

A generic anti-QCD jet tagger

J.A. Aguilar-Saavedra,^a Jack Collins^{b,c} and Rashmish K. Mishra^{b,d}

^a*Departamento de Física Teórica y del Cosmos, Universidad de Granada,
E-18071 Granada, Spain*

^b*Maryland Center for Fundamental Physics, Department of Physics, University of Maryland,
College Park, MD 20742, U.S.A.*

^c*Department of Physics and Astronomy, Johns Hopkins University,
Baltimore, MD 21218, U.S.A.*

^d*INFN, Pisa, Italy and Scuola Normale Superiore,
Piazza dei Cavalieri 7, 56126 Pisa, Italy*

E-mail: jaas@ugr.es, jhc296@umd.edu, rashmish@pi.infn.it

ABSTRACT: New particles beyond the Standard Model might be produced with a very high boost, for instance if they result from the decay of a heavier particle. If the former decay hadronically, then their signature is a single massive fat jet which is difficult to separate from QCD backgrounds. Jet substructure and machine learning techniques allow for the discrimination of many specific boosted objects from QCD, but the scope of possibilities is very large, and a suite of dedicated taggers may not be able to cover every possibility — in addition to making experimental searches cumbersome. In this paper we describe a generic model-independent tagger that is able to discriminate a wide variety of hadronic boosted objects from QCD jets using N -subjettiness variables, with a significance improvement varying between 2 and 8. This is in addition to any improvement that might come from a cut on jet mass. Such a tagger can be used in model-independent searches for new physics yielding fat jets. We also show how such a tagger can be applied to signatures over a wide range of jet masses without sculpting the background distributions, allowing to search for new physics as bumps on jet mass distributions.

KEYWORDS: Jets

ARXIV EPRINT: [1709.01087](https://arxiv.org/abs/1709.01087)

Contents

1	Introduction	1
2	Framework	3
3	A first approach to anti-QCD tagging	5
4	Dedicated versus generic taggers	9
5	Mass decorrelation	11
6	Discussion	15
A	How much information is in a multi-pronged jet?	17
B	Effect of signal composition on training	18
C	Background composition: effect of quark to gluon ratio	19
D	NN architecture	21
E	Is the tagger learning shape or kinematics?	21
F	Pythia versus Herwig	24
G	Boosted coloured jets	25

1 Introduction

Jet tagging algorithms have become an essential tool to explore the high energy frontier at the Large Hadron Collider (LHC). New physics processes are expected to involve the production of highly-boosted top quarks, W/Z , and Higgs bosons, which in their hadronic decay give rise to a ‘fat jet’ J where the decay products are highly collimated. In order to distinguish such jets from the background jets resulting from quarks and gluons, generically denoted as ‘QCD jets’, several jet substructure analysis techniques have been developed [1–13]. These are extensively used in searches for $W' \rightarrow tb$ [14, 15], $t\bar{t}$ resonances [16], diboson resonances [17–20], vector-like quarks [21, 22], dark matter produced in association with gauge bosons [23, 24] and new light bosons [25].

Experimental analyses carried out by the ATLAS and CMS collaborations use dedicated taggers in addition to the jet mass, to search for beyond the Standard Model (BSM) scenarios that can give rise to boosted top quarks, W/Z , or Higgs bosons. For example, shape variables such as the N -subjettiness ratio $\tau_{21}^{(1)}$ [7] and the energy correlation function $D_2^{(\beta=1)}$ [11] are very effective in distinguishing between QCD jets and two-pronged decays from W/Z , and the performance can be further improved by using a more complete set of jet substructure variables and a multivariate analysis [13]. As another example, the subjettiness ratio $\tau_{32}^{(1)}$ is used to identify jets from top quark decays. However, the inherent drawback in this approach is that, while these dedicated taggers are efficient in the

discrimination of top quarks and W/Z hadronic decays from QCD jets, they may not be able to identify fat jets arising from the decay of BSM boosted particles.

New particles near the electroweak scale may exist and evade direct detection, for example, if their couplings to quarks and gauge bosons are small. They can still be produced in the decay of heavier particles and may have dominant decays into hadronic final states. Examples of such cases are neutral (pseudo-)scalars in models with left-right symmetry [26] and warped extra dimensional models with more than 2 branes [27, 28] (see also refs. [29, 30]). An explicit example of the limitations of dedicated jet taggers has been given in ref. [31], by considering a new ‘stealth boson’ S with a mass in the 100 GeV range and undergoing a cascade decay $S \rightarrow AA \rightarrow b\bar{b}b\bar{b}$ mediated by a lighter particle A . When S is boosted, so that the four b quarks merge into a single jet, the $\tau_{21}^{(1)}$ and $D_2^{(\beta=1)}$ variables used to tag massive SM bosons would ‘see’ the resulting four-pronged fat jet as a QCD jet. Consequently, new physics searches involving boosted hadronically-decaying W or Z bosons, e.g. diboson resonance searches, can be relatively blind to the analogous new physics processes (diboson-like resonances) involving one or two S particles of a mass around $M_{W,Z}$.

It is highly desirable that ATLAS and CMS searches are not restricted to a few simple benchmark models, but rather cover as many new physics signatures as possible. A broader scope for LHC searches becomes of the utmost importance given the absence of any convincing hint of new physics beyond the SM, as we still do not know how new physics may manifest at collider experiments. With that purpose, a generic ‘anti-QCD’ tagger that distinguishes QCD jets not only from W/Z hadronic decays, but also from generic BSM boosted objects, would be a useful tool. In this paper we address this problem and provide a proof of concept that this kind of tool can be developed (see also ref. [32] which pursues related ideas). With this goal, we perform a multivariate analysis using a neural network (NN) that is trained to discriminate QCD jets from fat jets with two-, three- and four-pronged structure, arising from the decay of relatively light boosted particles. After describing our framework in section 2, we perform a simple analysis in section 3, to demonstrate the discrimination power for several examples of fat jets from boosted new particles against QCD jets. A comparison between the performance of generic and dedicated taggers is given in section 4. The decorrelation between the background rejection with tagging based on jet substructure and the jet mass requires a slightly more sophisticated analysis, which is presented in section 5. Our results are discussed in section 6. Some appendices are devoted to additional details of our analysis. In appendix A we study the dependence of the results on the number of input variables for the NNs. In appendices B and C we discuss how the results change when we modify the signal flavour composition, and the quark/gluon background composition, respectively. The dependence of the results on the NN architecture is explored in appendix D. In appendix E we examine the issue of whether the taggers only learn jet shapes or they also learn about different signal and background kinematics. A related issue is the dependence of the results on the specific model for hadronisation and showering; this is addressed in appendix F, where we compare the results using two Monte Carlo simulation codes. Finally, in appendix G we study for completeness the signals of light coloured boosted objects.

2 Framework

Following ref. [13], we characterise the jet substructure by a set of generalised N -subjettiness [10] variables

$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_i p_{Ti} \min \left\{ \Delta R_{1i}^\beta, \Delta R_{2i}^\beta, \dots, \Delta R_{Ni}^\beta \right\}, \quad (2.1)$$

with i labelling the particles in the jet, p_{Ti} their transverse momenta, ΔR_{Ki} their lego-plot distance to the axis $K = 1, \dots, N$ and p_{TJ} the jet transverse momentum. As in ref. [13], in the computation of these variables we use the axes defined by exclusive k_T algorithm [33, 34] with standard E -scheme recombination [35]. Ref. [13] proposed the following basis of observables,

$$\left\{ \tau_1^{(1/2)}, \tau_1^{(1)}, \tau_1^{(2)}, \dots, \tau_{M-2}^{(1/2)}, \tau_{M-2}^{(1)}, \tau_{M-2}^{(2)}, \tau_{M-1}^{(1)}, \tau_{M-1}^{(2)} \right\}, \quad (2.2)$$

motivated by the requirement to be able to fully reconstruct the $(3M - 4)$ -dimensional phase space for a decay into M particles. They found that the discriminating power for Z -jets versus gluon and quark jets was saturated by considering up to 4-body phase space. Because we are interested also in higher pronged decays we explore a larger 17-dimensional basis with $M = 7$. This specific choice is motivated in appendix A. It is likely that a smaller, more carefully selected basis of substructure variables could be used with little degradation in discrimination power, but in this work we do not attempt to optimise this. We remark that, equivalently, a set of energy correlation functions [11] $\text{ECF}(N, \beta)$ could also be used, but the calculations are much more computationally-demanding when one considers higher N , as is required for the identification of multi-pronged boosted jets.

The values of these variables are used as the input to a NN trained to discriminate quark/gluon jets from multi-pronged decays of boosted colour singlet particles. Quark and gluon jets are obtained by generating the parton-level processes $pp \rightarrow Zg$ and $pp \rightarrow Zq$, with $Z \rightarrow \nu\nu$, using MADGRAPH5 [36]. Event generation is followed by hadronisation and parton showering with PYTHIA 8 [37]. The detector response is simulated with DELPHES 3.4 [38] using the CMS detector card. Jets are reconstructed using the anti- k_T algorithm [39] with radius $R = 0.8$, as implemented in FASTJET 3.2 [40]. For the signal we use fat jets resulting from the decay of neutral, colour-singlet particles into two, three and four quarks. For these, we consider the six processes

$$\begin{aligned} pp \rightarrow Z' \rightarrow S Z(\rightarrow \nu\nu), & \quad S \rightarrow u\bar{u} \quad \text{and} \quad S \rightarrow b\bar{b}, \\ pp \rightarrow Z' \rightarrow F Z(\rightarrow \nu\nu)\nu, & \quad F \rightarrow udd \quad \text{and} \quad F \rightarrow ubb, \\ pp \rightarrow Z' \rightarrow S Z(\rightarrow \nu\nu), & \quad S \rightarrow u\bar{u}u\bar{u} \quad \text{and} \quad S \rightarrow b\bar{b}b\bar{b}, \end{aligned} \quad (2.3)$$

with S a scalar and F a fermion. These processes are generated with PROTOS [41] and, in order to remain as model-agnostic as possible, we implement decays of S and F with a flat matrix element, so that the decay weight of the different kinematical configurations only corresponds to the two-, three- or four-body phase space. We will refer to these Monte Carlo data as Model Independent (MI) data in the following. Our choice is motivated

by the need to sample phase space without model prejudice. For example, any specific choice of four-body decay topology, such as $1 \rightarrow 1 + 1 \rightarrow 2 + 2$, combined with a choice of masses for the intermediate particles, would only sample a part of four-body phase space, which varies with those mass choices. Therefore, training on specific cascade modes would introduce a model bias. Our choice to train on both light and b quarks is also with the same aim, of making the tagger as model-agnostic as possible. Variations on this choice, either removing final states with b quarks, or adding signal processes with gluons (e.g. $S \rightarrow gg$) in the training, are discussed in appendix B.

Several new physics signal processes are generated to test whether the NN correctly identifies jets resulting from boosted multi-pronged particle decays, including some for which it is not trained. We use seven such processes,

$$\begin{aligned}
 pp \rightarrow W' \rightarrow W Z(\rightarrow \nu\nu), & \quad W \rightarrow q\bar{q}', \\
 pp \rightarrow Z' \rightarrow H_1^0 Z(\rightarrow \nu\nu), & \quad H_1^0 \rightarrow gg, \\
 pp \rightarrow Z' \rightarrow H_1^0 Z(\rightarrow \nu\nu), & \quad H_1^0 \rightarrow A^0 A^0, \\
 pp \rightarrow Z' \rightarrow H_1^0 Z(\rightarrow \nu\nu), & \quad H_1^0 \rightarrow t\bar{t}, \\
 pp \rightarrow Z' \rightarrow H_1^0 Z(\rightarrow \nu\nu), & \quad H_1^0 \rightarrow W^+ W^-, \\
 pp \rightarrow Z' \rightarrow H_1^0 Z(\rightarrow \nu\nu), & \quad H_1^0 \rightarrow Z A^0, \\
 pp \rightarrow W' \rightarrow H^\pm Z(\rightarrow \nu\nu), & \quad H^\pm \rightarrow tb,
 \end{aligned} \tag{2.4}$$

with H_1^0 a heavy scalar and A^0 a pseudo-scalar, H^\pm a charged scalar and Z', W' additional vector bosons. All these new particles arise, for example, in left-right models. We consider hadronic decays of the top quarks, W/Z bosons and pseudo-scalars resulting from the H_1^0 and H^\pm decays.¹ We note that for W and Z hadronic decays the jet shapes are very similar, so for brevity we only consider the former. These processes are generated with MADGRAPH5 implementing the relevant interactions [26] in FEYNRULES [43], and using the universal Feynrules output [44] to interface with the event generator.

We treat the search for boosted BSM objects as a binary classification problem, with quark and gluon jets labelled as background and jets originating from boosted massive objects labelled as signal. Our NN classifiers are multilayer perceptrons, a simple fully connected architecture that is well suited for use with unstructured input data. These are implemented using Keras [45] with a TensorFlow backend [46]. We choose an architecture with two hidden layers, the first containing 512 nodes and the second containing 32 nodes, all using rectifier activation functions. (See appendix D for a few examples using alternative NN architectures.) The output layer is a single node with sigmoid activation. The input consists of the 17 $\tau_N^{(\beta)}$ variables, with some preprocessing applied. We use two kinds of preprocessing which we discuss in more detail in section 3 and section 5 respectively. The first is a simple standardisation of the inputs, which we find significantly improves the training time, stability over variations of the initial seed, and discrimination performance

¹When the W and Z bosons decay leptonically, the resulting jets have energetic leptons that can be further used for background rejection [42]. Therefore, we restrict ourselves to the most difficult scenario of fully hadronic decays.

of the trained tagger for simple architectures like the ones used here. The second approach relies on a more complicated transformation of the input data and allows a tagger to be sensitive to signals over a broad range of masses, while decorrelating background rejection from jet mass and p_T .

Except where otherwise specified, we train on equal numbers of background and signal events. Background is divided equally between quark and gluon jets (see appendix C for variations in this choice). Signal training data is divided equally between the six categories of MI events described in eqs. (2.3). 20% of this signal and background data is set aside for validation. We choose binary cross entropy as the loss function to be optimised, using the Root Mean Square Propagation (RMSProp) algorithm with a learning rate of 10^{-3} . Additionally, if the loss as measured on validation data does not improve over three epochs, the learning rate is reduced by a factor of 10. Training is stopped after 100 epochs, or when validation loss has not improved in five epochs. Typically, we find that training in this manner takes several tens of epochs. We train five different copies of each NN in the same way but with different starting seeds, and pick the one which has the best performance as measured by area under the receiver operating characteristic (ROC) curve with validation data. All machine learning calculations were performed with an Intel(R) Core(TM) i5-6300U CPU @ 2.40 GHz with 8 GB of RAM. Our NNs take 1-10 minutes to train.

3 A first approach to anti-QCD tagging

For this first simple analysis each of the N -subjettiness variables in eq. (2.1) is standardised by a linear transformation,

$$\tau_N^{(\beta)} \rightarrow \tau_N^{\text{std}(\beta)} = a_N^{(\beta)} + b_N^{(\beta)} \tau_N^{(\beta)}, \quad (3.1)$$

with $a_N^{(\beta)}$ and $b_N^{(\beta)}$ constant, so that the resulting $\tau_N^{\text{std}(\beta)}$ distribution for the QCD background (composed of equal parts of quark and gluon jets) has zero mean and unit standard deviation. We consider three benchmarks for the jet transverse momentum p_{TJ} and mass m_J , and for each one we also select the Z' resonance mass in eqs. (2.3) to yield MI data with a p_{TJ} distribution close to the threshold and similar to the background (see appendix E). One tagger is built in each case,

- (a) Tagger ‘std500’: $p_{TJ} > 500$ GeV, $m_J \in [65 - 105]$ GeV, $M_{Z'} = 1100$ GeV.
- (b) Tagger ‘std1000’: $p_{TJ} > 1000$ GeV, $m_J \in [65 - 105]$ GeV, $M_{Z'} = 2200$ GeV.
- (c) Tagger ‘std1500’: $p_{TJ} > 1500$ GeV, $m_J \in [350 - 450]$ GeV, $M_{Z'} = 3300$ GeV.

In the first two cases, the MI data used for training is generated with boosted particle masses $M_{S,F} = 80$ GeV, and in the last case with masses of 400 GeV. The jet mass m_J and transverse momentum p_{TJ} used here are of the ungroomed jet, for reasons that we specify at the end of this section. The number of events used for the training is collected in table 1.

	std500	std1000	std1500
Training sample size	284,016	249,206	144,884
Validation sample size	71,004	62,302	36,220

Table 1. Training sample sizes for `std` taggers.

The `std500` and `std1000` taggers are used to investigate the discrimination power for jets coming from BSM boosted particles with a mass around the W mass. We use two regimes of p_{TJ} ($p_T > 500$ GeV and $p_T > 1000$ GeV) to check the differences, and the extent to which the results are specific — or not — to a kinematical region. For each benchmark, a NN is trained and validated with MI and QCD data, and is then tested on a number of boosted jet topologies,²

$$\begin{aligned}
 &W \rightarrow q\bar{q}, \\
 &H_1^0 \rightarrow gg, \quad M_{H_1^0} = 80 \text{ GeV}, \\
 &H_1^0 \rightarrow A^0 A^0 \rightarrow b\bar{b}b\bar{b}, \quad M_{H_1^0} = 80 \text{ GeV}, \quad M_{A^0} = 30 \text{ GeV}, \\
 &H_1^0 \rightarrow A^0 A^0 \rightarrow b\bar{b}b\bar{b}, \quad M_{H_1^0} = 80 \text{ GeV}, \quad M_{A^0} = 15 \text{ GeV}, \\
 &H_1^0 \rightarrow A^0 A^0 \rightarrow u\bar{u}u\bar{u}, \quad M_{H_1^0} = 80 \text{ GeV}, \quad M_{A^0} = 30 \text{ GeV}, \quad (3.2)
 \end{aligned}$$

setting the parent Z' resonance mass responsible for the processes in eq. (3.2) to 1100 GeV and 2200 GeV for the $p_T > 500$ GeV and $p_T > 1000$ GeV test samples respectively (the p_T distributions of the QCD and signal jets generated in this way are very similar, see appendix E). The third and fourth line in eq. (3.2) are two examples of the stealth boson S in ref. [31]. The results for the ROC curves giving the signal efficiency versus background rejection are presented in figure 1. To better illustrate the effect of the tagging on the signal-to-background significance S/\sqrt{B} , we define significance improvement as the factor multiplying the luminosity-dependent ratio S/\sqrt{B} due to the tagging, and indicate the lines (in dashed gray) that correspond to a significance improvement of 1, 2, 4 and 8. For comparison we also include the efficiency curve for the dedicate tagger $\tau_{21}^{(1)}$, applied to fat jets from W bosons. Several comments are in order.

1. The taggers perform better for jets with light quarks, either from W bosons or from stealth bosons decaying to four u quarks.
2. For W bosons the anti-QCD taggers represent a significant improvement over the dedicated tagger $\tau_{21}^{(1)}$, as also observed in ref. [13].
3. The discriminating power of the `std1000` tagger, when applied to all jet topologies, outperforms the discriminating power of $\tau_{21}^{(1)}$ when applied to W bosons.
4. Although the discrimination power of the taggers is worse for the two stealth boson examples giving four b quarks, it is far better than the one that would be achieved

²We consider here colour-singlet new particles, as they are the most likely ones that could be mainly seen from the decay of a heavier one, and not from their direct production. An example of the application of the tagger to a coloured particle is presented in appendix G.

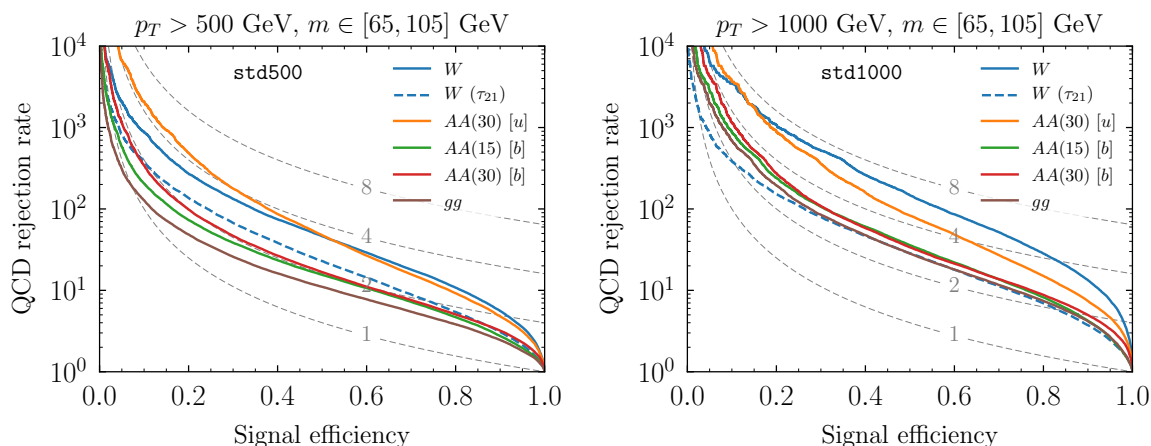


Figure 1. Signal efficiency versus background rejection for the anti-QCD `std500` (left) and `std1000` (right) taggers. Significance improvement due to shape-variable tagging is indicated by the dashed grey contours. Also indicated is the efficiency curve for $\tau_{21}^{(1)}$ for hadronic W decay (dashed blue).

with $\tau_{21}^{(1)}$, which is specifically designed for W bosons and actually may reduce the S/\sqrt{B} ratio for this type of signals [31].

5. The taggers have a good discrimination for fat jets from $H_1^0 \rightarrow gg$, for which they are not trained.

The `std1500` tagger is used to test the performance at higher jet masses and also the ability to distinguish more complex boosted signatures,

$$\begin{aligned}
 H_1^0 &\rightarrow WW \rightarrow q\bar{q}'q\bar{q}', & M_{H_1^0} &= 400 \text{ GeV}, \\
 H_1^0 &\rightarrow gg, & M_{H_1^0} &= 400 \text{ GeV}, \\
 H_1^0 &\rightarrow t\bar{t} \rightarrow WbW\bar{b}, \rightarrow q\bar{q}'bq\bar{q}'\bar{b}, & M_{H_1^0} &= 400 \text{ GeV}, \\
 H^{\pm} &\rightarrow t\bar{b}/\bar{t}b \rightarrow Wb\bar{b} \rightarrow q\bar{q}'b\bar{b}, & M_{H^{\pm}} &= 400 \text{ GeV}, \\
 H_1^0 &\rightarrow A^0A^0 \rightarrow b\bar{b}b\bar{b}, & M_{H_1^0} &= 400 \text{ GeV}, \quad M_{A^0} = 80 \text{ GeV}, \\
 H_1^0 &\rightarrow ZA^0 \rightarrow q\bar{q}b\bar{b}, & M_{H_1^0} &= 400 \text{ GeV}, \quad M_{A^0} = 160 \text{ GeV}, \quad (3.3)
 \end{aligned}$$

using a Z'/W' resonance mass of 3300 GeV. These topologies include a $1 \rightarrow 1 + 1 \rightarrow 1 + 3$ asymmetric cascade decay (tb), a $1 \rightarrow 1 + 1 \rightarrow 2 + 2$ cascade decay with different intermediate particle masses (ZA^0) and even six-pronged fat jets ($t\bar{t}$) for which the tagger is not trained. The ROC curves are presented in figure 2. As in the previous cases, the discrimination power is best for jets with light quarks, with a significance improvement up to a factor of 8 for WW . In addition, it is very good for the rest of signals except for a resonance decaying to gg , for which it is not trained. The performance for $t\bar{t}$ is remarkable, especially if one considers that the tagger is trained with up to four-pronged MI data, and a merged $t\bar{t}$ jet has six quarks.

The significance improvement from the tagging of shape variables adds to that gained from a jet mass cut. For illustration, we show in table 2 the significance improvement for

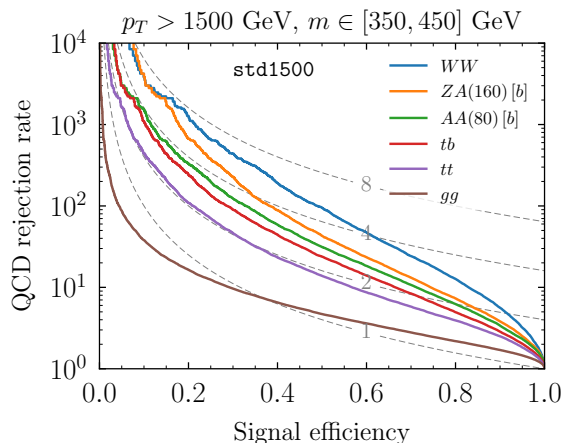


Figure 2. Signal efficiency versus background rejection for the `std1500` anti-QCD tagger. Significance improvement due to shape-variable tagging is indicated by the dashed grey contours.

the signals that is achieved with the cuts $m_J \in [65, 105]$ GeV for $p_{TJ} > 500$ (1000) GeV, as in the `std500` (`std1000`) tagger, and $m_J \in [350, 450]$ GeV for $p_{TJ} > 1500$ GeV, as in the `std1500` tagger. The full significance improvement that can be achieved by the combination of jet mass and shape variables is obtained by multiplying the numbers in table 2 using the ungroomed jet mass with those that can be read from figures 1 and 2. The improvement is modest at low m_J because the ungroomed jet mass distribution for QCD events is large there.

Finally, let us comment about our choice for ungroomed jet masses and p_{TJ} . There are several methods [47–50] to improve signal mass resolution by removing soft particles within the jet. This also tends to improve signal and background separation by shifting the mass spectrum of QCD jets to lower values. However, the results depend on the choice of algorithm and set of parameters, and choices that are optimised for boosted SM particles are often not satisfactory for complex and massive multi-pronged jet topologies such as those considered in this paper. To make this more precise, we compare the significance improvement coming from jet mass cut in table 2, between ungroomed jets and jets trimmed [49] with the parameters $R_{\text{sub}} = 0.2$, $f_{\text{cut}} = 0.05$, commonly used for massive vector bosons W/Z . This algorithm and parameter choice work well for W bosons, as can be observed from table 2, but is too aggressive for many of the multi-pronged boosted objects for which this groomer can significantly broaden and shift the signal peak. For example, for stealth bosons with $p_{TJ} > 500$ GeV (row 3, top panel in table 2) this groomer degrades the mass resolution. For all the complex signals from H_1^0 and H^\pm decays with masses $M = 400$ GeV (lower panel, table 2), this degradation is more pronounced. Jet pruning [48] and soft drop [50] have a similar performance [31]. In any case, the selection of a grooming algorithm and parameter choice that works well for generic BSM objects is another interesting and unrelated issue, which deserves a dedicated study.

	$p_{TJ} > 500$ GeV		$p_{TJ} > 1000$ GeV	
	ungroomed	trimmed	ungroomed	trimmed
W	1.54	2.13	1.52	2.24
$H_1^0 \rightarrow gg$	1.48	1.67	1.50	1.96
$H_1^0 \rightarrow A^0 A^0$ (30) [b]	1.48	1.46	1.46	1.83
$H_1^0 \rightarrow A^0 A^0$ (15) [b]	1.43	1.65	1.44	1.86
$H_1^0 \rightarrow A^0 A^0$ (30) [u]	1.51	1.64	1.54	1.92

	$p_{TJ} > 1500$ GeV	
	ungroomed	trimmed
$H_1^0 \rightarrow WW$	2.64	2.55
$H_1^0 \rightarrow gg$	2.74	2.49
$H_1^0 \rightarrow t\bar{t}$	2.43	1.33
$H^\pm \rightarrow t\bar{b}/\bar{t}b$	2.37	1.99
$H_1^0 \rightarrow A^0 A^0$ (80) [b]	2.53	2.34
$H_1^0 \rightarrow ZA^0$ (160)	2.39	2.07

Table 2. Enhancement of the significance S/\sqrt{B} due to the jet mass cuts $m_J \in [65, 105]$ GeV (top), $m_J \in [350, 450]$ GeV (bottom), with m_J either the ungroomed or trimmed jet mass.

4 Dedicated versus generic taggers

It is naturally expected that a NN jet tagger trained on a specific signal will achieve a better discrimination for that signal than a generic tagger, but it will also have a worse performance than the generic tagger on other types of signals. In order to quantify these statements, we have trained two dedicated taggers:

- (a) Tagger ‘std1000.W’: $p_{TJ} > 1000$ GeV, $m_J \in [65-105]$ GeV, $M_{Z'} = 2200$ GeV, trained on $W \rightarrow q\bar{q}'$ and the QCD background.
- (b) Tagger ‘std1500.WW’: $p_{TJ} > 1500$ GeV, $m_J \in [350 - 450]$ GeV, $M_{Z'} = 3300$ GeV, trained on $H_1^0 \rightarrow WW \rightarrow q\bar{q}'q\bar{q}'$, with $M_{H_1^0} = 400$ GeV and the QCD background.

We show our results in figure 3. On the left panel we can observe that the dedicated tagger `std1000_W` has a slightly better discrimination power than the generic tagger `std1000` for the W bosons it is trained on, but somewhat worse for four-pronged stealth bosons. We find that for final states with W bosons, the performance loss by using a generic tagger is rather small, and is more than compensated by the broader sensitivity to new physics signals. On the right panel it is apparent that the dedicated tagger is significantly better than the generic tagger for the $H_1^0 \rightarrow WW$ signal it is trained on, but it is considerably worse for other signals. Firstly, even though this tagger is specifically trained on a four-pronged signal (WW), its performance on a different four-pronged signal (ZA^0)

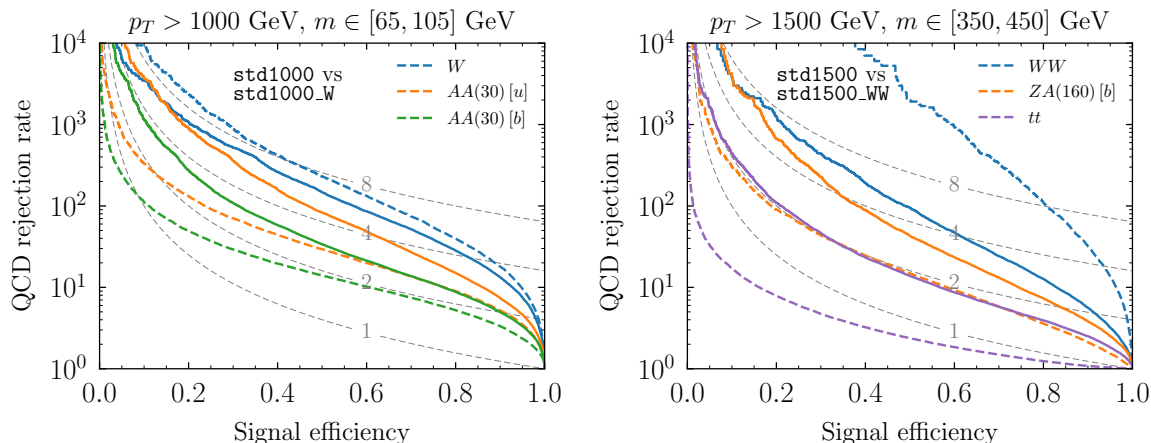


Figure 3. Performance of dedicated taggers (dashed lines) compared to generic taggers (solid lines). Left: a dedicated W -tagger is tested on three signatures at the W -mass. Right: a dedicated 400 GeV WW -tagger is tested on three 400 GeV signatures.

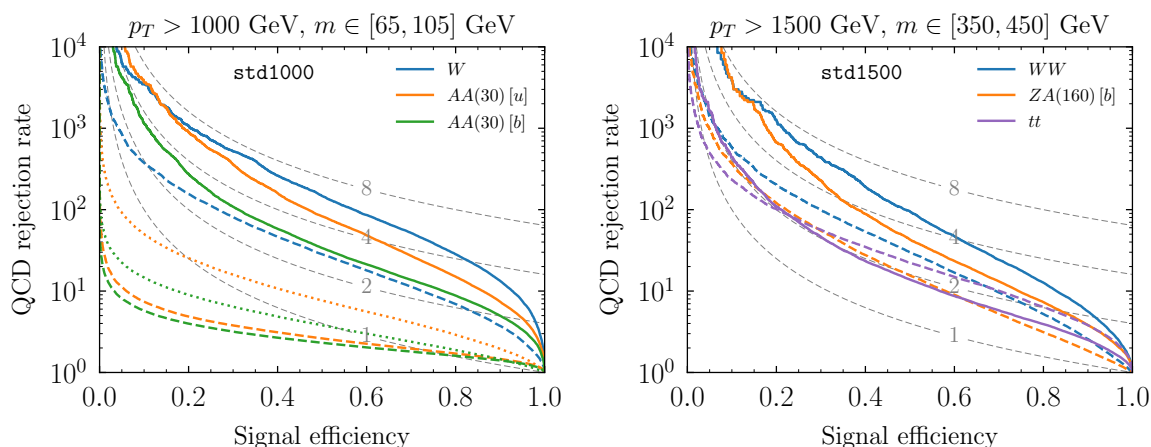


Figure 4. Comparison of generic taggers (solid) to selected τ -ratios (dashed, dotted), for several boosted signals. Left: the generic tagger `std1000` is compared to $\tau_{21}^{(1)}$ (dashed) for W and two stealth boson signatures, and also with $\tau_{42}^{(1)}$ (dotted) for the latter. Right: the generic tagger `std1500` is compared to $\tau_{43}^{(1)}$ for WW and ZA , and $\tau_{63}^{(1)}$ for $t\bar{t}$.

is even worse than with the generic multi-pronged tagger. This fact illustrates that there can be large differences between signals even if they happen to share the same number of prongs, and justifies our choice to train on MI data rather than specific signal models. Secondly, the sensitivity to $t\bar{t}$ is completely degraded by using the dedicated WW tagger, which actually deteriorates rather than enhances sensitivity to this signal.

Generic multivariate taggers are also found to discriminate the various signals from the QCD background better than the simple τ -ratios that have commonly been used in new physics searches. In the left panel of figure 4 we compare the performance of the `std1000` generic tagger (solid lines) to those of $\tau_{MN}^{(1)} = \tau_M^{(1)}/\tau_N^{(1)}$ ratios (dashed) which have been selected for each signal. For W -discrimination we compare with $\tau_{21}^{(1)}$ and for four-pronged

stealth boson signatures we also use $\tau_{42}^{(1)}$, which has also been used for boosted hadronic $H \rightarrow WW^*$ discrimination by the CMS collaboration [53]. In all the three cases, the performance of the generic tagger is much better, but this is especially apparent for stealth bosons, in agreement with previous results [31]. In the right panel we do the comparison for more massive jets using the `std1500` tagger and various selected τ -ratios. Only for a $t\bar{t}$ signal, for which this tagger is not trained, the performances are comparable.

Altogether, the comparison of generic taggers with dedicated ones and simple τ -ratios is very illustrative. For jet masses around the weak boson masses, there is a remarkable improvement for non- W signals with respect to a W -dedicated tagger, keeping nearly the same performance for W bosons, and in all cases, quite an improvement over $\tau_{MN}^{(1)}$. For heavier jet masses, the advantage of a generic tagger is still the broader sensitivity, though the performance of a dedicated tagger can be significantly better.

5 Mass decorrelation

It is desirable, although not compulsory, that a tagger based on jet substructure is decorrelated from the jet mass, in the sense that the tagging efficiency for the background has little dependence on m_J . When this happens, the jet tagging does not shape the m_J distribution of the SM background [52]. This allows for data-driven background estimation using jet mass sidebands, and for the application of bump-hunting strategies on a jet mass distribution. The NNs described in the previous sections, when combined with a choice of threshold on the NN output, act as cuts in the 17-dimensional $\tau_N^{(\beta)}$ space. The distribution of QCD events in $\tau_N^{(\beta)}$ -space varies with both jet mass and p_T , which is illustrated in the first two columns of figure 5. The left column shows a few two-dimensional distributions (there are 136 in total, for the 17 variables considered) in the original $\tau_N^{(\beta)}$ variables, for three intervals of m_J . The middle column corresponds to the same distributions, but for the rescaled variables $\tau_N^{\text{std}(\beta)}$ which were used as NN inputs in the previous section. The efficiency of the tagger (with fixed threshold) on QCD events will necessarily vary with jet mass and p_T , and will result in a sculpting of jet mass distributions in ways that depend sensitively on the mass and p_T of the jets on which it was trained.

In order to build a tagger with an efficiency on QCD jets not varying strongly with jet mass or p_T , there are three obvious possibilities. The simplest one is to apply to the NN output the approach utilised already in [25]. In this case, the threshold on the NN output would be adjusted with jet mass and p_T , in such a way that background rejection is fixed. This approach has many advantages (first and foremost being simplicity), but a tagger used in this way that is optimised for signal discrimination at one mass will tend to have suboptimal performance for signals at different masses as the shapes of the input distributions vary, and the tagger might sculpt signal shapes and shift signal mass peaks. It might be required that a suite of taggers are trained, optimised at different mass points. This problem could be ameliorated if a basis of variables is found which are only weakly correlated with jet mass and p_T . A second approach, to be adopted in this section, involves preprocessing the $\tau_N^{(\beta)}$ variables in such a way that the QCD distributions of the transformed input variables no longer exhibit strong dependence on mass or p_T . This will introduce greater complexity in an experimental analysis which grows with the number of

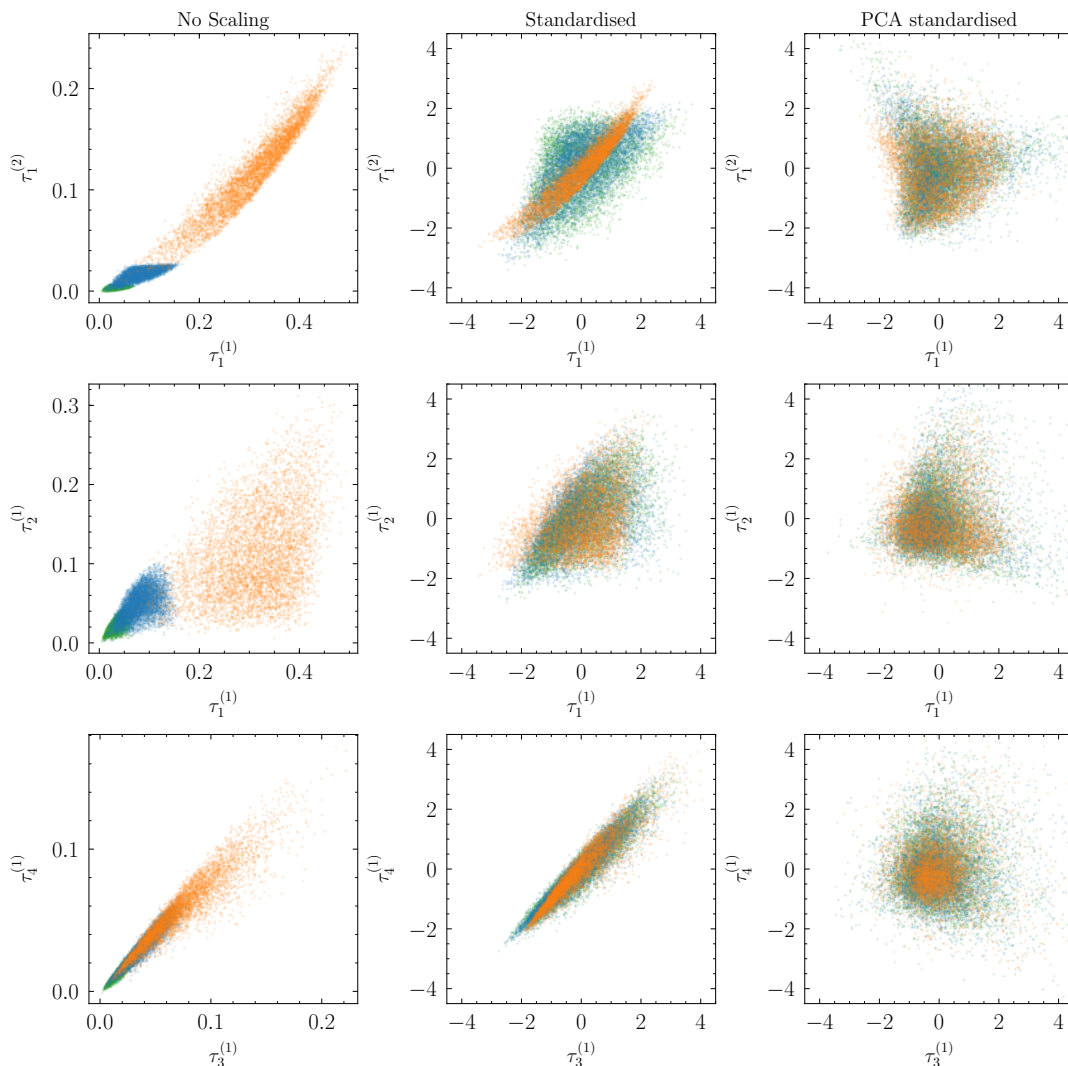


Figure 5. N-subjettiness variables for QCD, plotted in three mass windows and with three different levels of processing. Green: $35 \text{ GeV} < m_J < 40 \text{ GeV}$, blue: $80 \text{ GeV} < m_J < 85 \text{ GeV}$, orange: $265 \text{ GeV} < m_J < 275 \text{ GeV}$. Left column: bare τ variables. Middle column: standardised τ variables, as described in section 3. Right column: standardised also along principal component axes. Each row is a different pair of τ variables.

input variables, but will have the advantage that a single tagger can be used with good signal discrimination over a wide range of masses and p_T . A third possibility would be to build a tagger that can learn to vary the region of $\tau_N^{(\beta)}$ -space to cut as a function of jet mass. This was achieved in ref. [51] using an adversarial strategy designed to maintain mass decorrelation on QCD jets. This would leave open the question, however, of how to sample signal masses in training, in such a way that the tagger is not biased towards particular signal masses.

Let us consider a set of $\tau_N^{\text{std}(\beta)}$ calculated using QCD jets selected within a certain jet mass and p_T bin. Arranging the 17 $\tau_N^{\text{std}(\beta)}$ variables into a 17-dimensional vector $\vec{\tau}^{\text{std}}$, we

define the following transformation for the $\tau_N^{\text{std}(\beta)}$ (in that bin)

$$\vec{\tau}^{\text{std}} \rightarrow \vec{\tau}^{\text{PCA}} = R^{-1}SR\vec{\tau}^{\text{std}} \quad (5.1)$$

with R and S being 17 by 17 square matrices. R is a rotation matrix that diagonalises the symmetric covariance matrix calculated from this τ -set. This matrix induces a rotation into a basis aligned with the principal component axes of the dataset. In this basis, all pairs of variables are linearly uncorrelated. This is equivalent to choosing a basis whose axes lie along the principal axes of a rigid body formed out of this distribution. We then standardise the data along these axes, so that along each principal axis the standard deviation of the data is 1. This is the action of the diagonal matrix S . We then invert the principal axis rotation with the action of R^{-1} .³ In practice, data should be binned according to jet mass and p_T , and a transformation matrix $M_i = R_i^{-1}S_iR_i$ must be determined for each bin i . Alternatively, one could define the PCA rescaling as a continuous function of p_T and m_J which could be fitted to binned data. In the third column of figure 5 we plot τ^{PCA} distributions for QCD. Firstly, it can be seen that much of the variation in these distributions with jet mass has been eliminated by the transformation. Second, thin directions have been stretched and fat directions have been squashed, as can be seen most clearly in the third row. This fact can aid in the training of the NN.

Therefore, the PCA tagger involves two different tasks:

1. To set up a transformation map between τ and τ^{PCA} , which requires a binning of MC data for the QCD background in m_J and p_{TJ} . This map is used both when training the NN (with signal and background events) and when applying the tagger to test data.
2. To train the NN using τ^{PCA} variables in some interval of m_J and p_{TJ} , which might only be a subset of the entire domain of the transformation map.

In order to test whether a tagger trained on input data with this preprocessing will sculpt QCD jet mass distributions, we generate as test data 1,081,834 QCD jets (evenly split between gluon and quark jets) with $p_T > 1000$ GeV, and with no mass cut. The jet mass distribution for this data is given by the solid black lines in figure 6. For the PCA preprocessing of the $\tau_N^{(\beta)}$ variables, we bin the data by jet mass with variable bin sizes (as indicated by the bin widths in figure 6), in order to have similar numbers of events in each bin, and define a PCA transformation for each bin calculated from this data in that bin.

An additional sample of QCD jets, generated in the same manner as above, is set aside for use in training two new taggers. These taggers are trained on the τ^{PCA} values of QCD and MI data selected only in a mass window, indicated by the shaded boxes in figure 6, to investigate if they will sculpt the QCD jet mass distribution around those windows and

³The reason for applying R^{-1} is as follows. R has a permutation ambiguity, and it is natural to choose this so that the eigenvectors are ordered by the corresponding eigenvalues of the covariance matrix. However, if there is an eigenvalue crossing between two adjacent bins, this will cause a discontinuity in the rescaling matrix SR which would spoil mass decorrelation. The action of R^{-1} removes the permutation ambiguity from the $\vec{\tau}^{\text{PCA}}$.

	PCA1000_80	PCA1000_200	PCA500_80
Training sample size	108,958	71,466	126,372
Validation sample size	27,238	17,866	31,592

Table 3. Training sample sizes for PCA taggers.

if they will still be sensitive to new physics signals outside of those windows. The cuts implemented on the training data and the parameters for the generation of the MI training data for these taggers are

- (a) Tagger ‘PCA1000_80’: $p_{TJ} > 1000$ GeV, $m_J \in [65 - 105]$ GeV, $M_{Z'} = 2200$ GeV, $M_S = 80$ GeV.
- (b) Tagger ‘PCA1000_200’: $p_{TJ} > 1000$ GeV, $m_J \in [170 - 230]$ GeV, $M_{Z'} = 2200$ GeV, $M_S = 200$ GeV.

The sizes of the event samples used for training these taggers are given in the first two columns of table 3. The solid coloured lines in the first two rows of figure 6 indicate the jet mass distribution for the QCD test sample after selection by the taggers at varying thresholds. We find that there are no new spurious features introduced by application of either tagger.

In order to test the sensitivity of these taggers to boosted resonance signals at different masses we simulate the following two signals,

$$\begin{aligned}
 H_1^0 \rightarrow A^0 A^0 \rightarrow b\bar{b}b\bar{b}, \quad M_{H_1^0} = 100 \text{ GeV}, \quad M_{A^0} = 40 \text{ GeV}, \\
 H_1^0 \rightarrow WW \rightarrow q\bar{q}'q\bar{q}', \quad M_{H_1^0} = 200 \text{ GeV},
 \end{aligned}
 \tag{5.2}$$

resulting from the decay of a 2.2 TeV resonance. The dashed lines in figure 6 indicate the results when these signals are injected into the QCD test sample, re-weighted to correspond to 1.2% and 0.7% of the size of the QCD sample, respectively. We see that both taggers not only succeed in not sculpting the QCD jet mass distribution, but they are also sensitive to BSM boosted objects outside of the mass range in which they were trained.

We also wish to test the effect of using a tagger in a p_T region in which it was not trained. We therefore generate QCD data in the range $500 \text{ GeV} < p_T < 1000 \text{ GeV}$, binned in jet mass in the same way as the $p_T > 1000$ GeV data above. The τ distributions of this data determine the PCA transformations for data falling into these bins. We also generate MI data on which to train the following tagger,

- (c) Tagger ‘PCA500_80’: $p_{TJ} > 500$ GeV, $m_J \in [65 - 105]$ GeV, $M_{Z'} = 1100$ GeV, $M_S = 80$ GeV.

The size of the event samples used for the training is given in the third column of table 3. This tagger is then applied to the data described above in the $p_T > 1000$ GeV bins. The results are shown in the third row of figure 6. We find that the performance of the tagger is not greatly sensitive to the p_T and mass spectrum of jets used to train the tagger, so long as the data has been properly standardised along the principal component axes.

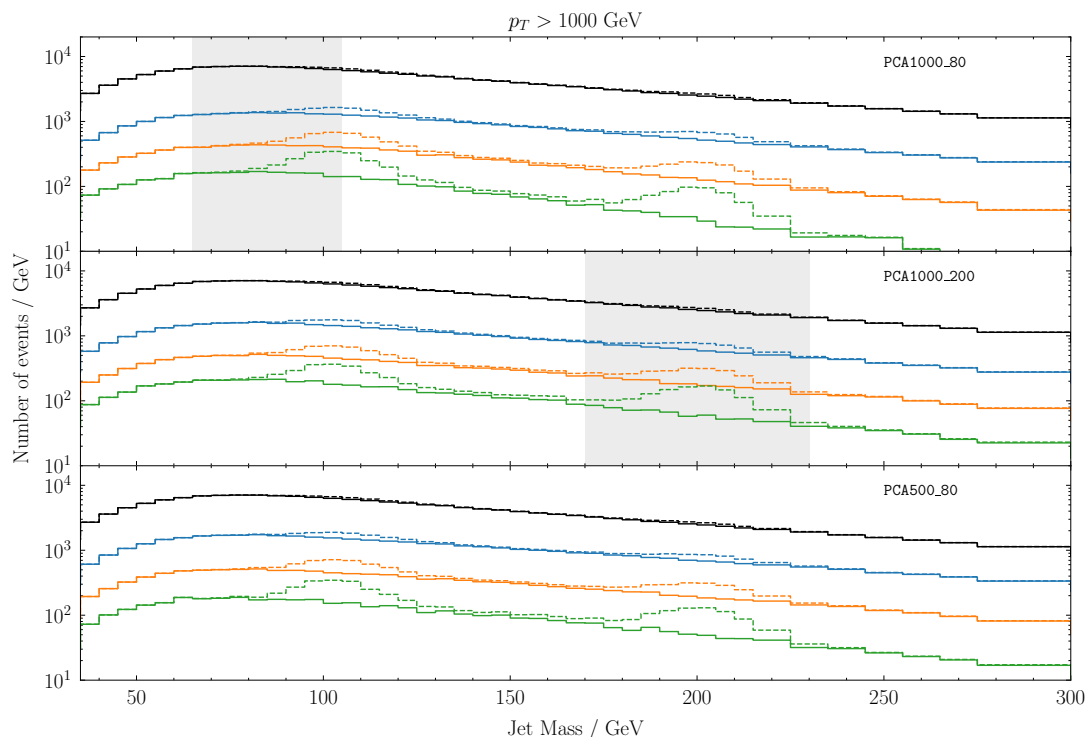


Figure 6. Jet mass distributions for $p_T > 1000$ GeV, selected with three taggers at various thresholds. Top: PCA1000_80 tagger. Middle: PCA1000_200 tagger. Bottom: PCA500 tagger. The solid lines correspond to the QCD jet background, and the dashed lines to the background plus injected signals. The shaded boxes show the jet mass intervals for which each tagger is trained.

6 Discussion

The generic anti-QCD taggers we have developed in this work provide an alternative to usual taggers in LHC searches for new physics in the boosted regime, with the main advantage being their broad sensitivity to multi-pronged boosted signatures. This feature is of great interest as we do not yet know how new physics might manifest at the LHC. Indeed, new relatively light particles beyond the SM might exist and be produced with very high boosts, for instance if they result from the decay of a heavier particle. If these particles decay hadronically then their signature is a single massive fat jet which might be difficult to separate from QCD backgrounds with existing tools.

A generic anti-QCD tagger entails a compromise between a high rejection of the QCD background and a broad sensitivity to a variety of signals. As we have shown, a dedicated tagger has a better performance for the specific signal it is trained on, but it can be rather blind to other types of signals. In particular,

1. For jet masses around the weak boson masses, there is a remarkable improvement for BSM boosted signals (exemplified by stealth bosons) with respect to a W -dedicated tagger analogous to the one in ref. [13], while keeping nearly the same performance for W bosons. Both for W and stealth bosons, the generic tagger provides quite an improvement over the simple ratio $\tau_{21}^{(1)}$ often used in experimental analyses.

2. For heavier jet masses of a few hundreds of GeV, the advantage of a generic tagger is still the sensitivity to several multi-pronged signals, though the performance of a dedicated tagger can be significantly better.

In either case, final states involving several b quarks are harder to distinguish from the QCD background than those involving light quarks, but b tagging could also be used as an additional independent tool. Overall, we observe that searches for new resonances would greatly benefit from a generic tagger for hadronic boosted objects, perhaps complementing dedicated ones. (Dedicated taggers also have their place in specific analyses where one is not interested in other possible signatures, for example in $t\bar{t}$ measurements in the boosted regime.)

A simple application of a generic tagger of this kind would be an extension of the existing searches for diboson resonances, which search for a resonance bump in a di-jet invariant mass distribution. The use of a generic tagger would allow to search for resonances decaying to a boosted SM boson and a boosted BSM boson. In this case, leptonic decays could be selected for the SM boson and the recoiling fat jet might be selected in a series of broad mass windows after selection by a generic tagger trained in each window. Alternatively, hadronic decays could also be selected for the SM boson, using standard tagging criteria. One recent example is given by ref. [54], which looks for XH decays of a heavy resonance, selecting $H \rightarrow b\bar{b}$ for the Higgs boson and a two-pronged decay $X \rightarrow q\bar{q}$ for X , with a set of overlapping mass windows for the new particle X and a standard tagger $D_2^{(\beta=1)}$. In this case, a generic tagger could be used to provide sensitivity not only to $X \rightarrow q\bar{q}$ but to other topologies as well. A search could also be carried out for di-BSM bosons, requiring both bosons to have similar mass, and doing a scan over a series of broad mass windows.

Going beyond the discrimination of various signals against the QCD background, it may also be desirable to have a fixed background rejection as a function of the jet mass, for example to allow for data-driven background estimation using jet mass sidebands, and for the application of bump-hunting strategies on a jet mass distribution. This is a solved problem, and can be achieved by applying existing decorrelation techniques to the NN output. However, doing this in such a way as to also maintain good sensitivity to signals over a broad range of masses with a single tagger and without signal-mass bias is a more difficult problem. We have demonstrated that an approach based on standardising along the principal component axes gives satisfactory results in simulation, which is implemented by building a ‘transformation map’ in the two-dimensional plane of m_J and p_{TJ} , using Monte Carlo simulation of the QCD background. This map relates the N -subjettiness variables $\tau_N^{(\beta)}$ to the PCA-scaled ones $\tau_N^{\text{PCA}(\beta)}$, which are the inputs to the tagger. This relation varies with the jet mass and p_T and, in practice, it is enough to consider suitable bins in m_J and p_{TJ} . This can be considered as an extension of the approach which has already been taken in a CMS search for light resonances decaying to quark pairs [25] to decorrelate the jet substructure tagger from the jet mass. In our case, the tagger is trained at some given m_J and p_{TJ} intervals, and it can be subsequently applied outside these intervals by using the map of transformations between $\tau_N^{(\beta)}$ and $\tau_N^{\text{PCA}(\beta)}$ for other values of m_J and p_{TJ} . As figure 6 demonstrates, this procedure is quite effective. And, in particular, one

does not need to train the tagger with various new physics signals at different jet masses and transverse momenta; only the QCD background prediction needs to be known in order to determine the transformations from $\tau_N^{(\beta)}$ to $\tau_N^{\text{PCA}(\beta)}$ at that jet mass and transverse jet momentum.

Although the number of variables (17×17 for the correlation matrix) used here for the transformation map of PCA-scaled taggers seems a formidable task for an experimental analysis, let us point out that a simpler approach will suffice. First, a five-body tagger nearly has the same performance, as seen in appendix A, which reduces the number of variables to 11×11 . Second, some optimisation by reducing the number of variables may be performed too, without sacrificing the performance. Indeed, in this work our goal has been to provide a proof of concept that anti-QCD taggers can be built, leaving the optimisation for future analyses.

Either in its simplest versions (as in section 3) with standardised input, or in its mass-decorrelated versions (as in section 5) with PCA scaling, a generic anti-QCD tagger is a novel tool, whose implementation seems feasible, and which could greatly benefit experimental analyses. The final goal is quite ambitious: to enlarge the scope of new physics searches with SM boosted objects, so as to be sensitive to new physics yielding BSM boosted objects. This will constitute a leap forward in new physics searches at the energy frontier, and is well worth the effort.

Acknowledgments

We would like to thank Siddharth Mishra-Sharma and Mathieu Cliche for discussions about the implementation of neural networks and python in the early stages of this work. We also thank Rafael Teixeira De Lima, Petar Maksimovic, Nhan Tran, Jesse Thaler and Matt Eelsey for discussions and comments on the manuscript. The UFO file used to generate the data used in appendix G was generously provided by Bogdan Dobrescu and Felix Yu. The work of JAAS is supported by MINECO Projects FPA 2016-78220-C3-1-P and FPA 2013-47836-C3-2-P (including ERDF), and by Junta de Andalucía Project FQM-101. The work of JHC and RKM is supported by NSF under Grant No. PHY-1620074 and by the Maryland Center for Fundamental Physics (MCFP).

A How much information is in a multi-pronged jet?

The number of N -subjettiness observables that provide additional information on the substructure of multi-pronged jets is found empirically, as in ref. [13], by training different taggers with $M = 2, 3, \dots$ in eq. (2.1) and comparing the ROC lines obtained for the signals of interest. In that reference, for two-pronged jets it was found that the results do not improve beyond $M = 4$. Because we also consider three- and four-pronged jets, we have found it is enough to consider $M = 7$. For illustration, we show in figure 7 the performance of different versions of the `std1000` tagger taking $M = 2, \dots, 7$, applied to a stealth boson signal $H_1^0 \rightarrow A^0 A^0 \rightarrow b\bar{b}b\bar{b}$ with $M_{H_1^0} = 80 \text{ GeV}$, $M_{A^0} = 30 \text{ GeV}$. For a given M , the input to the NN for the corresponding tagger is a $3M - 4$ dimensional vector, as defined in ref. [13]. The ROC curves indicate that the performance saturates before $M = 7$.

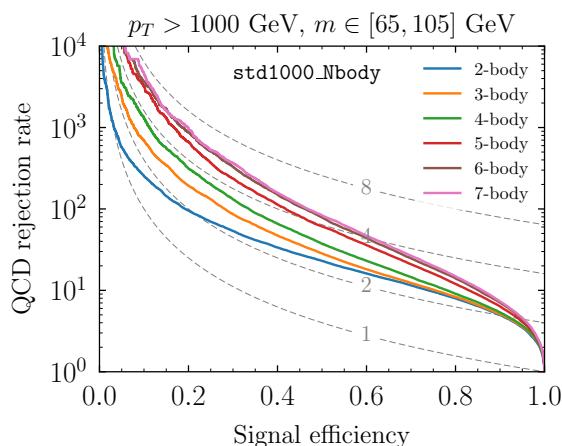


Figure 7. Signal efficiency versus background rejection for different versions of the `std1000` tagger taking $M = 2, \dots, 7$ in eq. (2.1).

B Effect of signal composition on training

The shape of a jet from a heavy quark such as a b quark is in general different from that of gluons and light quarks. We have included light and b quark jets in our MI data earlier, in an attempt to capture all possible shapes, but for simplicity we have not included gluons. In this appendix we show how the results are affected if (i) one doesn't include b quarks in the training data, or (ii) if one also includes gluons. For each case, we train taggers with the MI data set modified — for case (i) we use the subset of processes in eq. (2.3) that do not involve b quark in final state, while for case (ii) we use all the processes in eq. (2.3) and in addition we add the process $H_1^0 \rightarrow gg$. In all cases, we continue to use equal numbers of events for each of the three or seven categories of training signal data. We perform these studies in two kinematic regimes corresponding to those used for the `std1000` and `std1500` taggers in section 3:

- (a) $p_{T,J} > 1000$ GeV, $m_J \in [65 - 105]$ GeV, $M_{Z'} = 2200$ GeV, and
- (b) $p_{T,J} > 1500$ GeV, $m_J \in [350 - 450]$ GeV, $M_{Z'} = 3300$ GeV,

We test the performance of the taggers on the signal processes

$$\begin{aligned}
 H_1^0 &\rightarrow gg, & M_{H_1^0} &= 80 \text{ (400) GeV}, \\
 H_1^0 &\rightarrow A^0 A^0 \rightarrow b\bar{b}b\bar{b}, & M_{H_1^0} &= 80 \text{ (400) GeV}, & M_{A^0} &= 30 \text{ (160) GeV}, \\
 H_1^0 &\rightarrow A^0 A^0 \rightarrow u\bar{u}u\bar{u}, & M_{H_1^0} &= 80 \text{ (400) GeV}, & M_{A^0} &= 30 \text{ (160) GeV}.
 \end{aligned} \tag{B.1}$$

The masses indicated for H_1^0 and A^0 correspond to cases (a) and (b) respectively. The results are shown in figure 8 (top panel for b quarks and bottom panel for gluons).

Focusing first on the case of inclusion of b quarks in the training data, we find that for decays with only light (u) quarks in the final state inclusion of b quark MI data has no effect on tagging performance. For other decays which include b quarks or gluons in the

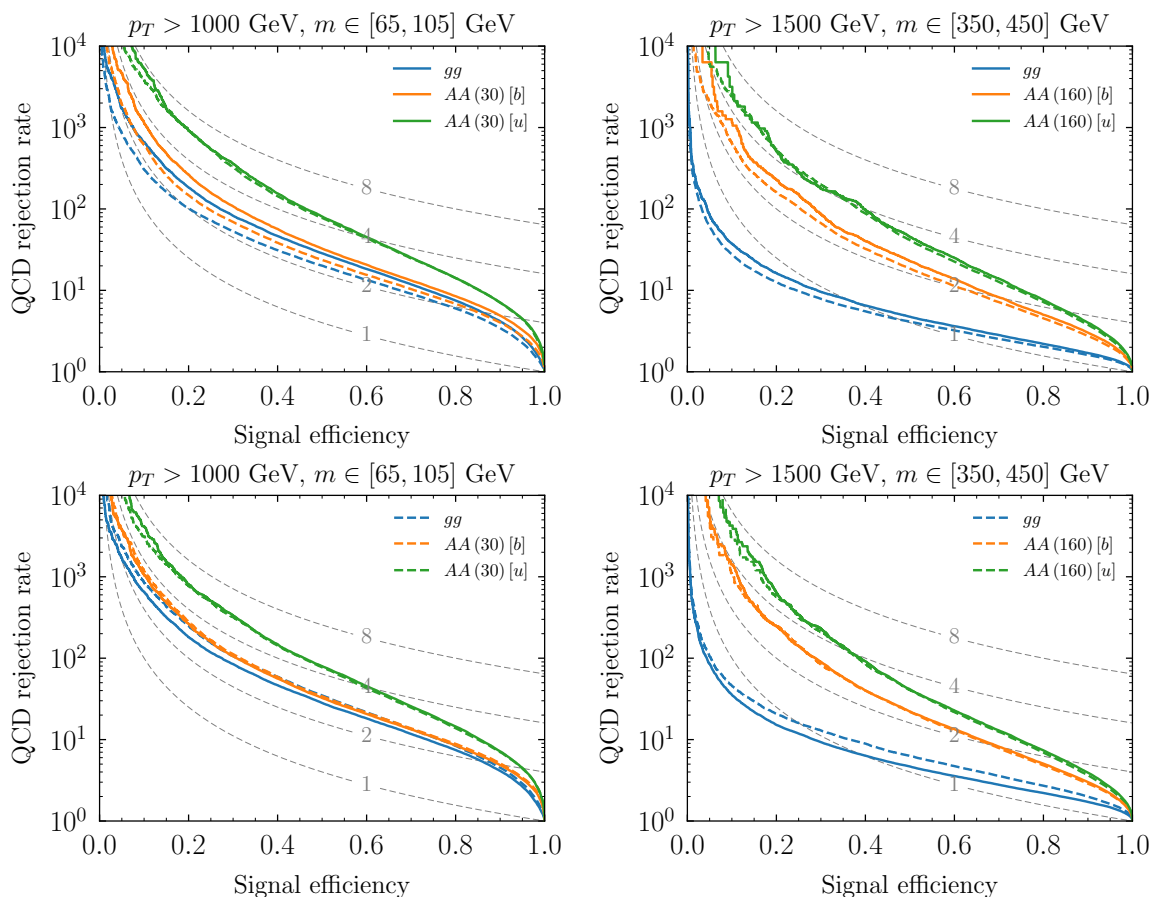


Figure 8. Top: effect of b quarks on training, on various test signals. Solid lines correspond to standard choice of MI data for training, and dashed lines correspond to MI data without b quarks in the final state for training. Bottom: effect of gluons on training, on various test signals. Solid lines correspond to standard choice of MI data for training while dashed lines correspond to MI data as well as gluons in the final state for training.

final state, taggers trained with b quarks do marginally better. Secondly, for the case when (gg) -jets are added to the training data, we find that including this process marginally improves the discrimination power for (gg) -jets in both kinematic regimes studied. For other processes that have b or u quarks in the final state, the inclusion of these jets in training has a negligible effect on the performance.

C Background composition: effect of quark to gluon ratio

For simplicity, in the training of our taggers and their testing on Monte Carlo data, we have assumed that the background is composed of equal parts of quarks and gluons. This is obviously not the case in a real analysis, in which the relative ratio will depend not only on the final state considered, but also on the energies involved. In this appendix we explore how sensitive the results are to the precise ratio of quarks and gluons.

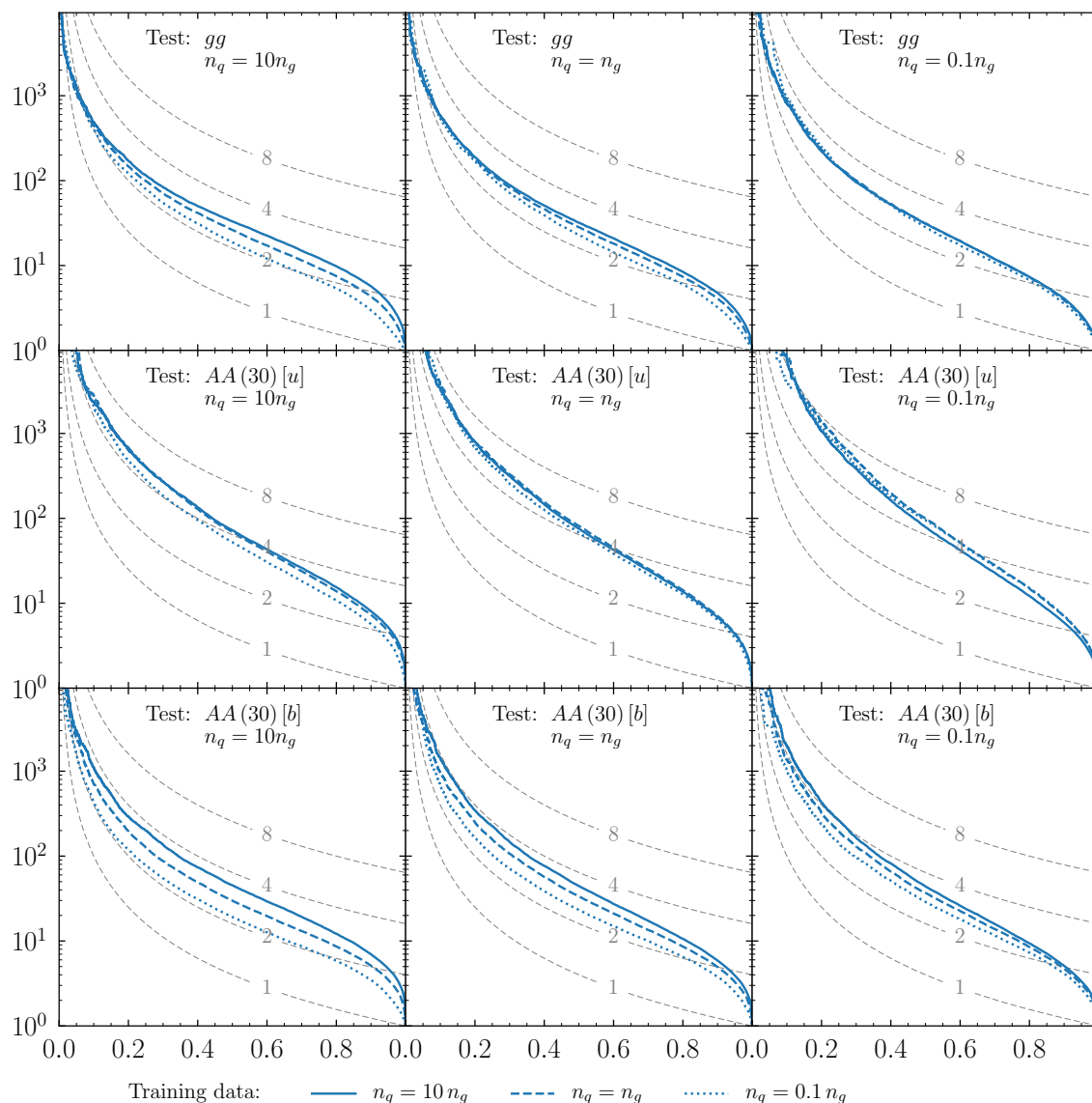


Figure 9. Effect of QCD quark to gluon ratio in training and test data. The performance of the different taggers (solid, dashed and dotted lines) is shown for several signals (in rows) and for different choices of quark to gluon ratios in the background test data (in columns).

We focus on $p_{T,J} > 1000 \text{ GeV}$, $m_J \in [65, 105]$, as considered for the `std1000` tagger, and train three taggers on the MI data in eq. (2.3) and the QCD background, with three ratios of quarks and gluons: $n_q = 10n_g$, $n_q = n_g$, $n_q = 0.1n_g$, corresponding to the solid, dashed and dotted lines in figure 9. We test these taggers for several signals, with a background composed of the same three ratios of quarks and gluons: $n_q = 10n_g$ (left column), $n_q = n_g$ (middle column), $n_q = 0.1n_g$ (right column). The signal processes

considered are

$$\begin{aligned}
 H_1^0 &\rightarrow gg, & M_{H_1^0} &= 80 \text{ GeV}, \\
 H_1^0 &\rightarrow A^0 A^0 \rightarrow u\bar{u}u\bar{u}, & M_{H_1^0} &= 80 \text{ GeV}, & M_{A^0} &= 30 \text{ GeV}, \\
 H_1^0 &\rightarrow A^0 A^0 \rightarrow b\bar{b}b\bar{b}, & M_{H_1^0} &= 80 \text{ GeV}, & M_{A^0} &= 30 \text{ GeV}.
 \end{aligned}
 \tag{C.1}$$

The main conclusion of this comparison is that the results actually do not depend much on the precise background composition, as seen from a glance at figure 9. In some cases the relative performance of the three taggers is as expected: for example, for the second process in (C.1) above, the tagger is (marginally) better when the background composition is the same in training and testing. But this is not the case for the third process in (C.1). For example, for $n_q = 0.1 n_g$, the tagger trained with the ‘inverse’ ratio $n_q = 10 n_g$ is slightly better. This suggests that changing the background composition also affects the way in which the tagger learns what is signal and what is background.

D NN architecture

The choice of architecture in any NN problem merits its own study. Throughout this paper we have used an architecture that gives robust results against variations in its depth (number of nodes in a hidden layer) and breadth (number of hidden layers). In this appendix we show that the results are very insensitive to variations on this choice. Apart from the architecture considered for our results in sections 3–5 (two fully connected hidden units with 512 nodes and 32 nodes respectively, henceforth referred as 512–32), we consider here two more architectures — 1024–32 and 512–512–32, in a self explanatory notation.

We consider the mass and $p_{T,J}$ ranges used in the definition of the `std1000` and `std1500` taggers, and train two taggers on the MI processes in eqs. (2.3), with Z' masses chosen as in section 3, and the QCD background. We test the first tagger on stealth bosons with masses $M_{H_1^0} = 80 \text{ GeV}$, $M_{A^0} = 30 \text{ GeV}$, and the second tagger with masses $M_{H_1^0} = 400 \text{ GeV}$, $M_{A^0} = 160 \text{ GeV}$. The results are shown in figure 10. We see practically no difference in the performance as the NN architecture is varied.

E Is the tagger learning shape or kinematics?

Although it has been shown that the taggers can efficiently discriminate various multi-pronged signals from the QCD background, a question remains whether this discrimination is solely based on jet shapes or there is also some effect from the different kinematics of the signals and the background. For example, we have already mentioned that the heavy Z' and W' resonance masses have been chosen in such a way that the $p_{T,J}$ distributions are similar to the background, but still there are some differences, which can be seen in figure 11 (left), between the distributions of the QCD background and two sample signals, W and stealth bosons. The same can be said about the jet mass, shown in the right panel: while the background distribution is rather flat, the signals concentrate around and slightly above the input resonance mass.

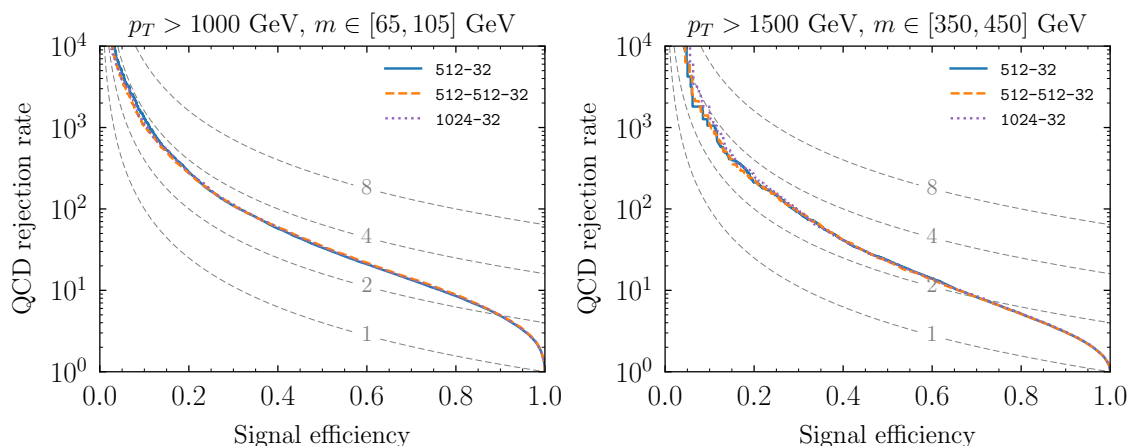


Figure 10. Performance of the taggers for different network architectures.

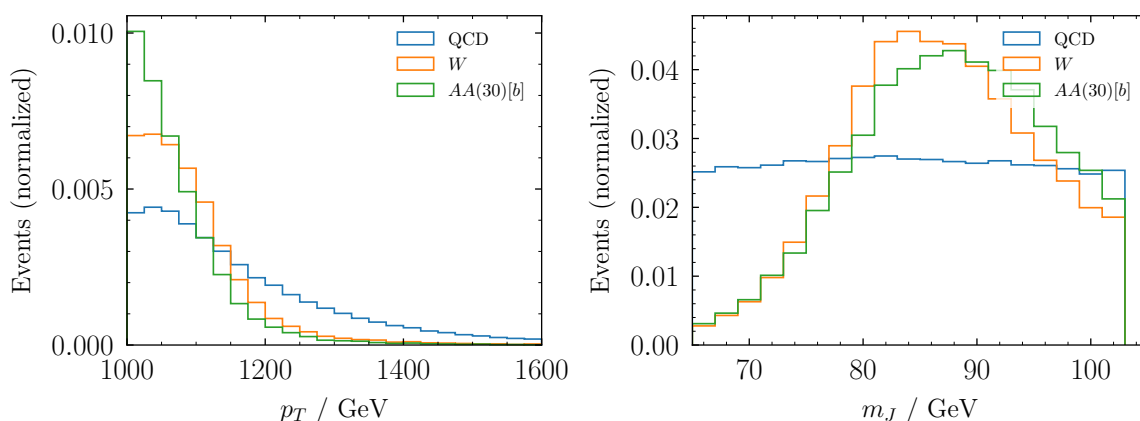


Figure 11. Transverse momentum (left) and jet mass (right) of the QCD background and two of the signals used to test the `std1000` tagger.

We have tested the effect of the different $p_{T,J}$ and m_J dependence by considering the discrimination of these two signals with reweighted distributions, using the `std1000` tagger. (The reweighting of the signals makes them have the same two-dimensional $(p_{T,J}, m_J)$ signal distributions as the background.) With this purpose, a two-dimensional binning in $p_{T,J}$ and m_J is applied, with 25 GeV bins in $p_{T,J}$ and 5 GeV bins in m_J . The ranges of these variables are restricted to $p_{T,J} \in [1000, 1250]$ GeV and $m_J \in [75, 105]$ GeV, in order to avoid the appearance of a few events with too large weights that might bias the results. As can be seen from figure 11, still within those intervals there is significant variation of these two variables.

The comparison between the ROC curves for the signals with the original and reweighted distributions, in both cases restricted to the mentioned $p_{T,J}$ and m_J intervals, is presented in figure 12. The left panel shows the results for W bosons, also including the curves for $\tau_{21}^{(1)}$, and the right panel shows the results for stealth bosons. In both cases we observe that the differences between the results with the original and re-weighted distributions are very small. Also, the ROC curves without re-weighting (but with restricted

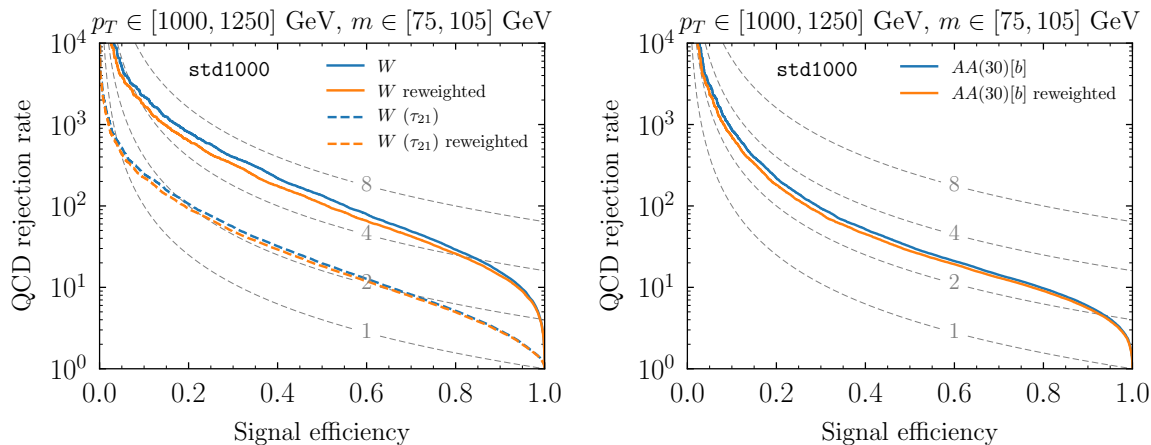


Figure 12. Comparison between signal efficiency versus background rejection for W boson (left) and stealth boson (right) signals, with their original and with re-weighted $p_{T,J}$ and m_J distributions.

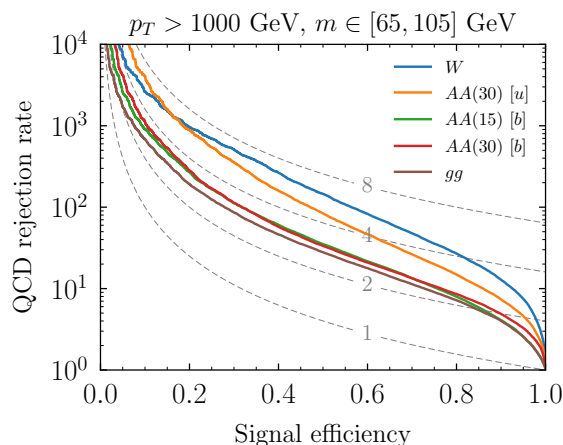


Figure 13. Signal efficiency versus background rejection for a variant of the `std1000` tagger which does not use $\tau_1^{(2)}$. The performance is identical to that of the `std1000` tagger in figure 1.

$p_{T,J}$ and m_J range) can be compared to those in figure 1 (right), to see that they are very similar. Overall, it is found that the influence of kinematics in the tagger learning, if any, is quite small.

Besides, we note that the variable $\tau_1^{(2)}$ which is an input to our taggers is closely related to $(m_J/p_T)^2$ [55]. From the discussions in section 5, it is our objective to avoid as much as possible jet mass and p_T being directly used as discriminating variables by the tagger. One may wonder whether this variable should have been excluded from our set in eq. (2.2). In order to test its influence on our results, we train a variant of the `std1000` tagger on all 7-body variables except $\tau_1^{(2)}$ and compare its performance to the `std1000` tagger in figure 13; we find no effect on the tagger performance. This can be understood because the leading dependence $\tau_1^{(2)} \sim (m_J/p_T)^2$ is the same for all signals and backgrounds at a given mass and $p_{T,J}$. Therefore, the standardisation of the inputs erases this dependence to a very large extent.

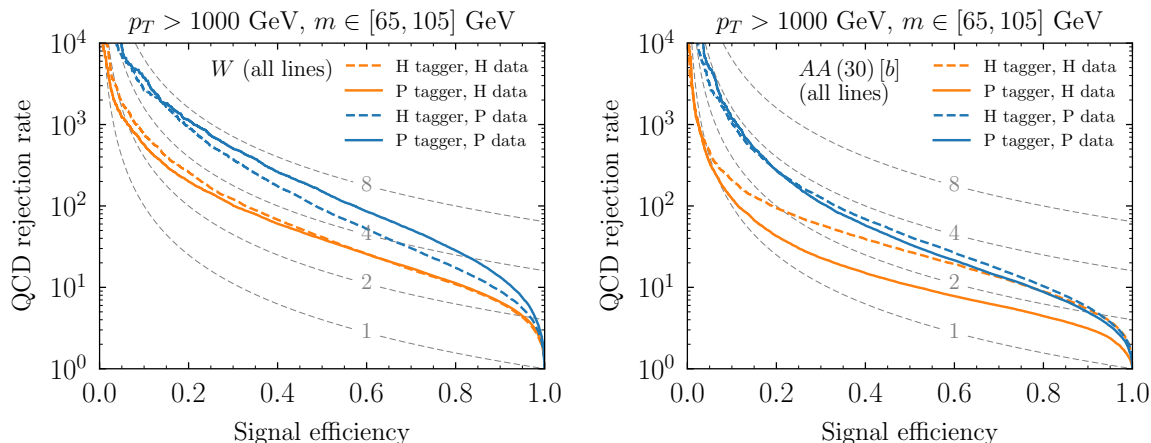


Figure 14. Comparison of results for anti-QCD taggers using Herwig and Pythia simulated showers, for W bosons (left) and stealth bosons (right).

F Pythia versus Herwig

A serious challenge in the application of machine learning to jet physics in a real collider experiment is the question of whether the distributions of substructure variables are correctly modelled by simulation, and whether the performance of the tagger is robust under mismodelling. Designing approaches to bypass mismodelling fragility is an active area of research [56–58], but beyond the scope of this work. In this appendix, we restrict ourselves to investigating the variation of tagging performance when using data hadronised with Pythia (as used for our results in sections 3–5) and Herwig.

We focus on $p_{TJ} > 1000 \text{ GeV}$, $m_J \in [65, 105]$, as used in the `std1000` tagger, for brevity. Two new taggers are trained on all processes in eqs. (2.3) and the background; one of them is trained on data using Pythia and the other one with data using Herwig. We test the performance of the two taggers on W bosons and stealth bosons with $M_{H_1^0} = 80 \text{ GeV}$, $M_{A^0} = 30 \text{ GeV}$. This test data (both signal and background) is generated twice, once with Pythia and once with Herwig. We show in figure 14 the results for in-sample tests (e.g. a Pythia trained tagger tested on Pythia data) as well as out of sample tests (e.g. a Pythia trained tagger tested on Herwig data).

We find that in general the performance is better on Pythia generated data than on Herwig generated data, though for the most part this is largely independent of which data the tagger was trained on. The exception is that the tagger trained on Pythia data has significantly worse performance on the Herwig data for stealth bosons, compared to the tagger trained on Herwig data. The differences between the Herwig and Pythia curves appear to arise mostly from the different modeling of the higher-order $\tau_N^{(\beta)}$, as suggested by figure 15 (left). This figure shows the ROC curves corresponding to $\tau_1^{(1)}$ and $\tau_4^{(1)}$ alone for the stealth boson signal, using Pythia and Herwig data. It is clearly seen that the differences due to showering are much more pronounced for $\tau_4^{(1)}$. This argument is confirmed by the study of the discriminating power of the $\tau_{21}^{(1)}$ ratio in figure 15 (right). Both for W and stealth bosons the discriminating power is similar using Pythia and Herwig data.

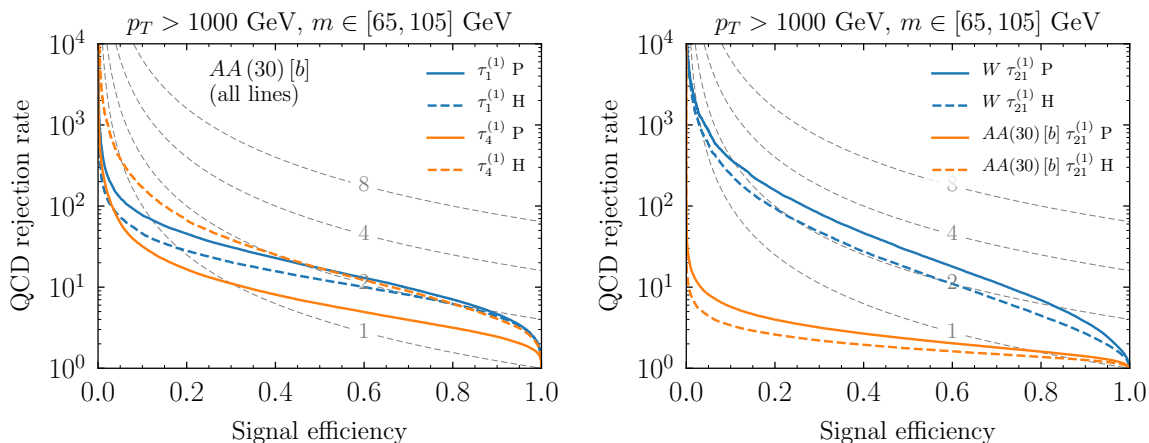


Figure 15. Comparison of results for selected $\tau_N^{(\beta)}$ (left) and the $\tau_{21}^{(1)}$ ratio (right) using Herwig (H) and Pythia (P) simulated showers.

Because it is of the utmost importance that the performance of the tagger on QCD data be very well understood, it might be best to train such a tagger with real QCD data, and especially, test its performance directly on data in suitable control regions. Significant uncertainties on signal efficiencies may remain, but these are less important than having an accurate prediction for the background.

G Boosted coloured jets

In this paper we have focused on jets resulting from the boosted decays of colour-singlet new particles, which might easily be missed in searches looking for their direct production due to a small production cross section. Light coloured particles have large production cross sections, and searches for signatures resulting from their direct pair production via QCD are typically highly constraining. However, in ref. [59] it was noted that for some decays of such particles, for example a vector-like quark (VLQ) decaying via a non-renormalisable operator into three light quarks, there are no meaningful LHC constraints from direct searches for these particles with masses between 100 GeV and 1000 GeV. For masses as low as a few hundred GeV, passing the LHC thresholds for jet-based searches may require these VLQs to be produced with high momentum, resulting in collimation of their decay products into a fat jet. Therefore, and also for completeness, it is of interest to see if the generic tagger which we have trained only on colour-singlet jets is sensitive also to these coloured jets. Further, this is the only three-pronged signal which we test our tagger on.

We simulate the pair production of VLQs $T\bar{T}$ with $T \rightarrow cc\bar{c}$ using an UFO file generously provided by the authors of ref. [59], setting the T mass to 400 GeV. We impose a generation level cut $H_T > 3$ TeV, where H_T is the scalar sum of p_T of the quark decay products. The detector level selection is made in the same way as described for the signals used to test the tagger `std1500` in section 3, selecting the hardest jet. In figure 16, we show the performance of the generic tagger on this signal (solid line), as well as the performance of a dedicated tagger trained to discriminate T -jets from QCD-jets. We find

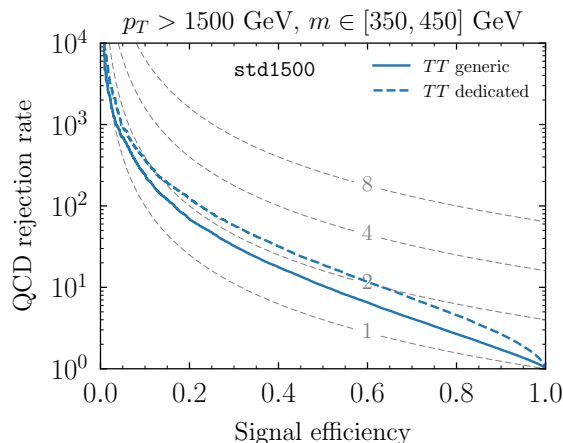


Figure 16. Signal efficiency versus QCD rejection for jets resulting from decays of 400 GeV vector-like quarks T , decaying as $T \rightarrow c\bar{c}$. Solid: generic `std1500` tagger; dashed: dedicated tagger.

that the generic tagger has a moderate performance for this signal, with approximately 10% QCD efficiency at 50% signal efficiency. One might wonder if the sensitivity to such signals is lost by training the generic tagger only on colour-singlet jets. However, the dedicated tagger has only marginally better performance, which suggests that the reason for the moderate performance of the generic tagger is that this type of signal is intrinsically hard to distinguish from QCD jets.

An outline of a possible search strategy for this signature could be as follows. Select back-to-back dijet events passing some high p_T threshold. Require the jet mass of the two jets to be close to each other, and after applying a tagger at some threshold, look for a bump in the average jet mass distribution of the two jets. This could either be a cut-and-count analysis in relatively wide jet mass bins, or as a bump hunting shape analysis on a smooth background distribution.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] 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** (2008) 242001 [[arXiv:0802.2470](https://arxiv.org/abs/0802.2470)] [[INSPIRE](https://inspirehep.net/literature/76005)].
- [2] J. Thaler and L.-T. Wang, *Strategies to Identify Boosted Tops*, *JHEP* **07** (2008) 092 [[arXiv:0806.0023](https://arxiv.org/abs/0806.0023)] [[INSPIRE](https://inspirehep.net/literature/77321)].
- [3] D.E. Kaplan, K. Rehermann, M.D. Schwartz and B. Tweedie, *Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks*, *Phys. Rev. Lett.* **101** (2008) 142001 [[arXiv:0806.0848](https://arxiv.org/abs/0806.0848)] [[INSPIRE](https://inspirehep.net/literature/77321)].
- [4] L.G. Almeida, S.J. Lee, G. Perez, G.F. Sterman, I. Sung and J. Virzi, *Substructure of high- p_T Jets at the LHC*, *Phys. Rev. D* **79** (2009) 074017 [[arXiv:0807.0234](https://arxiv.org/abs/0807.0234)] [[INSPIRE](https://inspirehep.net/literature/78001)].

- [5] T. Plehn, G.P. Salam and M. Spannowsky, *Fat Jets for a Light Higgs*, *Phys. Rev. Lett.* **104** (2010) 111801 [[arXiv:0910.5472](#)] [[INSPIRE](#)].
- [6] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, *Stop Reconstruction with Tagged Tops*, *JHEP* **10** (2010) 078 [[arXiv:1006.2833](#)] [[INSPIRE](#)].
- [7] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N -subjettiness*, *JHEP* **03** (2011) 015 [[arXiv:1011.2268](#)] [[INSPIRE](#)].
- [8] A. Hook, M. Jankowiak and J.G. Wacker, *Jet Dipolarity: Top Tagging with Color Flow*, *JHEP* **04** (2012) 007 [[arXiv:1102.1012](#)] [[INSPIRE](#)].
- [9] M. Jankowiak and A.J. Larkoski, *Jet Substructure Without Trees*, *JHEP* **06** (2011) 057 [[arXiv:1104.1646](#)] [[INSPIRE](#)].
- [10] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N -subjettiness*, *JHEP* **02** (2012) 093 [[arXiv:1108.2701](#)] [[INSPIRE](#)].
- [11] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [[arXiv:1305.0007](#)] [[INSPIRE](#)].
- [12] I. Moult, L. Necib and J. Thaler, *New Angles on Energy Correlation Functions*, *JHEP* **12** (2016) 153 [[arXiv:1609.07483](#)] [[INSPIRE](#)].
- [13] K. Datta and A. Larkoski, *How Much Information is in a Jet?*, *JHEP* **06** (2017) 073 [[arXiv:1704.08249](#)] [[INSPIRE](#)].
- [14] ATLAS collaboration, *Search for $W' \rightarrow tb \rightarrow qqbb$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **75** (2015) 165 [[arXiv:1408.0886](#)] [[INSPIRE](#)].
- [15] CMS collaboration, *Searches for W' bosons decaying to a top quark and a bottom quark in proton-proton collisions at 13 TeV*, *JHEP* **08** (2017) 029 [[arXiv:1706.04260](#)] [[INSPIRE](#)].
- [16] CMS collaboration, *Search for $t\bar{t}$ resonances in highly boosted lepton+jets and fully hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **07** (2017) 001 [[arXiv:1704.03366](#)] [[INSPIRE](#)].
- [17] CMS collaboration, *Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks*, *Eur. Phys. J. C* **76** (2016) 371 [[arXiv:1602.08762](#)] [[INSPIRE](#)].
- [18] ATLAS collaboration, *Searches for heavy diboson resonances in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **09** (2016) 173 [[arXiv:1606.04833](#)] [[INSPIRE](#)].
- [19] CMS collaboration, *Search for massive resonances decaying into WW , WZ or ZZ bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **03** (2017) 162 [[arXiv:1612.09159](#)] [[INSPIRE](#)].
- [20] CMS collaboration, *Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **77** (2017) 636 [[arXiv:1707.01303](#)] [[INSPIRE](#)].
- [21] CMS collaboration, *Search for pair-produced vectorlike B quarks in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *Phys. Rev. D* **93** (2016) 112009 [[arXiv:1507.07129](#)] [[INSPIRE](#)].
- [22] CMS collaboration, *Search for single production of a heavy vector-like T quark decaying to a Higgs boson and a top quark with a lepton and jets in the final state*, *Phys. Lett. B* **771** (2017) 80 [[arXiv:1612.00999](#)] [[INSPIRE](#)].

- [23] CMS collaboration, *Search for dark matter in proton-proton collisions at 8 TeV with missing transverse momentum and vector boson tagged jets*, *JHEP* **12** (2016) 083 [Erratum *ibid.* **08** (2017) 035] [[arXiv:1607.05764](#)] [[INSPIRE](#)].
- [24] ATLAS collaboration, *Search for dark matter produced in association with a hadronically decaying vector boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **763** (2016) 251 [[arXiv:1608.02372](#)] [[INSPIRE](#)].
- [25] CMS collaboration, *Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV*, [arXiv:1710.00159](#) [[INSPIRE](#)].
- [26] J.A. Aguilar-Saavedra and F.R. Joaquim, *Multiboson production in W' decays*, *JHEP* **01** (2016) 183 [[arXiv:1512.00396](#)] [[INSPIRE](#)].
- [27] K. Agashe, P. Du, S. Hong and R. Sundrum, *Flavor Universal Resonances and Warped Gravity*, *JHEP* **01** (2017) 016 [[arXiv:1608.00526](#)] [[INSPIRE](#)].
- [28] K.S. Agashe, J. Collins, P. Du, S. Hong, D. Kim and R.K. Mishra, *LHC Signals from Cascade Decays of Warped Vector Resonances*, *JHEP* **05** (2017) 078 [[arXiv:1612.00047](#)] [[INSPIRE](#)].
- [29] R.N. Mohapatra and Y. Zhang, *LHC accessible second Higgs boson in the left-right model*, *Phys. Rev. D* **89** (2014) 055001 [[arXiv:1401.0018](#)] [[INSPIRE](#)].
- [30] U. Ellwanger and M. Rodriguez-Vazquez, *Simultaneous Search for Extra Light and Heavy Higgs Bosons via Cascade Decays*, [arXiv:1707.08522](#) [[INSPIRE](#)].
- [31] J.A. Aguilar-Saavedra, *Stealth multiboson signals*, *Eur. Phys. J. C* **77** (2017) 703 [[arXiv:1705.07885](#)] [[INSPIRE](#)].
- [32] A. Chakraborty, A.M. Iyer and T.S. Roy, *A Universal Framework for Finding Anomalous Objects at the LHC*, [arXiv:1707.07084](#) [[INSPIRE](#)].
- [33] S. Catani, Y.L. Dokshitzer, M.H. Seymour and B.R. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, *Nucl. Phys. B* **406** (1993) 187 [[INSPIRE](#)].
- [34] S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* **48** (1993) 3160 [[hep-ph/9305266](#)] [[INSPIRE](#)].
- [35] G.C. Blazey et al., *Run II jet physics*, [hep-ex/0005012](#) [[INSPIRE](#)].
- [36] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [[arXiv:1405.0301](#)] [[INSPIRE](#)].
- [37] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [[arXiv:0710.3820](#)] [[INSPIRE](#)].
- [38] DELPHES 3 collaboration, J. de Favereau et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [[arXiv:1307.6346](#)] [[INSPIRE](#)].
- [39] M. Cacciari, G.P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](#)] [[INSPIRE](#)].
- [40] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [[arXiv:1111.6097](#)] [[INSPIRE](#)].
- [41] J.A. Aguilar-Saavedra, *PROTOS, a PROgram for TOp Simulations*, <http://jaguilar.web.cern.ch/jaguilar/protos/>.

- [42] C. Brust, P. Maksimovic, A. Sady, P. Saraswat, M.T. Walters and Y. Xin, *Identifying boosted new physics with non-isolated leptons*, *JHEP* **04** (2015) 079 [[arXiv:1410.0362](#)] [[INSPIRE](#)].
- [43] A. Alloul, N.D. Christensen, C. Degrande, C. Duhr and B. Fuks, *FeynRules 2.0 — A complete toolbox for tree-level phenomenology*, *Comput. Phys. Commun.* **185** (2014) 2250 [[arXiv:1310.1921](#)] [[INSPIRE](#)].
- [44] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer and T. Reiter, *UFO — The Universal FeynRules Output*, *Comput. Phys. Commun.* **183** (2012) 1201 [[arXiv:1108.2040](#)] [[INSPIRE](#)].
- [45] F. Chollet, *Keras: Deep Learning for Python (2015)*, <https://github.com/fchollet/keras>.
- [46] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems (2015)*, <http://tensorflow.org/>.
- [47] S.D. Ellis, C.K. Vermilion and J.R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, *Phys. Rev. D* **80** (2009) 051501 [[arXiv:0903.5081](#)] [[INSPIRE](#)].
- [48] 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** (2010) 094023 [[arXiv:0912.0033](#)] [[INSPIRE](#)].
- [49] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084 [[arXiv:0912.1342](#)] [[INSPIRE](#)].
- [50] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146 [[arXiv:1402.2657](#)] [[INSPIRE](#)].
- [51] C. Shimmin et al., *Decorrelated Jet Substructure Tagging using Adversarial Neural Networks*, *Phys. Rev. D* **96** (2017) 074034 [[arXiv:1703.03507](#)] [[INSPIRE](#)].
- [52] J. Dolen, P. Harris, S. Marzani, S. Rappoccio and N. Tran, *Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure*, *JHEP* **05** (2016) 156 [[arXiv:1603.00027](#)] [[INSPIRE](#)].
- [53] CMS collaboration, *Search for a massive resonance decaying into a Higgs boson and a W or Z boson in hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **02** (2016) 145 [[arXiv:1506.01443](#)] [[INSPIRE](#)].
- [54] ATLAS collaboration, *A search for resonances decaying into a Higgs boson and a new particle X in the $XH \rightarrow q\bar{q}b\bar{b}$ final state with the ATLAS detector*, [arXiv:1709.06783](#) [[INSPIRE](#)].
- [55] P. Gras et al., *Systematics of quark/gluon tagging*, *JHEP* **07** (2017) 091 [[arXiv:1704.03878](#)] [[INSPIRE](#)].
- [56] L.M. Dery, B. Nachman, F. Rubbo and A. Schwartzman, *Weakly Supervised Classification in High Energy Physics*, *JHEP* **05** (2017) 145 [[arXiv:1702.00414](#)] [[INSPIRE](#)].
- [57] T. Cohen, M. Freytsis and B. Ostdiek, *(Machine) Learning to Do More with Less*, [arXiv:1706.09451](#) [[INSPIRE](#)].
- [58] E.M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: Learning from mixed samples in high energy physics*, [arXiv:1708.02949](#) [[INSPIRE](#)].
- [59] B.A. Dobrescu and F. Yu, *Exotic Signals of Vectorlike Quarks*, [arXiv:1612.01909](#) [[INSPIRE](#)].