


# Supervised jet clustering with graph neural networks for Lorentz boosted bosons

Xiangyang Ju<sup>\*</sup> and Benjamin Nachman<sup>†</sup>

*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

 (Received 25 August 2020; accepted 23 September 2020; published 13 October 2020)

Jet clustering is traditionally an unsupervised learning task because there is no unique way to associate hadronic final states with the quark and gluon degrees of freedom that generated them. However, for uncolored particles like  $W$ ,  $Z$ , and Higgs bosons, it is possible to approximately (though not exactly) associate final state hadrons to their ancestor. By labeling simulated final state hadrons as descending from an uncolored particle, it is possible to train a supervised learning method to create boson jets. Such a method would operate on individual particles and identify connections between particles originating from the same uncolored particle. Graph neural networks are well-suited for this purpose as they can act on unordered sets and naturally create strong connections between particles with the same label. These networks are used to train a supervised jet clustering algorithm. The kinematic properties of these graph jets better match the properties of simulated Lorentz-boosted  $W$  bosons. Furthermore, the graph jets contain more information for discriminating  $W$  jets from generic quark jets. This work marks the beginning of a new exploration in jet physics to use machine learning to optimize the construction of jets and not only the observables computed from jet constituents.

DOI: [10.1103/PhysRevD.102.075014](https://doi.org/10.1103/PhysRevD.102.075014)

## I. INTRODUCTION

Lorentz-boosted massive bosons are a common feature of theories that extend the Standard Model (SM) of particle physics. In particular, new heavy particles introduced to solve one of the challenges with the SM may predominately decay into bosons and if there is a large mass hierarchy between the heavy particle and the bosons, the latter will be produced in the lab frame with a significant Lorentz boost. Singly produced bosons can also have significant Lorentz boost when produced in association with initial state radiation. The ATLAS and CMS collaborations have performed extensive searches involving boosted bosons decaying hadronically in the  $VV$  [1–4],  $Vh$  [5,6],  $hh$  [7–9],  $VX$  [10],  $Xh$  [11],  $XY$  [12], single- $X$  [13–17], and single- $h$  [18,19] channels, where  $V \in \{W^\pm, Z\}$ ,  $h$  is the SM Higgs boson, and  $X/Y$  are beyond the SM bosons.

A variety of jet substructure techniques have been developed to enhance Lorentz boosted boson tagging [20–27]. These methods range from physically motivated features such as groomed jet mass [28–32],  $N$ -subjettiness [33,34]

and  $D_2$  [35] to complex observables built using machine learning [21]. ATLAS and CMS have integrated and extended these methods as well as studied them using collision data [36–40]. One feature that all of these algorithms have in common is that they start from a collection of constituents selected using a jet clustering algorithm. Various studies have investigated optimizing the jet clustering algorithm by considering many options [41–43]. While important for converging on a method in the traditional paradigm, these approaches are fundamentally limited by the discreteness of the algorithm types and the flexibility offered by the tunable parameters of a given algorithm.

The most common approach for forming the initial Lorentz boosted boson candidate jets is the anti- $k_r$  algorithm [44]. This algorithm is a form of *unsupervised* learning because no per-particle labels are used to form the jets.<sup>1</sup> Instead, a distance measure motivated by the fragmentation of quarks and gluons is used to collect constituents that were likely produced from the same initiating high-energy quark or gluon. This last sentence does not have a precise meaning because quark and gluon jets are not well-defined objects [47,48]. Due to the strength of the strong force, the energy flows from outgoing quarks and gluons are interconnected with each other and with the beam remnants. In contrast, the quarks and gluons from color singlet massive bosons are

<sup>\*</sup>xju@lbl.gov

<sup>†</sup>bpnachman@lbl.gov

*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>Previous attempts at combining jet finding with *unsupervised* machine learning have been studied in the past [45,46], but do not have the benefits of the supervised approaches discussed here.

isolated from the rest of the event. In the limit that the number of colors  $N_c \rightarrow \infty$  or the width of the boson resonance  $\Gamma \rightarrow 0$ , there is a unique mapping between final state hadrons and ancestor color singlet. The corrections to this picture are suppressed by at least  $(1/N_c)^2$  (“color reconnection”) and by powers of  $\Gamma/\Lambda_{\text{QCD}}$ .

Given the approximate (but not exact) mapping between hadrons and color singlets, it makes sense to ask if one could construct a *supervised* approach to forming jets. In particular, a machine could be trained to label individual particles as originating from a color singlet or not based on the particle kinematic properties as well as the relationship with other particles in the event. While such an approach may give up the calculability afforded by algorithms like anti- $k_r$ , it may provide an optimal approach to constructing jets for searches where calculability is not necessarily required. If the jets are constructed optimally, then their substructure should contain as much information as possible for identifying their origin. One could even co-optimize the jet construction and the jet classification in an end-to-end approach [49,50], but there are many benefits to first building jets, such as the jet energy calibration.

Modern machine learning has proven to be a powerful toolkit for jet substructure. For example, a wide range of architectures and applications have been studied for tagging the origin of jets [50–103]. To construct a supervised jet clustering algorithm, a machine learning architecture is needed that can process variable length sets as input. Multiple such *point cloud* methods have been studied for jet substructure [70,72,73,78,79,104], but the structure chosen here is the graph neural network (GNN) (see Refs. [70,72,73,79,104–109]). This is because GNNs not only can process variable length sets, but they can also label the relationship between elements (not unique to GNNs, but natural given their construction). This property is critical for labeling particles as originating from the color singlet ancestor or not. Labeling constituents is also known as *semantic segmentation* and has been studied for other tasks in high energy physics ranging from pileup particle identification [72,110] to liquid argon time projection chamber labeling [111,112]. In addition, a recent study [113] shows that GNNs can be executed with a latency of less than 1  $\mu\text{s}$  on an field-programmable gate arrays, making such networks very promising for real-time data learning and filtering.

This paper is organized as follows. Section II introduces the simulated samples used to train the supervised jet clustering algorithm, where Lorentz-boosted  $W$  bosons provide a reoccurring example. The graph neural network methods are described in Sec. III and numerical results are presented in Sec. IV. The paper ends with outlook and conclusions in Sec. V.

## II. SIMULATION

Proton-proton collisions are simulated with PYTHIA8.183 [114,115] at a center-of-mass-energy of  $\sqrt{s} = 13$  TeV.

Lorentz boosted  $W$  bosons are generated from the decay of a hypothetical  $W'$  boson with a mass of 600 GeV that decays 100% of the time to a  $W$  boson and a  $Z$  boson. The  $W$  boson is forced to decay hadronically and the  $Z$  boson decays into neutrinos. To simulate a quark jet with nearly the same kinematic properties, a hypothetical excited quark  $q^*$  with a mass of 600 GeV is generated and decays 100% of the time into a quark and a  $Z$  boson. This  $Z$  boson then is forced to decay into neutrinos. The widths of the  $W'$ ,  $q^*$ , and  $W$  boson are set to 0.01 GeV. In total, 100,000  $W'$  and  $q^*$  events were generated.

As a leading  $N_c$  generator such as PYTHIA, it is possible to uniquely trace final state hadrons to the  $W$  boson. Individual final state hadrons are then labeled based on the existence (or not) of a real  $W$  boson in their ancestry from the event record. This is illustrated for one event in Fig. 1.

To compare with the graph neural network-based clustering scheme described in the next section, jets are clustered using the anti- $k_r$  algorithm [44] with radius parameter  $R = 1.0$  implemented in FASTJET3.0.3 [116,117]. Jets are only kept if they have  $p_T > 100$  GeV. These jets are subsequently trimmed [30] by keeping only  $R = 0.2$  subjets with at least 5% of the ungroomed jet’s transverse momentum. Trimming is not the only jet grooming algorithm [28–32], but it is widely used (see, e.g., Refs. [41,42]).

Figure 2 presents histograms of basic quantities in  $W$  events. The number of detector-stable particles with a  $W$  ancestor is about the same as the number of constituents inside the leading jet clustered by the anti- $k_r$  algorithm, however, it only accounts for about 10% of the total number of detector-stable particles in the event. The mass computed from the detector-stable particles originating from a  $W$  boson is nearly exactly  $m_W$  while leading jet mass is peaked around  $m_W$  with a broad width. On the other hand, there are many non- $W$  particles in the event, giving rise to an event mass far from the  $W$  boson mass. Therefore, it is nontrivial for a machine to find the  $W$  decay products in order to reconstruct the  $W$  boson mass. In the leading jet case, the low-mass peak corresponds to cases where both quarks from the  $W$  decay are not mostly contained within the leading jet or the leading jet is unrelated to the quarks from the  $W$  decay. Figure 3 shows that the kinematic properties of the jets in  $W'$  and  $q^*$  events are similar. The jet transverse momentum spectra are not identical because the radiation pattern outside of the jet cone is different for the color singlet  $W$  and color triplet quarks.

## III. GRAPH NEURAL NETWORK METHODS

A graph contains a set of nodes, a set of edges with each connecting a pair of nodes, and a set of node-, edge- and graph-level attributes, collectively called graph attributes. Graph neural networks (GNN) are trainable functions that operate on a graph to learn latent graph attributes as well as to form a parametrized message-passing by which

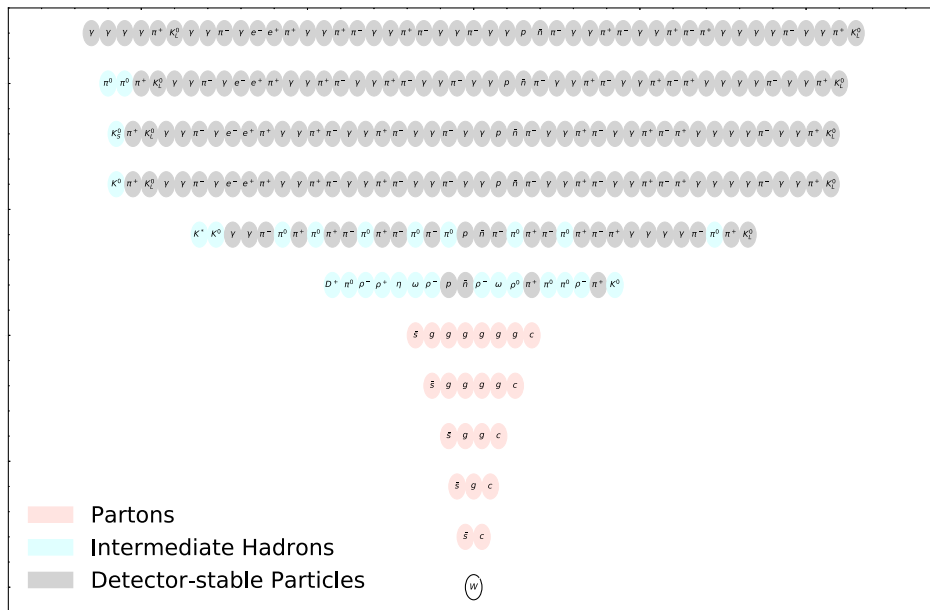


FIG. 1. An illustration of the  $W \rightarrow c\bar{s}$  decay tracing for a single event. At each step, every nondetector-stable particle is replaced with their immediate descendants from the PYTHIA event record. The order per row is arbitrary.

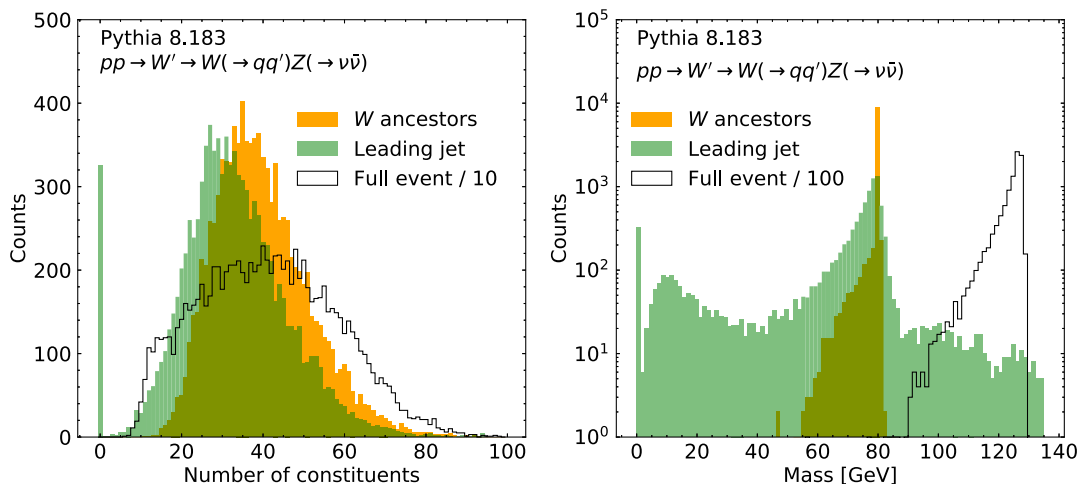


FIG. 2. Left: a histogram of the number of detector-stable particles originating from the  $W$  boson, inside the leading jet, and in the full event from the  $W'$  production. The leading jet is constructed from the anti- $k_T$  algorithm. The spike at 0 corresponds to events with no jet with  $p_T > 100$  GeV. For the full event, the number of constituents is divided by 10. Right: a histogram of the mass from detector-stable particles originating from the  $W$  boson, inside the leading jet, and in the full event from the  $W'$  production. For the full event, the mass is divided by 100.

information is propagated across the graph, ultimately learning sophisticated graph attributes.

Each collision event is represented as a fully connected bidirectional graph in which the nodes are the final state particles and the edges are the connections between all pairs of particles. The node-level attributes are the four-momenta of the particles and the edge- and graph-level attributes will be learned by a GNN. The GNN architecture

is same as the one in Ref. [71], which is based on the model in Ref. [118], composed of four trainable components:

- (1) a node encoder which transforms the node-level attributes into their latent representations;
- (2) an edge encoder which transforms the aggregated latent attributes of its neighbouring nodes into their latent representations;
- (3) an interaction network [119];

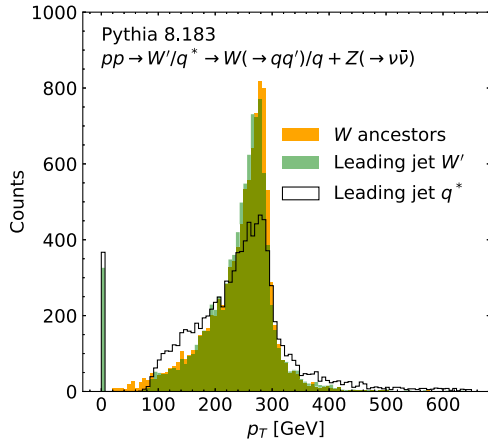


FIG. 3. A histogram of the  $p_T$  of the vector sum of the four momenta of the detector-stable particles originating from the  $W$  boson and inside the leading jet in the  $W'$  and  $q^*$  events.

TABLE I. The edge efficiency and edge purity as a function of the threshold on the edge scores. The definition of the edge efficiency and edge purity can be found in the text.

Threshold	0.1	0.5	0.8
Edge efficiency	0.965	0.896	0.824
Edge purity	0.715	0.908	0.960

(4) and a decoder that computes graph- or edge-level classification scores.

The encoders and the decoder use basic deep learning building blocks including multilayer perceptrons.

The boosted  $W$  boson is reconstructed by training a GNN, namely the *edge classifier*, to learn the relational information of the final state hadrons. Specifically, the edge-level attributes of the simulated  $W$  boson events are labeled as 1 if two hadrons come from the same  $W$  boson

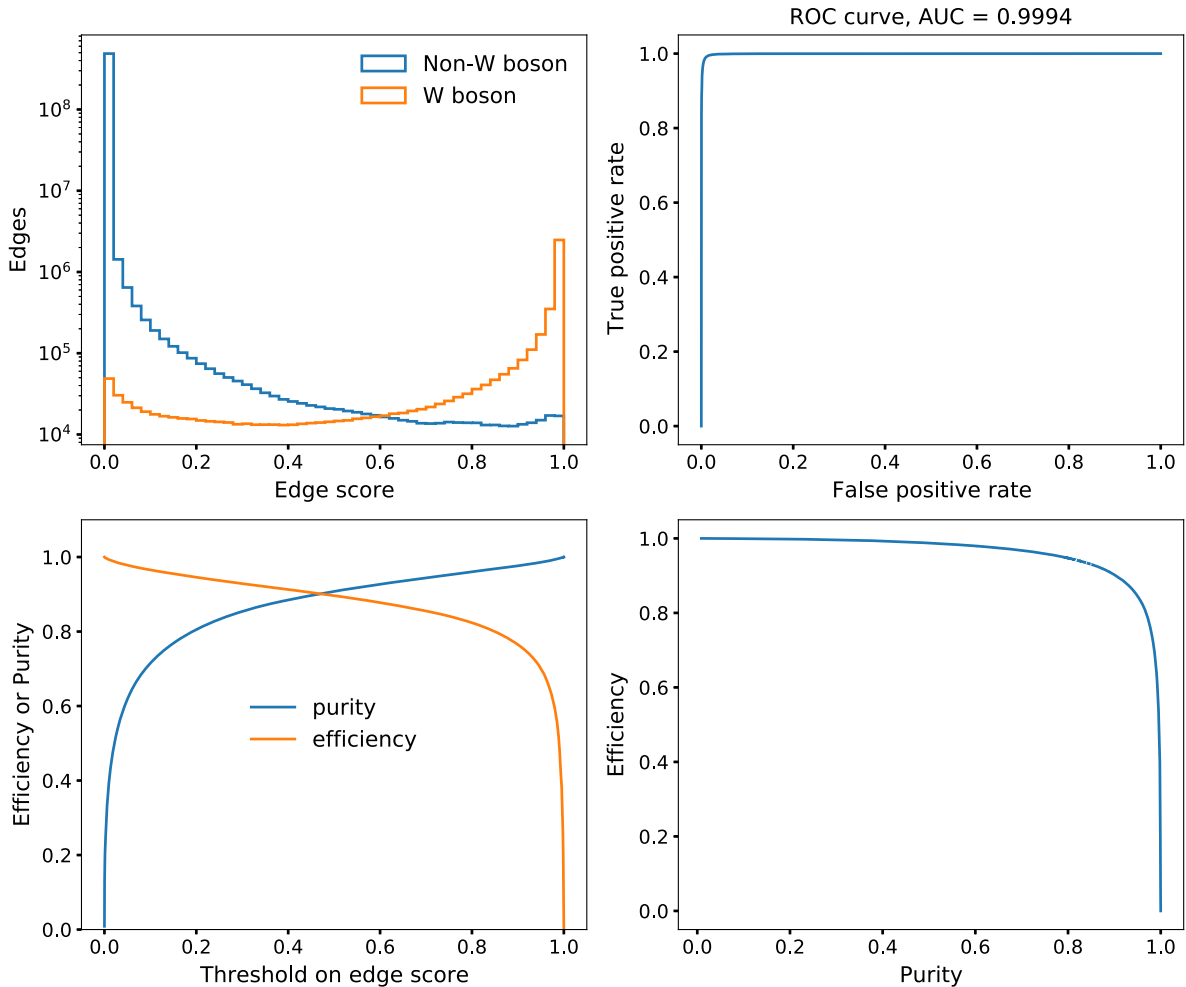


FIG. 4. Metrics used in evaluating classification performance of the edge classifier. Upper left: the distribution of the edge score for edges that connect the hadrons coming from the  $W$  boson in yellow (true edges) and the edges that do not in blue (fake edges). Upper right: the receiver operating characteristic (ROC) curve. AUC is the area under the ROC curve. Bottom left: the edge efficiency and edge purity as a function of the threshold on the edge score. The definition of the edge efficiency and purity can be found in the text. Bottom right: the edge efficiency versus the edge purity.

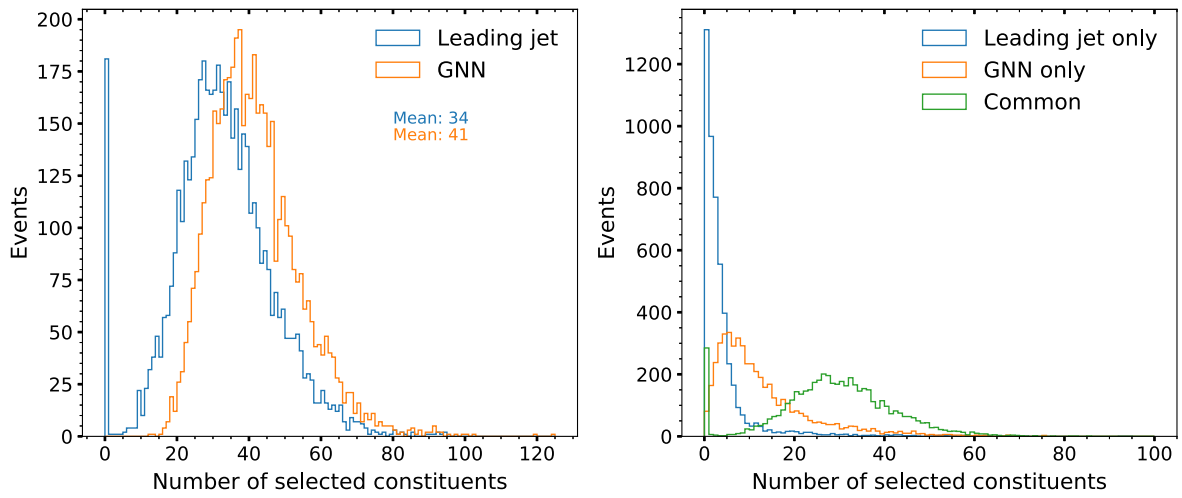


FIG. 5. Left: comparison of the number of selected constituents by the edge classifier (GNN) and the constituents inside the leading jet constructed from the anti- $k_r$  algorithm. Right: Decomposition of the selected constituents from the two methods into the ones selected by both methods in green, only by the leading jet in blue and only by the edge classifier in yellow.

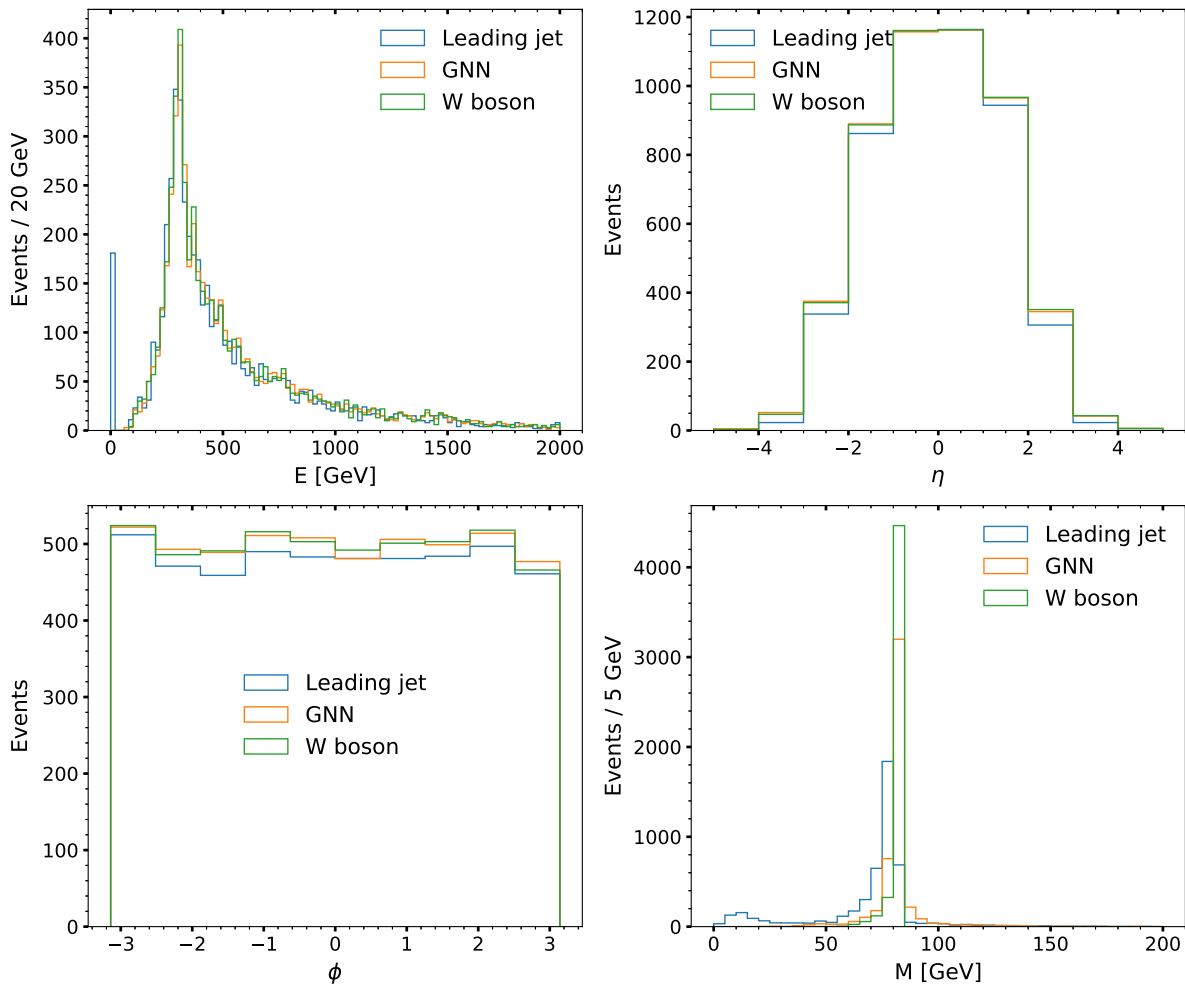


FIG. 6. Comparisons of the four-momenta of the reconstructed  $W$  boson candidates among the anti- $k_r$  jet clustering in blue, the GNN-based jet clustering in yellow and the truth-level  $W$  boson in green. The spike at zero in the top left plot corresponds to events with no anti- $k_r$  jet with  $p_T > 100$  GeV. Such events are removed from the other plots.

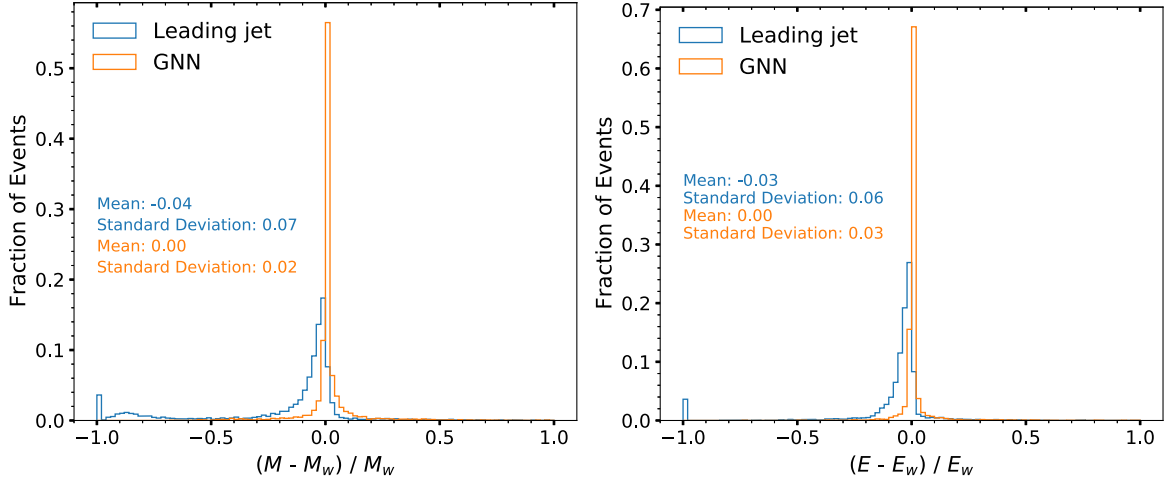


FIG. 7. Comparison of the mass resolution (left) and the energy fraction (right) between the anti- $k_t$  jet clustering and the GNN-based jet clustering. The mean and standard deviation are calculated in the range of  $-0.25$  and  $0.25$  in all cases.

and 0 otherwise. The edge classifier outputs edge-level classification scores, abbreviated as edge scores, which are compared with the edge labeling using the binary cross-entropy loss. Trainable parameters in the classifier are optimized by the gradient-based stochastic optimizer, Adam [120]. The reconstructed  $W$  boson candidate for each event is built from the hadrons that are connected by edges with scores larger than a threshold of 0.5. The threshold is a hyper-parameter that can be tuned for a specific problem. The four-momenta of the reconstructed  $W$  boson candidate are the sum of the four-momenta of the selected hadrons. The “edge classifier” was trained with 90,000 simulated  $W$  boson events and tested with 5,000 events.

The reconstructed  $W$  boson candidates from the GNN-based edge classifier carry unique information which other machine learning architectures (or traditional jet substructure observables) can use in order to separate the  $W$  boson events from background events, such as the  $q^*$  events. In this study, another GNN with the same architecture as the edge classifier is used, namely the *event classifier*. The input graphs are the fully connected bidirectional graphs constructed from the hadrons selected by the trained edge classifier. The graph-level attributes are labeled as 1 for the  $W$  boson events and 0 for the  $q^*$  events. The event classifier outputs the graph-level classification score, abbreviated as event scores, which are compared with the graph labeling using the binary cross-entropy loss. Trainable parameters in the classifier are optimized by the gradient-based stochastic optimizer, Adam. The event classifier was trained with 90,000  $W$  boson events and 90,000  $q^*$  events, and tested with other 5,000  $W$  boson events and 5,000  $q^*$  events. As a comparison, the GNN is also trained with the inputs from the anti- $k_t$  algorithm. In this case, the input graphs are the fully connected bidirectional graphs constructed from the hadrons inside the leading jet which in turn is constructed

from the anti- $k_t$  algorithm. To facilitate the discussions below, the GNN trained with the inputs from the anti- $k_t$  algorithm is called *tGNN* while that trained with the inputs from the trained edge classifier is called *eGNN*. All training was performed on an NVIDIA V100 GPU.

#### IV. RESULTS

The edge classifier was trained for 30 epochs, after which no improvement was seen when the model was evaluated on the testing data. The performance of the edge classifier is showed in Fig. 4. Two important metrics are the edge efficiency, defined as the ratio of the number of true edges passing the threshold over the number of total true edges, and the purity, defined as the ratio of the number of true edges passing the threshold over the number of total

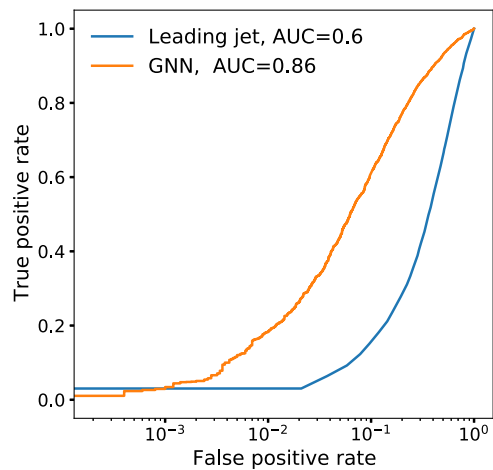


FIG. 8. Comparison of the ROC curve from the GNNs trained with the inputs from the anti- $k_t$  based jet clustering and the inputs from the trained edge classifier. Note that the small inefficiency from the  $p_T$  requirement for the anti- $k_t$  jets is not included.



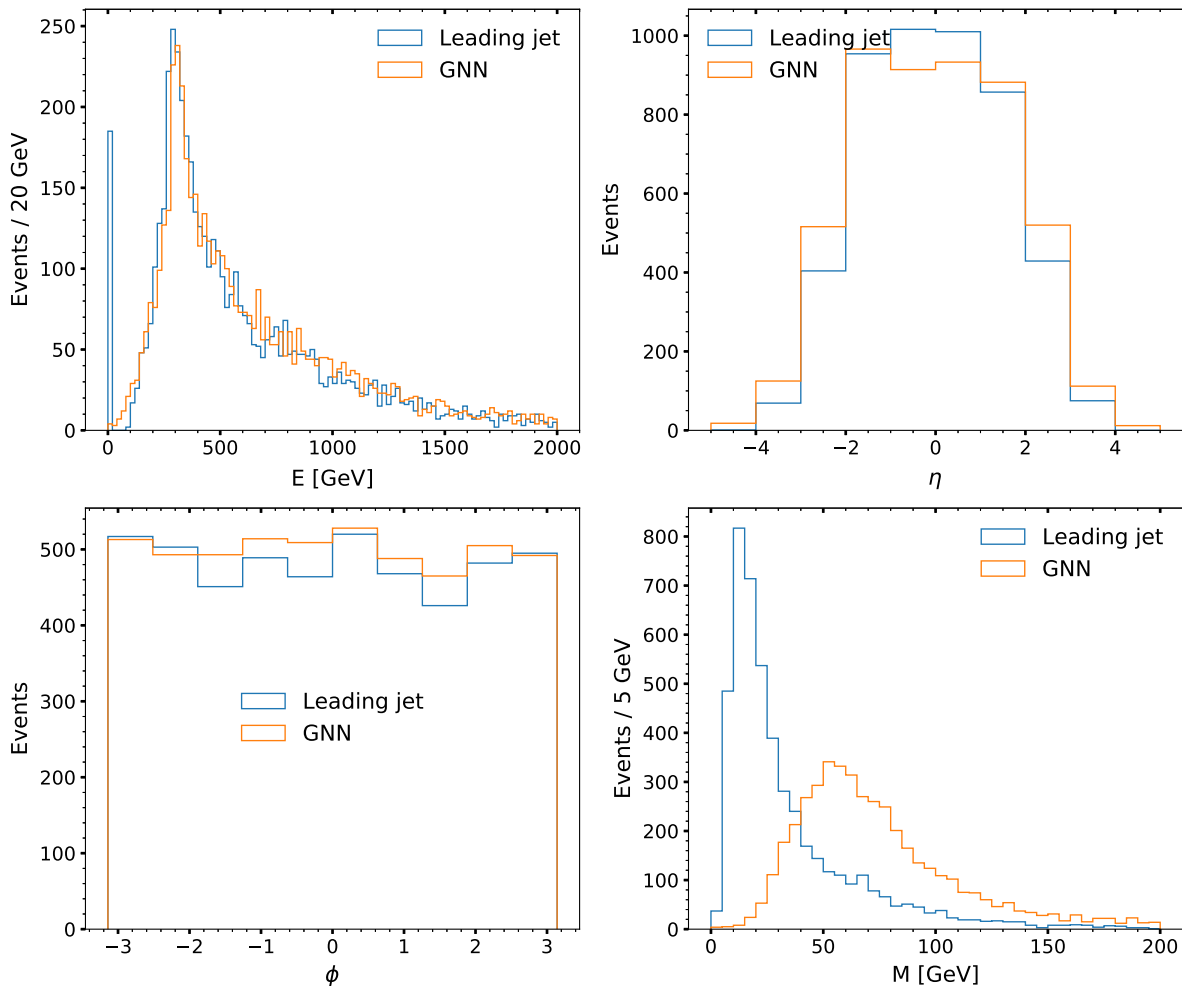


FIG. 9. Comparisons of the four-momenta of the reconstructed jet for the  $q^*$  events between the anti- $k_T$  jet clustering and the GNN jet clustering.

edges passing the threshold. Varying the threshold in the edge scores results in different values of edge efficiency and purity. Table I shows the edge efficiency and purity for three different thresholds on the edge scores.

The nodes that are connected by the edges passing a threshold of 0.5 are considered as the hadrons coming from the  $W$  bosons. The four-momenta of the reconstructed  $W$  boson is the sum of these surviving hadrons. Figure 5 compares the number of hadrons selected by the edge classifier and the anti- $k_t$  algorithm. On average, the number of hadrons selected by the edge classifier is about 20% more than that the anti- $k_t$  jet includes, disregarding the events with no anti- $k_t$  jet with  $p_T > 100$  GeV. There are also many particles chosen by one algorithm but not by the other. It will be interesting in the future to examine the properties of such particles to identify which features the GNN is learning differently than anti- $k_t$  (and vice versa). Furthermore, Fig. 6 compares the kinematic distributions of the  $W$  boson candidates reconstructed from hadrons selected by the edge classifier or the anti- $k_t$  algorithm or the truth-labeled. About 3% of the time, there is no

reconstructed jet with  $p_T > 100$  GeV, which results in the spike at zero. In addition, the fraction of the reconstructed  $W$  energy/mass over the total  $W$  energy/mass are compared between the two methods in Fig. 7. In both cases, the GNN-based method significantly outperforms the anti- $k_t$  based method in reconstructing the boosted  $W$  bosons.<sup>2</sup>

The event classifiers were trained for 25 epochs for the tGNN and 15 epochs for the eGNN. In both cases, no improvement were seen after these epochs when the GNNs were evaluated on the testing data. Figure 8 shows a comparison of the receiver operating characteristic curve (ROC curve) of the two trained GNNs as well as the area under the ROC curve (AUC). The GNN trained with the inputs from the edge classifier outperforms the GNN

<sup>2</sup>There is no correct answer for generic quark jets, but the GNN-based jet clustering is applied to the  $q^*$  events and the four-momenta of the reconstructed jet is compared with the leading jet from the anti- $k_T$  algorithm in the Appendix (Fig. 9). There is a small tendency of the jet mass to be near the  $W$  mass, but it is not as sharp as for  $W$  events.

trained with inputs from the traditional anti- $k_t$  algorithm by more than 40% in AUC.

## V. CONCLUSIONS

Traditional jet clustering based on unsupervised learning has proven to be an effective tool for studying hadronic final states at the LHC. In particular, the widely-used anti- $k_t$  algorithm is both theoretically and experimentally powerful for studying the SM and searching for physics beyond the SM. A wide variety of jet substructure techniques using these jets with and without machine learning are being developed and many have already been deployed in data analysis. However, there is a unique opportunity with color singlet decays to reexamine the construction of jets.

In particular, we have exploited the precise mapping between color singlet particles and final-state hadrons to construct a supervised jet clustering based on graph neural network. These jets match the kinematic properties of true  $W$  bosons more precisely than the leading anti- $k_t$  jet. Furthermore, we have shown that there is more information contained in the graph network jets about the originating particle than anti- $k_t$  jets. In particular, a classifier trained using jet constituents to distinguish  $W$  boson jets from quark jets is more effective for GNN jets than for anti- $k_t$  jets.

This work marks the beginning of a new exploration in jet physics to use machine learning to optimize the construction of jets and not only the observables computed from jet constituents. Tagging Lorentz-boosted color singlet jets is an integral part of measurement and search efforts at the LHC and so further developments in this area

have a significant potential to enhance the sensitivity of the LHC physics program. A variety of further studies will be required to integrate supervised jets into the experimental workflow. In particular, future work will investigate how event topology effects GNN jets (i.e., what happens when there are more ( $W$ ) jets in the event). Furthermore, it is important to study the impact of detector-effects and to investigate how well such jets could be calibrated, including pileup stability.

The studies presented in this paper have only considered boosted  $W$  bosons, but the same ideas could be applied to any color-singlet particles and it will be interesting to see how GNN jets can be integrated with additional information such as  $b$ -jet tagging in the case of Higgs bosons. Examining the structure of the supervised jets may also provide useful physical insight about where the information about the initiating particle is embedded in the event radiation pattern. Finally, it may be that the ultimate performance is achievable when supervised learning is combined with unsupervised techniques and this could lead to new insight for traditional quark and gluon jet reconstruction.

## ACKNOWLEDGMENTS

We would like to thank Andrew Larkoski, Zach Marshall, Ian Moutl, and Jesse Thaler for useful feedback on the manuscript. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

- 
- [1] ATLAS Collaboration, Search for diboson resonances in hadronic final states in  $139 \text{ fb}^{-1}$  of  $pp$  collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector, *J. High Energy Phys.* **09** (2019) 091; Erratum, *J. High Energy Phys.* **06** (2020) 042.
  - [2] ATLAS Collaboration, Search for heavy diboson resonances in semileptonic final states in  $pp$  collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector, [arXiv:2004.14636](https://arxiv.org/abs/2004.14636).
  - [3] CMS Collaboration, A multi-dimensional search for new heavy resonances decaying to boosted  $WW$ ,  $WZ$ , or  $ZZ$  boson pairs in the dijet final state at  $13 \text{ TeV}$ , *Eur. Phys. J. C* **80**, 237 (2020).
  - [4] CMS Collaboration, Combination of CMS searches for heavy resonances decaying to pairs of bosons or leptons, *Phys. Lett. B* **798**, 134952 (2019).
  - [5] ATLAS Collaboration, Search for resonances decaying into a weak vector boson and a Higgs boson in the fully hadronic final state produced in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector, [arXiv:2007.05293](https://arxiv.org/abs/2007.05293).
  - [6] CMS Collaboration, Search for heavy resonances decaying into two Higgs bosons or into a Higgs boson and a  $W$  or  $Z$  boson in proton-proton collisions at  $13 \text{ TeV}$ , *J. High Energy Phys.* **01** (2019) 051.
  - [7] ATLAS Collaboration, Reconstruction and identification of boosted di- $\tau$  systems in a search for Higgs boson pairs using  $13 \text{ TeV}$  proton-proton collision data in ATLAS, [arXiv:2007.14811](https://arxiv.org/abs/2007.14811).
  - [8] CMS Collaboration, Search for resonances decaying to a pair of Higgs bosons in the  $b\bar{b}q\bar{q}'\ell\nu$  final state in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$ , *J. High Energy Phys.* **10** (2019) 125.
  - [9] CMS Collaboration, Search for a massive resonance decaying to a pair of Higgs bosons in the four  $b$  quark final state in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$ , *Phys. Lett. B* **781**, 244 (2018).
  - [10] ATLAS Collaboration, Search for Higgs boson decays into a  $Z$  boson and a light hadronically decaying resonance



- using 13 TeV  $pp$  collision data from the ATLAS detector, [arXiv:2004.01678](https://arxiv.org/abs/2004.01678).
- [11] ATLAS Collaboration, A search for resonances decaying into a Higgs boson and a new particle  $X$  in the  $XH \rightarrow qqbb$  final state with the ATLAS detector, *Phys. Lett. B* **779**, 24 (2018).
- [12] ATLAS Collaboration, Dijet Resonance Search with Weak Supervision Using  $\sqrt{s} = 13$  TeV  $pp$  Collisions in the ATLAS Detector, *Phys. Rev. Lett.* **125**, 131801 (2020).
- [13] ATLAS Collaboration, Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector, *Phys. Lett. B* **788**, 316 (2019).
- [14] CMS Collaboration, Search for Low Mass Vector Resonances Decaying to Quark-Antiquark Pairs in Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV, *Phys. Rev. Lett.* **119**, 111802 (2017).
- [15] CMS Collaboration, Search for low-mass resonances decaying into bottom quark-antiquark pairs in proton-proton collisions at  $\sqrt{s} = 13$  TeV, *Phys. Rev. D* **99**, 012005 (2019).
- [16] CMS Collaboration, Search for Low-Mass Quark-Antiquark Resonances Produced in Association with a Photon at  $\sqrt{s} = 13$  TeV, *Phys. Rev. Lett.* **123**, 231803 (2019).
- [17] CMS Collaboration, Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at  $\sqrt{s} = 13$  TeV, *Phys. Rev. D* **100**, 112007 (2019).
- [18] ATLAS Collaboration, Search for boosted resonances decaying to two b-quarks and produced in association with a jet at  $\sqrt{s} = 13$  TeV with the ATLAS detector, CERN Report No. ATLAS-CONF-2018-052, 2018, <http://cds.cern.ch/record/2649081>.
- [19] CMS Collaboration, Inclusive Search for a Highly Boosted Higgs Boson Decaying to a Bottom Quark-Antiquark Pair, *Phys. Rev. Lett.* **120**, 071802 (2018).
- [20] R. Kogler, B. Nachman, A. Schmidt *et al.*, Jet substructure at the large hadron collider: Experimental review, *Rev. Mod. Phys.* **91**, 045003 (2019).
- [21] I. M. A. J. Larkoski 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).
- [22] D. Adams *et al.*, Towards an understanding of the correlations in jet substructure, *Eur. Phys. J. C* **75**, 409 (2015).
- [23] A. Altheimer *et al.*, Boosted objects and jet substructure at the LHC, *Eur. Phys. J. C* **74**, 2792 (2014).
- [24] A. Abdesselam *et al.*, Boosted objects: A probe of beyond the standard model physics, *Eur. Phys. J. C* **71**, 1661 (2011).
- [25] A. Altheimer *et al.*, Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks, *J. Phys. G* **39**, 063001 (2012).
- [26] G. P. Salam, Towards jetography, *Eur. Phys. J. C* **67**, 637 (2010).
- [27] S. Marzani, G. Soyez, and M. Spannowsky, *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-Object Phenomenology* (Springer, New York, 2019), Vol. 958.
- [28] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Jet Substructure as a New Higgs Search Channel at the LHC, *Phys. Rev. Lett.* **100**, 242001 (2008).
- [29] 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).
- [30] D. Krohn, J. Thaler, and L.-T. Wang, Jet trimming, *J. High Energy Phys.* **02** (2010) 084.
- [31] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, Soft drop, *J. High Energy Phys.* **05** (2014) 146.
- [32] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, Towards an understanding of jet substructure, *J. High Energy Phys.* **09** (2013) 029.
- [33] J. Thaler and K. Van Tilburg, Identifying boosted objects with N-subjettiness, *J. High Energy Phys.* **03** (2011) 015.
- [34] J. Thaler and K. Van Tilburg, Maximizing boosted top identification by minimizing N-subjettiness, *J. High Energy Phys.* **02** (2012) 093.
- [35] A. J. Larkoski, I. Moulton, and D. Neill, Power counting to better jet observables, *J. High Energy Phys.* **12** (2014) 009.
- [36] ATLAS Collaboration, Identification of boosted Higgs bosons decaying into  $b$ -quark pairs with the ATLAS detector at 13 TeV, *Eur. Phys. J. C* **79**, 836 (2019).
- [37] ATLAS Collaboration, Performance of top-quark and  $W$ -boson tagging with ATLAS in Run 2 of the LHC, *Eur. Phys. J. C* **79**, 375 (2019).
- [38] ATLAS Collaboration, In situ calibration of large-radius jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector, *Eur. Phys. J. C* **79**, 135 (2019).
- [39] CMS Collaboration, Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques, *J. Instrum.* **15**, P06005 (2020).
- [40] CMS Collaboration, Jet algorithms performance in 13 TeV data, CERN Report No. CMS-PAS-JME-16-003, <https://cds.cern.ch/record/2256875>.
- [41] ATLAS Collaboration, Impact of alternative inputs and jet grooming on large-R jet performance, CERN Report No. ATL-PHYS-PUB-2019-027, 2019, <http://cds.cern.ch/record/2683619>.
- [42] ATLAS Collaboration, Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton-proton collisions, CERN Report No. ATLAS-CONF-2020-021, 2020, <http://cds.cern.ch/record/2723736>.
- [43] A. Chakraborty, S. Dasmahapatra, H. Day-Hall, B. Ford, S. Jain, S. Moretti, E. Olaiya, and C. Shepherd-Themistocleous, Revisiting jet clustering algorithms for new Higgs boson searches in hadronic final states, [arXiv:2008.02499](https://arxiv.org/abs/2008.02499).
- [44] M. Cacciari, G. P. Salam, and G. Soyez, The anti- $k_r$  jet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [45] L. Mackey, B. Nachman, A. Schwartzman, and C. Stansbury, Fuzzy jets, *J. High Energy Phys.* **06** (2016) 010.
- [46] D. Grigoriev, E. Jankowski, and F. Tkachov, Optimal jet finder, *Comput. Phys. Commun.* **155**, 42 (2003).
- [47] P. Gras, S. Hche, D. Kar, A. Larkoski, L. Lnnblad, S. Pitzer, A. Sidmok, P. Skands, G. Soyez, and J. Thaler, Systematics of quark/gluon tagging, *J. High Energy Phys.* **07** (2017) 091.
- [48] S. Bright-Thonney and B. Nachman, Investigating the topology dependence of quark and gluon jets, *J. High Energy Phys.* **03** (2019) 098.

- [49] M. Andrews, J. Alison, S. An, P. Bryant, B. Burkle, S. Gleyzer, M. Narain, M. Paulini, B. Poczoz, and E. Usai, End-to-end jet classification of quarks and gluons with the CMS open data, *Nucl. Instrum. Methods Phys. Res., Sect. A* **977**, 164304 (2020).
- [50] M. Andrews, M. Paulini, S. Gleyzer, and B. Poczoz, End-to-end physics event classification with CMS open data: Applying image-based deep learning to detector data for the direct classification of collision events at the LHC, *Comput. Software Big Sci.* **4**, 6 (2020).
- [51] A. J. Larkoski, I. Moulton, 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).
- [52] D. Guest, K. Cranmer, and D. Whiteson, Deep learning and its application to LHC physics, *Annu. Rev. Nucl. Part. Sci.* **68**, 161 (2018).
- [53] G. Kasieczka, T. Plehn *et al.*, The machine learning landscape of top taggers, *SciPost Phys.* **7**, 014 (2019).
- [54] HEP ML Community, A living review of machine learning for particle physics, <https://iml-wg.github.io/HEPML-LivingReview/>.
- [55] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman, Jet-images: Computer vision inspired techniques for jet tagging, *J. High Energy Phys.* **02** (2015) 118.
- [56] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee, and M. Perelstein, Playing tag with ANN: Boosted top identification with pattern recognition, *J. High Energy Phys.* **07** (2015) 086.
- [57] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images deep learning edition, *J. High Energy Phys.* **07** (2016) 069.
- [58] ATLAS Collaboration, Quark versus gluon jet tagging using jet images with the ATLAS detector, CERN Report No. Tech. Report No. ATL-PHYS-PUB-2017-017, 2017, <http://cds.cern.ch/record/2275641>.
- [59] J. Lin, M. Freytsis, I. Moulton, and B. Nachman, Boosting  $H \rightarrow b\bar{b}$  with machine learning, *J. High Energy Phys.* **10** (2018) 101.
- [60] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, Learning to classify from impure samples with high-dimensional data, *Phys. Rev. D* **98**, 011502 (2018).
- [61] J. Barnard, E. N. Dawe, M. J. Dolan, and N. Rajcic, Parton shower uncertainties in jet substructure analyses with deep neural networks, *Phys. Rev. D* **95**, 014018 (2017).
- [62] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, Deep learning in color: Towards automated quark/gluon jet discrimination, *J. High Energy Phys.* **01** (2017) 110.
- [63] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, Deep-learning top taggers or the end of QCD?, *J. High Energy Phys.* **05** (2017) 006.
- [64] S. Macaluso and D. Shih, Pulling out all the tops with computer vision and deep learning, *J. High Energy Phys.* **10** (2018) 121.
- [65] T. Q. Nguyen, D. Weitekamp, D. Anderson, R. Castello, O. Cerri, M. Pierini, M. Spiropulu, and J.-R. Vlimant, Topology classification with deep learning to improve real-time event selection at the LHC, *Comput. Software Big Sci.* **3**, 12 (2019).
- [66] ATLAS Collaboration, Convolutional neural networks with event images for pileup mitigation with the ATLAS detector, Tech. Report No. ATL-PHYS-PUB-2019-028, CERN, <http://cds.cern.ch/record/2684070>.
- [67] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, Jet flavor classification in high-energy physics with deep neural networks, *Phys. Rev. D* **94**, 112002 (2016).
- [68] G. Louppe, K. Cho, C. Becot, and K. Cranmer, QCD-aware recursive neural networks for jet physics, *J. High Energy Phys.* **01** (2019) 057.
- [69] T. Cheng, Recursive neural networks in quark/gluon tagging, *Comput. Software Big Sci.* **2**, 3 (2018).
- [70] I. Henrion, K. Cranmer, J. Bruna, K. Cho, J. Brehmer, G. Louppe, and G. Rochette, Neural Message Passing for Jet Physics, *Deep Learning for Physical Sciences Workshop at the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA* (2017), [https://dl4physicsciences.github.io/files/nips\\_dlps\\_2017\\_29.pdf](https://dl4physicsciences.github.io/files/nips_dlps_2017_29.pdf).
- [71] X. Ju *et al.*, Graph neural networks for particle reconstruction in high energy physics detectors, *Machine Learning and the Physical Sciences Workshop at the 33rd Conference on Neural Information Processing Systems, Vancouver, Canada* (2020), <https://ml4physicsciences.github.io/2019>.
- [72] J. Arjona Martinez, O. Cerri, M. Pierini, M. Spiropulu, and J.-R. Vlimant, Pileup mitigation at the large hadron collider with graph neural networks, *Eur. Phys. J. Plus* **134**, 333 (2019).
- [73] E. A. Moreno, O. Cerri, J. M. Duarte, H. B. Newman, T. Q. Nguyen, A. Perival, 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).
- [74] S. R. Qasim, J. Kieseler, Y. Iiyama, and M. Pierini, Learning representations of irregular particle-detector geometry with distance-weighted graph networks, *Eur. Phys. J. C* **79**, 608 (2019).
- [75] A. Chakraborty, S. H. Lim, and M. M. Nojiri, Interpretable deep learning for two-prong jet classification with jet spectra, *J. High Energy Phys.* **19** (2020) 135.
- [76] A. Chakraborty, S. H. Lim, M. M. Nojiri, and M. Takeuchi, Neural network-based top tagger with two-point energy correlations and geometry of soft emissions, *J. High Energy Phys.* **07** (2020) 111.
- [77] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer, and A. Mück, Casting a graph net to catch dark showers, [arXiv:2006.08639](https://arxiv.org/abs/2006.08639).
- [78] P. T. Komiske, E. M. Metodiev, and J. Thaler, Energy flow networks: Deep sets for particle jets, *J. High Energy Phys.* **01** (2019) 121.
- [79] H. Qu and L. Gouskos, ParticleNet: Jet tagging via particle clouds, *Phys. Rev. D* **101**, 056019 (2020).
- [80] K. Datta, A. Larkoski, and B. Nachman, Automating the construction of jet observables with machine learning, [arXiv:1902.07180](https://arxiv.org/abs/1902.07180).
- [81] K. Datta and A. Larkoski, How Much Information is in a Jet? *J. High Energy Phys.* **06** (2017) 073.
- [82] K. Datta and A. J. Larkoski, Novel jet observables from machine learning, *J. High Energy Phys.* **03** (2018) 086.
- [83] P. T. Komiske, E. M. Metodiev, and J. Thaler, Energy flow polynomials: A complete linear basis for jet substructure, *J. High Energy Phys.* **04** (2018) 013.

- [84] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, Deep-learned top tagging with a Lorentz layer, *SciPost Phys.* **5**, 028 (2018).
- [85] Y.-C. J. 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 (2020).
- [86] K. Fraser and M. D. Schwartz, Jet charge and machine learning, *J. High Energy Phys.* **10** (2018) 093.
- [87] K. Datta, A. Larkoski, and B. Nachman, Automating the construction of jet observables with machine learning, *Phys. Rev. D* **100**, 095016 (2019).
- [88] E. A. Moreno, T. Q. Nguyen, J.-R. Vlimant, O. Cerri, H. B. Newman, A. Periwal, M. Spiropulu, J. M. Duarte, and M. Pierini, Interaction networks for the identification of boosted  $H \rightarrow b\bar{b}$  decays, *Phys. Rev. D* **102**, 012010 (2020).
- [89] M. Stoye, J. Kieseler, M. Verzetti, H. Qu, L. Gouskos, and A. Stakia (CMS Collaboration), DeepJet: Generic physics object based jet multiclass classification for LHC experiments, *Deep Learning for Physical Sciences Workshop at the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA* (2017), [https://dl4physicsciences.github.io/files/nips\\_dlps\\_2017\\_10.pdf](https://dl4physicsciences.github.io/files/nips_dlps_2017_10.pdf).
- [90] Y.-T. Chien and R. Kunnawalkam Elayavalli, Probing heavy ion collisions using quark and gluon jet substructure, [arXiv:1803.03589](https://arxiv.org/abs/1803.03589).
- [91] G. Kasieczka, N. Kiefer, T. Plehn, and J. M. Thompson, Quark-gluon tagging: Machine learning vs detector, *SciPost Phys.* **6**, 069 (2019).
- [92] G. Kasieczka, S. Marzani, G. Soyez, and G. Stagnitto, Towards machine learning analytics for jet substructure, [arXiv:2007.04319](https://arxiv.org/abs/2007.04319).
- [93] S. Diefenbacher, H. Frost, G. Kasieczka, T. Plehn, and J. M. Thompson, CapsNets continuing the convolutional quest, *SciPost Phys.* **8**, 023 (2020).
- [94] Y. Nakai, D. Shih, and S. Thomas, Strange jet tagging, [arXiv:2003.09517](https://arxiv.org/abs/2003.09517).
- [95] CMS Collaboration, Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV, *J. Instrum.* **13**, P05011 (2018).
- [96] J. Bielcikoa, R. K. Elayavalli, G. Ponimatkin, J. H. Putschke, and J. Sivic, Identifying heavy-flavor jets using vectors of locally aggregated descriptors, [arXiv:2005.01842](https://arxiv.org/abs/2005.01842).
- [97] P. Baldi, P. Sadowski, and D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, *Nat. Commun.* **5**, 4308 (2014).
- [98] CMS Collaboration, A deep neural network to search for new long-lived particles decaying to jets, *Mach. Learn. Sci. Technol.* **1**, 035012 (2020).
- [99] J. Alimena, Y. Iiyama, and J. Kieseler, Fast convolutional neural networks for identifying long-lived particles in a high-granularity calorimeter, [arXiv:2004.10744](https://arxiv.org/abs/2004.10744).
- [100] L. De Oliveira, B. Nachman, and M. Paganini, Electromagnetic showers beyond shower shapes, *Nucl. Instrum. Methods Phys. Res., Sect. A* **951**, 162879 (2020).
- [101] M. Paganini, L. de Oliveira, and B. Nachman, Survey of machine learning techniques for high energy electromagnetic shower classification, *Deep Learning for Physical Sciences Workshop at the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA* (2017), [https://dl4physicsciences.github.io/files/nips\\_dlps\\_2017\\_24.pdf](https://dl4physicsciences.github.io/files/nips_dlps_2017_24.pdf).
- [102] B. Hooberman, A. Farbin, G. Khattak, V. Pacela, M. Pierini, J.-R. Vlimant, M. Spiropulu, W. Wei, M. Zhang, and S. Vallecorsa, Calorimetry with deep learning: particle classification, energy regression, and simulation for high-energy physics, *Deep Learning for Physical Sciences Workshop at the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA* (2017), [https://dl4physicsciences.github.io/files/nips\\_dlps\\_2017\\_15.pdf](https://dl4physicsciences.github.io/files/nips_dlps_2017_15.pdf).
- [103] D. Belayneh *et al.*, Calorimetry with deep learning: Particle simulation and reconstruction for collider physics, *Eur. Phys. J. C* **80**, 688 (2020).
- [104] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer, and A. Mück, Casting a graph net to catch dark showers, [arXiv:2006.08639](https://arxiv.org/abs/2006.08639).
- [105] M. Abdughani, D. Wang, L. Wu, J. M. Yang, and J. Zhao, Probing triple Higgs coupling with machine learning at the LHC, [arXiv:2005.11086](https://arxiv.org/abs/2005.11086).
- [106] J. Ren, L. Wu, and J. M. Yang, Unveiling  $CP$  property of top-Higgs coupling with graph neural networks at the LHC, *Phys. Lett. B* **802**, 135198 (2020).
- [107] M. Abdughani, J. Ren, L. Wu, and J. M. Yang, Probing stop pair production at the LHC with graph neural networks, *J. High Energy Phys.* **08** (2019) 055.
- [108] J. Shlomi, P. Battaglia, and J.-R. Vlimant, Graph neural networks in particle physics, [arXiv:2007.13681](https://arxiv.org/abs/2007.13681).
- [109] N. Choma *et al.*, Track seeding and labelling with embedded-space graph neural networks, [arXiv:2007.00149](https://arxiv.org/abs/2007.00149).
- [110] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, Pileup mitigation with machine learning (PUMML), *J. High Energy Phys.* **12** (2017) 051.
- [111] C. Adams *et al.* (MicroBooNE Collaboration), Deep neural network for pixel-level electromagnetic particle identification in the MicroBooNE liquid argon time projection chamber, *Phys. Rev. D* **99**, 092001 (2019).
- [112] D. H. Koh *et al.*, Scalable, Proposal-free instance segmentation network for 3D pixel clustering and particle trajectory reconstruction in liquid argon time projection chambers, [arXiv:2007.03083](https://arxiv.org/abs/2007.03083).
- [113] Y. Iiyama *et al.*, Distance-weighted graph neural networks on FPGAs for real-time particle reconstruction in high energy physics, [arXiv:2008.03601](https://arxiv.org/abs/2008.03601).
- [114] T. Sjostrand, S. Mrenna, and P. Z. Skands, A brief introduction to PYTHIA8.1, *Comput. Phys. Commun.* **178**, 852 (2008).
- [115] T. Sjöstrand, S. Mrenna, and P. Z. Skands, PYTHIA6.4 physics and manual, *J. High Energy Phys.* **05** (2006) 026.
- [116] M. Cacciari, G. P. Salam, and G. Soyez, FASTJET user manual, *Eur. Phys. J. C* **72**, 1896 (2012).
- [117] M. Cacciari and G. P. Salam, Dispelling the  $N^3$  myth for the  $k_t$  jet-finder, *Phys. Lett. B* **641**, 57 (2006).
- [118] P. W. Battaglia *et al.*, Relational inductive biases, deep learning, and graph networks, [arXiv:1806.01261](https://arxiv.org/abs/1806.01261).
- [119] P. W. Battaglia, R. Pascanu, M. Lai, D. Rezende, and K. Kavukcuoglu, Interaction networks for learning about objects, relations and physics, [arXiv:1612.00222](https://arxiv.org/abs/1612.00222).
- [120] D. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).