

## Isolating color-singlet boson jets at the LHC using telescoping jet substructure

Yang-Ting Chien<sup>1,\*</sup>, Zachary Montague,<sup>2</sup> Alex Emerman,<sup>3</sup> Shih-Chieh Hsu,<sup>2</sup> and Samuel Meehan<sup>2</sup>

<sup>1</sup>*C.N. Yang Institute for Theoretical Physics, Stony Brook University, Stony Brook, New York 11794, USA*

<sup>2</sup>*Department of Physics, University of Washington, Seattle, Washington 98195, USA*

<sup>3</sup>*Department of Physics, Columbia University, New York, New York 10027, USA*

 (Received 25 December 2017; revised manuscript received 22 October 2019; accepted 26 May 2020; published 9 June 2020)

We demonstrate that jets from hadronic decays of moderately boosted color-singlet bosons at the LHC can be isolated. The telescoping jet substructure method is used to carve out dominant energy flows along subjects and identify the radiation surrounding the dipole, and it achieves excellent performance in identifying boosted  $W$  boson and top quark jets at the LHC energy. We show that telescoping jet variables efficiently isolate subjects, and we compare with standard  $N$ -prong taggers and isolation cone methods.

DOI: [10.1103/PhysRevD.101.114006](https://doi.org/10.1103/PhysRevD.101.114006)

The Large Hadron Collider (LHC) has been probing the physics above the electroweak scale, and the energies of produced massive Standard Model particles such as the weak bosons or top quarks can be much larger than their invariant masses. This results in hadronic jets with substructure of multiple prongs, and a variety of useful variables have been designed and combined in multivariate analysis [1–7] in order to identify such jets. Color-flow variables [8–12] which probe the color connections among energy flow directions have also been studied. Additional discrimination power can be gained by including the variable *volatility*  $v^{\mathcal{O}}$  [13–15] of a jet observable  $\mathcal{O}$  (e.g., jet mass  $\mathcal{M}$ ) with a parameter  $a$  (e.g., jet radius  $R$ ). It is defined as the ratio of the standard deviation and the mean of the set of values  $\{\mathcal{O}_{a_i}\}$  from a sampling of telescoping parameter  $\{a_i\}$ ,

$$v^{\mathcal{O}} = \frac{\sigma(\mathcal{O})}{\langle \mathcal{O} \rangle}. \quad (1)$$

The volatility  $v^{\mathcal{O}}$  measures the change of the observable  $\mathcal{O}$  with respect to the change of the phase-space boundary set by the parameter  $a$ . It gives a trend of the observable variation which itself can be used as a distinguishing feature to classify jets.

It is well known that, in the boosted regime the decay products of a color-singlet heavy particle are localized

within an angle of  $\theta_0 \sim \mathcal{O}(2M/p_T)$  along the particle direction where  $M$  and  $p_T$  are the mass and transverse momentum of the particle. Therefore these particles can be straightforwardly isolated within a cone of radius  $R_{\text{iso}} \gg \theta_0$ . This forms the basis of the isolation of hadronic  $\tau$  jets [16,17]<sup>1</sup> and extremely boosted weak boson jets at the proposed 100 TeV future circular collider [18,19] where  $\theta_0 \sim 0.02$ . Note that even in the extremely boosted case the use of a number of telescoping radii increases the performance significantly [19], which was pioneered and discovered in [15,20,21].

In this paper, we show that hadronic jets from decays of even only moderately boosted, color-singlet bosons at the LHC (i.e.,  $\theta_0 \sim 0.2$ – $0.4$ ) can be isolated from the rest of the event using telescoping jets and volatility. In order to demonstrate that hadrons from boson decays are isolated, one needs to separate such radiation from event-wide soft radiation contributions such as initial-state radiation, final-state radiation from other jets, underlying events and pileup. Since the volatility probes the variation of the jet radiation pattern surrounding dominant energy flows along subjects with multiple angular resolutions, it is sensitive to the isolation effect and the total color charge of the jet-initiating boson.

We will show that the volatility exhibits intrinsic properties of boson isolation. Indeed, boson isolation does not require resolving the two-prong structure [19]; therefore it is a new class of jet observable distinct from two-prong taggers. We will check this feature using “pseudocalorimeter cells” that the tagging performance of the volatility is robust against the finite angular segmentation existing in

\*yang-ting.chien@stonybrook.edu

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

<sup>1</sup>Since  $m_\tau \approx 1.8$  GeV is too small for extra QCD radiation to be emitted, the few hadronic  $\tau$  decay products are typically extremely boosted and can be easily isolated.

realistic detectors, which smears out jet substructure beyond this angular resolution. While detailed studies of detector effects and the use of tracks with higher angular resolution are practical issues, these are beyond the scope of this paper.

We focus on the jet mass volatility  $v^{\mathcal{M}}$  (or simply denoted as  $v$ ) as defined in the *telescoping subjets* procedure.<sup>2</sup> A number of  $N$  subjets of radius  $R_T$  are reconstructed within the Voronoi regions [20,22–24] of a jet partitioned by dominant energy flow directions. We choose the subjet axes to be the  $N$ -subjettiness axes [25] with  $\beta = 1$ . Particles are assigned to the nearest axis according to the distance  $\Delta R_{ij}$  between the axis  $\hat{n}_i$  and particle  $p_j$ ,

$$\text{subj}_i = \{p_j | \Delta R_{ij} < R_T \text{ and } \Delta R_{ij} < \Delta R_{kj}, \forall k \neq i\}, \quad (2)$$

where  $k$  is the index of the other axes  $\hat{n}_k$ . The jet mass  $\mathcal{M}_{R_T}$  is then defined as the invariant mass of the sum of  $N$  subjets of radius  $R_T$ , and the volatility  $v_N$  is defined with a sampling of the telescoping parameter  $a = R_T \in (0.1, 1.0) \times R$  uniformly within the range. Note that we scan through the entire catchment area of the jet. We focus on  $N = 2$  and 3 for  $W$  tagging, and  $N = 2, 3$ , and 4 for top tagging. However,  $N$  could be chosen differently for different boosted topologies.

The ability to reconstruct the volatility accurately in experiment may be obscured by the presence of underlying events and pileup. However, these contributions can be efficiently mitigated through a grooming procedure [26–29] or background subtraction technique [30–33]. Photon isolation has also been shown to be successful in heavy ion collisions after mitigating an enormous underlying event background [34]. In this paper we focus on analyzing the intrinsic boson isolation feature at the LHC, and the jets in our analysis are groomed using the trimming algorithm with  $R_{\text{sub}} = 0.3$  and  $f_{\text{cut}} = 0.05$ . We leave the detailed study of the impact of underlying events and pileup for future work.

Note that the generality of the telescoping algorithm allows a variety of other applications which are beyond the scope of this paper. Nevertheless, in *telescoping pruning* one can fix  $z_{\text{cut}}$  and construct  $v_{\text{prun}}$ , the volatility of the pruned jet mass with the telescoping parameter  $a$  in  $D_{\text{cut}} = a 2m_{\text{jet}}/p_{T_{\text{jet}}}$ . In *telescoping trimming*, one can fix the subjet radius  $R_{\text{sub}}$  and construct  $v_{\text{trim}}$ , the volatility of the trimmed jet mass with the telescoping parameter  $a = f_{\text{cut}}$ . One can also construct  $v_{\tau_N}$ , the volatility of the  $N$ -subjettiness with the telescoping parameter  $a = \beta$  [35]. Correlations among the observable variations with respect to multiple varied parameters can also be explored.

Our analysis is performed using samples generated from Monte Carlo simulations of proton-proton collisions at  $\sqrt{s} = 13$  TeV using PYTHIA8 [36]. Particles are clustered

<sup>2</sup>A similar method using the leading subjets in a reclustered jet was explored in [21].

into jets with FASTJET3 [37] using the anti- $k_T$  algorithm [38] with  $R = 1.0$ , and the jets are required to be central with a pseudorapidity  $|\eta| < 1.2$ . We consider two kinematic regimes where the jet  $p_T$  is either between 350 GeV and 500 GeV or 800 GeV and 1 TeV. Signal  $W$  boson and top quark jets are generated using decays of heavy Kaluza-Klein gravitons with invariant masses at 1 or 2 TeV for the two  $p_T$  bins in fully hadronic  $G^* \rightarrow W^+W^-$  and  $G^* \rightarrow t\bar{t}$  processes. Background QCD jets are generated from the Standard Model inclusive jet processes. To quantify the impact of finite angular segmentation, we compare with the results using the “cell particles” clustered in “pseudocalorimeter cells” of size  $(\eta, \phi)$ :  $0.1 \times 0.1$ , with each cell particle momentum constructed from the total energy deposited in the cell with zero mass and direction from the primary vertex. Note that the  $N$ -prong structure with angles among the prongs smaller than 0.1 is smeared away. A selection on the trimmed jet mass is made between 70 GeV and 90 GeV for  $W$  tagging and between 160 GeV and 190 GeV for top tagging.

For top tagging, in order to exploit the three-prong structure of top jets we also consider the ratio  $v_{N2}$  between  $v_N$  and  $v_2$  for  $N = 3, 4$  inspired by the  $N$ -subjettiness ratios,

$$v_{N2} = \frac{v_N}{v_2}. \quad (3)$$

Shown in Fig. 1 are the distributions of  $v_2$ ,  $v_3$ , and  $v_{32}$  for top and QCD jets. We find that top jets have a broader  $v_2$  distribution and a narrower  $v_3$  distribution compared to those of QCD jets. The large variation of the top jet mass when telescoping around two subjet axes is mostly caused by the partial reconstruction of the  $W$  boson. There is not an intrinsic mass scale dictating the hard emissions in QCD jets. On the other hand, the three prongs of top jets are quark-initiated subjets, two of them coming from the decay of the color-singlet  $W$  which is isolated, whereas the subjets in QCD jets can be initiated by gluons. Since the color charge of quarks is smaller than that of gluons, quark subjets are narrower than gluon subjets and the  $v_3$  of top jets tend to be smaller. While intuitively one expects that the ratio  $v_3/v_2$  improves the tagging performance, interestingly we find that  $v_{32}$  achieves a very similar performance as the combination of  $\{v_2, v_3\}$  using boosted decision trees (BDTs) implemented in TMVA [39]. This suggests that the  $v_{32}$  ratio approaches the optimal performance of the two volatilities.

Note an interesting feature of  $v_{32}$ : it cuts off naturally at 1, most clearly seen in QCD jets. Crucially,  $v_3 \leq v_2$  in the collinear limit. The two-prong structure in QCD jets implies that  $v_2$  and  $v_3$  contain similar information. The third energy flow axis can not locate far away from the two axes determined at  $N = 2$ . Hence, limited information is extracted from constructing a third subjet, and the distribution of  $v_{32}$  for QCD jets peaks at 1. In the case where

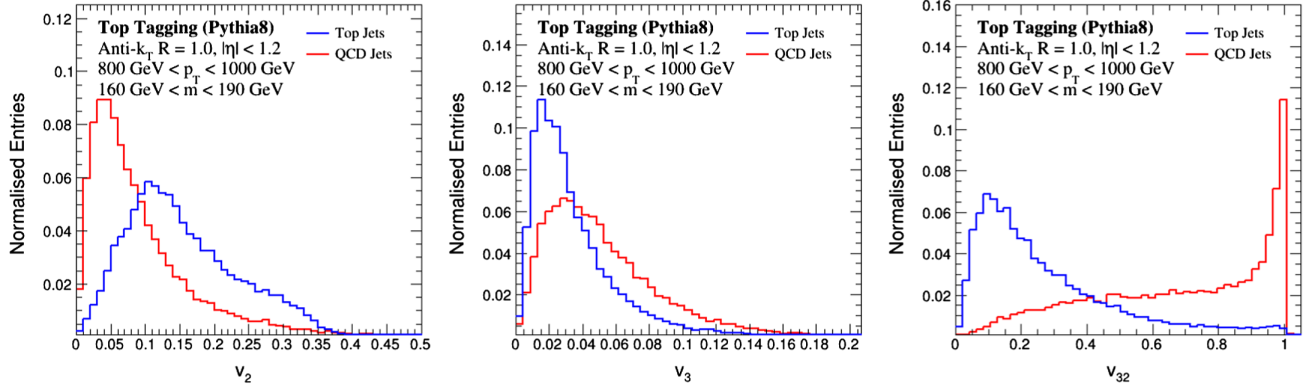


FIG. 1. The distributions of the volatilities  $v_2$  (left panel) and  $v_3$  (middle panel), as well as their ratio  $v_{32}$  (right panel) for top and QCD jets with  $800 \text{ GeV} < p_T < 1 \text{ TeV}$  and  $160 \text{ GeV} < m < 190 \text{ GeV}$  using the truth-particle information.

there is a third, semihard emission, such emission is captured by the telescoping subjets at  $N = 3$ ; therefore  $v_3 < v_2$ . In general, for larger  $N$  more particles are by-default captured, so the volatility is expected to decrease ( $v_{N+1} \leq v_N$ ).

The background efficiency as a function of the signal efficiency is illustrated by receiver operating characteristic (ROC) curves, where a lower curve indicates a better tagging performance. Shown in Fig. 2 are the ROC curves of  $v_3$  and a standard two-prong tagger such as the  $N$ -subjettiness ratio  $\tau_{21} = \tau_2/\tau_1$  for  $W$  tagging as a comparison. The upper and lower panels correspond respectively to two jet  $p_T$  regions of (350 GeV, 500 GeV) and (800 GeV, 1 TeV). Overall, the tagging performance increases at higher  $p_T$ , demonstrating the general advantage of applying telescoping jets to the boosted regime. In the lower panel, we compare the tagging performance using truth particles and pseudocalorimeter clusters, which degrade information about structures smaller than the cell size. We find excellent performance of  $v_3$ . Also, the  $v_3$  observable is robust against angular segmentation especially at high  $p_T$ . The telescoping procedure in the  $v_3$  observable utilizes the rapid depletion of radiation around the  $W$  at larger angles in the boosted regime. This  $W$  isolation effect is the manifestation of the fact that the  $W$  is a color singlet which affects the color structure of the subjets and the radiation pattern at large angles. The time dilation that occurs before  $W$  hadronically decays can also result in a period of time, about  $\mathcal{O}(1) \text{ fm}/c$  [40], during which no QCD radiation can be emitted. There is not such a delay in the jet formation process for QCD jets. Note that the observable  $v_3$  captures any significant third, semihard emission in  $W$  and QCD jets; therefore it can better quantify the isolation of  $W$  jets.

Shown in Fig. 3 are the ROC curves for top tagging performance including  $v_{42}$  and a standard three-prong tagger of the  $N$ -subjettiness ratio  $\tau_{32} = \tau_3/\tau_2$  as a comparison. Again, the upper and lower panels correspond to the two kinematic regimes  $p_T \in (350 \text{ GeV}, 500 \text{ GeV})$  and  $p_T \in (800 \text{ GeV}, 1 \text{ TeV})$ , and we note that the tagging performance increases at higher  $p_T$ . In the lower panel, the

ROC curves plot both truth-particle and pseudocalorimeter information. We find the excellent performance of  $v_{42}$  and its robustness against angular segmentation especially at high  $p_T$ . The  $v_{42}$  observable can capture any significant fourth, semihard emission, and it seems to exploit the  $W$

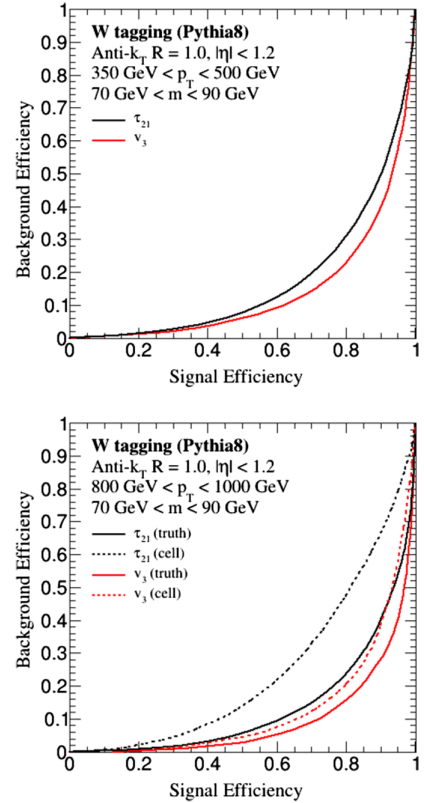


FIG. 2. The  $W$  tagging ROC curves of the volatility  $v_3$  and the two-prong tagger  $\tau_{21} = \tau_2/\tau_1$  in the (300 GeV, 500 GeV) jet  $p_T$  bin (upper panel) and the (800 GeV, 1 TeV) bin (lower panel). Solid curves correspond to the ones with the truth-particle information, and the dashed curves are the ones using the pseudocalorimeter cell particle information.

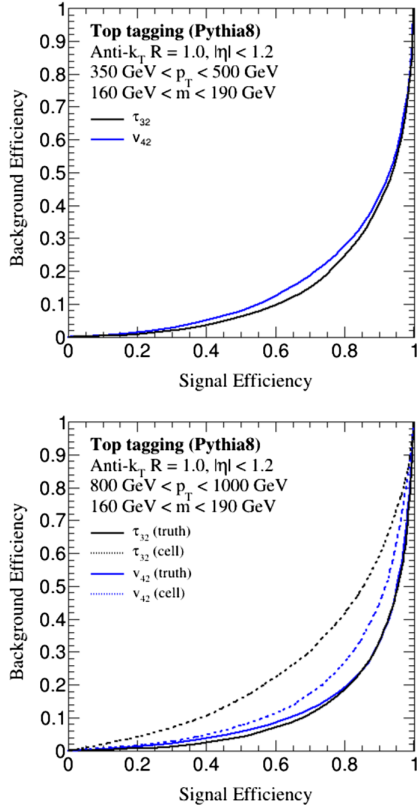


FIG. 3. The top tagging ROC curves of the volatility ratio  $v_{42}$  and the three-prong tagger  $\tau_{32} = \tau_3/\tau_2$  in the (300 GeV, 500 GeV) jet  $p_T$  bin (upper panel) and the (800 GeV, 1 TeV) bin (lower panel). Solid curves correspond to the ones with the truth-particle information, and the dashed curves are the ones using the pseudocalorimeter cell particle information.

isolation within the top jet in the boosted regime since  $v_{42}$  exhibits corresponding properties as the  $v_3$  in  $W$  tagging.

While the insensitivity of volatility to  $N$ -prong structure suggests that the volatility carries new jet substructure information, in Fig. 4 we show the scatter plots of  $v_3$  and  $\tau_{21}$  of  $W$  (upper panel) and QCD (lower panel) jets in the (350 GeV, 500 GeV) jet  $p_T$  region. We can see that  $v_3$  is quite correlated with  $\tau_{21}$ . However, in the higher  $p_T$  region of (800 GeV, 1 TeV) (upper panel of Fig. 5) where the isolation effect is more prominent, the correlation between  $v_3$  and  $\tau_{21}$  for  $W$  jets decreases (with correlation coefficient 0.463). We also check that the combination of  $v_3$  and  $\tau_{21}$  using BDTs significantly increases the performance. In contrast, the lower panel of Fig. 5 shows that the 2-prong taggers  $D_2$  [41] and  $\tau_{21}$  are significantly correlated (with correlation coefficient 0.721). On the other hand, while isolation cone variables have been successful for identifying localized objects, the volatility variable is able to exploit isolation of an object with substructure.

To conclude, we have demonstrated that the isolation of color-singlet boson jets is possible at the LHC. It is an important feature which, to the best of our knowledge, has not been exploited in jet substructure tools used at the LHC

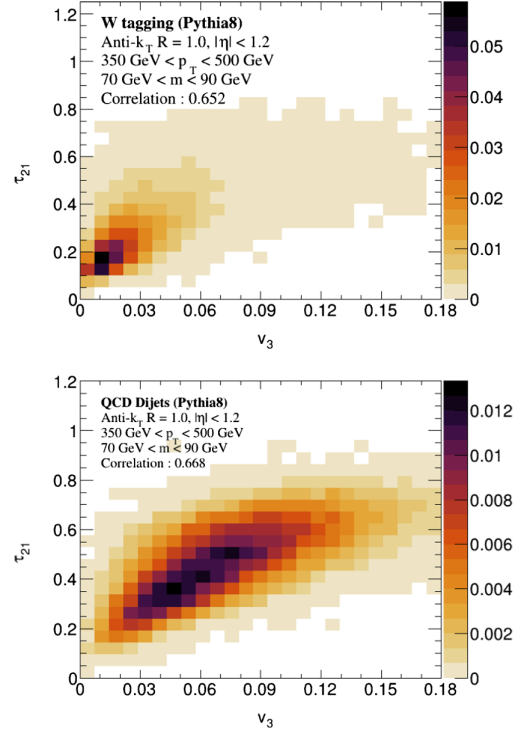


FIG. 4. The scatter plots of  $v_3$  and  $\tau_{21}$  of  $W$  (upper panel) and QCD (lower panel) jets in the (350 GeV, 500 GeV) jet  $p_T$  region.

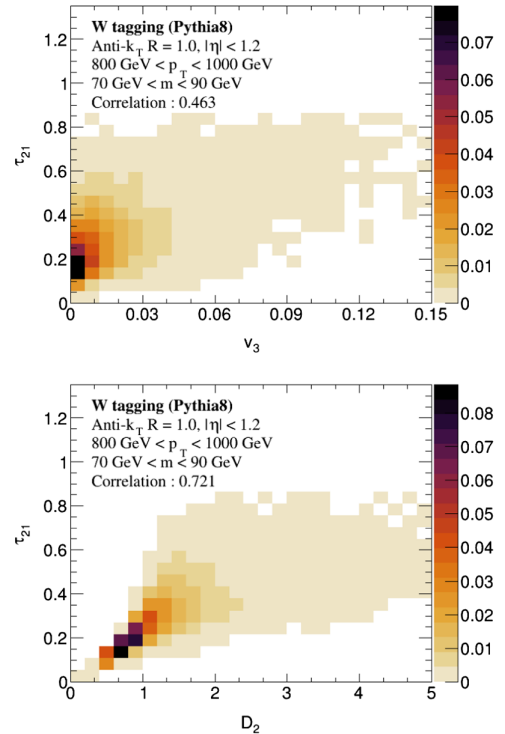


FIG. 5. The scatter plots of  $(v_3, \tau_{21})$  (upper panel) and  $(D_2, \tau_{21})$  (lower panel) of  $W$  jets in the (800 GeV, 1 TeV) jet  $p_T$  region.

although it has been known to be even more important at future, higher energy colliders. Such isolation feature is qualitatively different from the  $N$ -prong hard jet substructure and roots from the color coherence of the jet-initiating dipole. We have demonstrated that boson isolation indeed becomes more prominent in the boosted regime and is robust against finite angular segmentation via a comparison of the performance between reconstruction from using truth particles and from pseudocalorimeter cell particles. These are the basic properties of isolation and serve as consistency checks of the nature of the isolation feature. In contrast, the performance of typical  $N$ -prong taggers such as  $N$ -subjettiness degrades dramatically because of detector smearing as one would expect.

We would like to emphasize that, in this paper, boson isolation at the LHC is exposed through the use of the telescoping jet substructure method which systematically organizes jet information along energy flow directions with multiple angular scales [42]. Specifically, we find excellent performance of telescoping subjects using the volatility observables  $v_3$  in  $W$  tagging and  $v_{42}$  in top tagging. This promising tagging strategy is expected to work also in the contexts with the production of heavy, color-singlet

Standard Model particles including the  $W$ ,  $Z$  and the Higgs boson at the LHC.

Y.-T. C. would like to thank the organizers of the BOOST2015 conference where telescoping jet substructure was first presented. The authors thank Steve Ellis, Phillip Harris, Benjamin Nachman, Matthew Schwartz, Jesse Thaler and Nhan Tran for useful comments and suggestions. Y.-T. C. was supported by the US Department of Energy (DOE), Office of Science under Contract No. DE-AC52-06NA25396, the DOE Early Career Program, the LHC Theory Initiative Postdoctoral Fellowship under the National Science Foundation Grant No. PHY-1419008, and the National Science Foundation Grant No. PHY-1915093. A. Emerman was supported by the National Science Foundation under Grant No. PHY-1707971. S.-C. H. and S. M. were supported by the DOE Office of Science, Office of High Energy Physics Early Career Research program under Award No. DE-SC0015971. Z. M. was supported by the University of Washington's Ernest M. Henley & Elaine D. Henley Endowed Fellowship. Y.-T. C. and A. E. thank the support from Los Alamos National Laboratory where part of this work was performed.

- 
- [1] A. Abdesselam *et al.*, *Eur. Phys. J. C* **71**, 1661 (2011).
  - [2] A. Altheimer *et al.*, *J. Phys. G* **39**, 063001 (2012).
  - [3] A. Altheimer *et al.*, *Eur. Phys. J. C* **74**, 2792 (2014).
  - [4] D. Adams *et al.*, *Eur. Phys. J. C* **75**, 409 (2015).
  - [5] A. J. Larkoski, I. Moult, and B. Nachman, *Phys. Rep.* **841**, 1 (2020).
  - [6] M. Aaboud *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **79**, 375 (2019).
  - [7] C. Collaboration (CMS Collaboration), *J. High Energy Phys.* **12** (2014) 017; **12** (2014) 043; Comments: Replaced with published version. Added journal reference and DOI.
  - [8] J. Gallicchio and M. D. Schwartz, *Phys. Rev. Lett.* **105**, 022001 (2010).
  - [9] A. Hook, M. Jankowiak, and J. G. Wacker, *J. High Energy Phys.* **04** (2012) 007.
  - [10] L. Mackey, B. Nachman, A. Schwartzman, and C. Stansbury, *J. High Energy Phys.* **06** (2016) 010.
  - [11] G. Aad *et al.* (ATLAS Collaboration), *Phys. Lett. B* **750**, 475 (2015).
  - [12] M. Aaboud *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **78**, 847 (2018).
  - [13] S. D. Ellis, A. Hornig, T. S. Roy, D. Krohn, and M. D. Schwartz, *Phys. Rev. Lett.* **108**, 182003 (2012).
  - [14] D. Kahawala, D. Krohn, and M. D. Schwartz, *J. High Energy Phys.* **06** (2013) 006.
  - [15] Y.-T. Chien, D. Farhi, D. Krohn, A. Marantan, D. Lopez Mateos, and M. Schwartz, *J. High Energy Phys.* **12** (2014) 140.
  - [16] A. Airapetian *et al.* (ATLAS Collaboration), Reports No. CERN-LHCC-99-15 and No. ATLAS-TDR-15, 1999.
  - [17] G. L. Bayatian *et al.* (CMS Collaboration), *J. Phys. G* **34**, 995 (2007).
  - [18] M. L. Mangano *et al.*, *CERN Yellow Rep.* **3**, 1 (2017).
  - [19] C. Helsens, D. Jamin, M. L. Mangano, T. G. Rizzo, and M. Selvaggi, *Eur. Phys. J. C* **79**, 569 (2019).
  - [20] Y.-T. Chien, *Phys. Rev. D* **90**, 054008 (2014).
  - [21] Y. Cui, Z. Han, and M. D. Schwartz, *Phys. Rev. D* **83**, 074023 (2011).
  - [22] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, *Phys. Rev. Lett.* **105**, 092002 (2010).
  - [23] I. W. Stewart, F. J. Tackmann, J. Thaler, C. K. Vermilion, and T. F. Wilkason, *J. High Energy Phys.* **11** (2015) 072.
  - [24] J. Thaler and T. F. Wilkason, *J. High Energy Phys.* **12** (2015) 051.
  - [25] J. Thaler and K. Van Tilburg, *J. High Energy Phys.* **02** (2012) 093.
  - [26] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Phys. Rev. D* **80**, 051501 (2009).
  - [27] D. Krohn, J. Thaler, and L.-T. Wang, *J. High Energy Phys.* **02** (2010) 084.
  - [28] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, *J. High Energy Phys.* **09** (2013) 029.
  - [29] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *J. High Energy Phys.* **05** (2014) 146.
  - [30] D. Krohn, M. D. Schwartz, M. Low, and L.-T. Wang, *Phys. Rev. D* **90**, 065020 (2014).
  - [31] P. Berta, M. Spusta, D. W. Miller, and R. Leitner, *J. High Energy Phys.* **06** (2014) 092.

- [32] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **75**, 59 (2015).
- [33] D. Bertolini, P. Harris, M. Low, and N. Tran, *J. High Energy Phys.* **10** (2014) 059.
- [34] S. Chatrchyan *et al.* (CMS Collaboration), *Phys. Lett. B* **710**, 256 (2012).
- [35] J. Thaler and K. Van Tilburg, *J. High Energy Phys.* **03** (2011) 015.
- [36] T. Sjostrand, S. Mrenna, and P. Z. Skands, *Comput. Phys. Commun.* **178**, 852 (2008).
- [37] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **72**, 1896 (2012).
- [38] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
- [39] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *Proc. Sci.*, ACAT2007 (2007) 040 [arXiv:physics/0703039].
- [40] L. Apolinario, J. G. Milhano, G. P. Salam, and C. A. Salgado, *Phys. Rev. Lett.* **120**, 232301 (2018).
- [41] A. J. Larkoski, I. Moul, and D. Neill, *J. High Energy Phys.* **12** (2014) 009.
- [42] Y.-T. Chien and R. Kunnawalkam Elayavalli, arXiv: 1803.03589.