z. Phys. C **69**, 543–560 (1996)
44. A.J. Larkoski, G.P. Salam, J. Thaler, Energy correlation functions for jet substructure. JHEP **06**, 108 (2013). arXiv:1305.0007
45. CMS Collaboration Collaboration, Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET, Tech. Rep. CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr, (2009)
46. D. Bertolini, P. Harris, M. Low, N. Tran, Pileup per particle identification. JHEP **10**, 059 (2014). arXiv:1407.6013
47. D.E. Soper, M. Spannowsky, Finding physics signals with event deconstruction Phys. Rev. D **89**(9), 094005 (2014). arXiv:1402.1189
48. CMS Collaboration, S. Chatrchyan et al., Modification of jet shapes in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Phys. Lett. B **730**, 243–263 (2014). arXiv:1310.0878
49. CMS Collaboration Collaboration, Jet Fragmentation Function in pPb Collisions at sqrt(sNN) = 5.02 TeV and pp Collisions at sqrt(s) = 2.76 and 7 TeV, Tech. Rep. CMS-PAS-HIN-15-004, CERN, Geneva, (2015)
50. CMS Collaboration Collaboration, Splitting function in pp and PbPb collisions at 5.02 TeV, Tech. Rep. CMS-PAS-HIN-16-006, CERN, Geneva, (2016)
51. G. Brooijmans, P. Schichtel, M. Spannowsky, Cosmic ray air showers from sphalerons. Phys. Lett. B **761**, 213–218 (2016). arXiv:1602.00647
52. Pierre Auger Collaboration, A. Aab et al., The Pierre Auger cosmic ray observatory. Nucl. Instrum. Methods A **798**, 172–213 (2015). [arXiv:1502.01323]