

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Anomaly Detection for Resonant New Physics with Machine Learning

### Permalink

<https://escholarship.org/uc/item/6k64m399>

### Journal

Physical Review Letters, 121(24)

### ISSN

0031-9007

### Authors

Collins, Jack  
Howe, Kiel  
Nachman, Benjamin

### Publication Date

2018-12-14

### DOI

10.1103/physrevlett.121.241803

Peer reviewed

**Anomaly Detection for Resonant New Physics with Machine Learning**Jack Collins,<sup>1,2,\*</sup> Kiel Howe,<sup>3,†</sup> and Benjamin Nachman<sup>4,5,‡</sup><sup>1</sup>*Maryland Center for Fundamental Physics, Department of Physics, University of Maryland, College Park, Maryland 20742, USA*<sup>2</sup>*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA*<sup>3</sup>*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*<sup>4</sup>*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*<sup>5</sup>*Simons Institute for the Theory of Computing, University of California, Berkeley, Berkeley, California 94720, USA*

(Received 30 May 2018; published 12 December 2018)

Despite extensive theoretical motivation for physics beyond the standard model (BSM) of particle physics, searches at the Large Hadron Collider have found no significant evidence for BSM physics. Therefore, it is essential to broaden the sensitivity of the search program to include unexpected scenarios. We present a new model-agnostic anomaly detection technique that naturally benefits from modern machine learning algorithms. The only requirement on the signal for this new procedure is that it is localized in at least one known direction in phase space. Any other directions of phase space that are uncorrelated with the localized one can be used to search for unexpected features. This new method is applied to the dijet resonance search to show that it can turn a modest  $2\sigma$  excess into a  $7\sigma$  excess for a model with an intermediate BSM particle that is not currently targeted by a dedicated search.

DOI: [10.1103/PhysRevLett.121.241803](https://doi.org/10.1103/PhysRevLett.121.241803)

The main goal of high energy physics is to identify the elementary building blocks of matter and to characterize the laws governing their motion. In order to achieve this goal, experiments at the energy frontier collide particles with extremely high momenta in a quest to directly produce the elementary particles and study their interactions. The collider currently able to directly probe the smallest distance scales is the Large Hadron Collider (LHC). Building on decades of effort at previous experiments, the ATLAS and CMS Collaborations at the LHC discovered the Higgs boson in 2012 [1,2], completing the standard model (SM) of particle physics. While the SM has been enormously successful, it is not a complete theory of nature as it lacks a description of dark matter and gravity, in addition to various technical or aesthetic problems. Despite an intensive and impressive program to search directly for physics beyond the SM at the LHC [3–7], there is still no direct evidence for any new structures in nature. However, there are numerous compelling theoretical motivations for physics beyond the SM (BSM) at the energies scales accessible by the LHC [8]. While it could be that the BSM particles are too massive or produced with too low a cross section to be discovered yet, it is also possible that the current search program is

simply not sensitive to the regions of phase space populated by BSM physics.

In order to mitigate the possibility of uncovered regions of phase space, collider experiments have implemented model-independent anomaly detection techniques. Traditionally, there are two such approaches: general searches and bump hunts. The idea of general searches is to compare data and simulations in a large number of event topologies, characterized by the number and type of various physics objects, such as leptons or hadronic jets, resulting from high energy quark and gluon production [9–20]. While this approach has a broad coverage, it is restricted to simple observables because it relies heavily on simulations for background estimation. In contrast, bump hunts [21] often do not use any simulation for background estimation, other than to motivate and validate the background fit procedure: after identifying a region of phase space where a signal is expected to be localized, the background is fit with a smooth function and interpolated to the signal-sensitive region. Excesses over this background prediction would be an indication of BSM physics. To enhance the resonance structure from a diobject invariant mass, modern classification tools [22] can be used select the target objects like the  $b$  quark [28,29], top quark [30,31],  $W/Z$  [30,31], or Higgs boson [32,33] jets from generic quark or gluon jets. However, these classifiers are trained in simulation and calibrated in data, which may lead to suboptimal classifiers. Furthermore, it is not possible in this paradigm to develop classifiers for BSM objects, since no calibration sample exists.

This Letter presents a new technique to search for BSM physics that significantly extends the bump hunt approach

---

*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>.*

that uses classifiers trained directly on data. Consider a signal that is localized in one kinematic variable (the resonant variable,  $m_{\text{res}}$ ) on top of a smoothly varying background, for example, a dijet resonance that can be reconstructed from the invariant mass of two jets. Suppose that each event has additional auxiliary information (such as the substructure in the two jets) that may provide additional discriminating power between signal and background, but the detailed signal characteristics in these auxiliary variables are unknown *a priori*. Our proposal is that a classifier can be trained to discern the auxiliary characteristics of the signal (if present) directly from data, without reference to any specific signal model hypothesis. The output of this classifier can then be used to select signal-like events and reject background events, producing a new distribution in the resonant variable that remains smooth in the case that no signal was present, but that may enhance the significance of the bump if a real signal is present. In the event that a signal is discovered, the output of the classifier can then be studied to infer the signal characteristics.

The key feature of resonant signals that are utilized in our approach is that their localization in one kinematic variable on top of a smoothly varying background allows the identification of potential signal-enhanced and signal-depleted signal and sideband regions, respectively, with almost identical background characteristics. A classifier trained to distinguish the auxiliary characteristics of the signal region events from those of the sideband may in principle be as powerful as a classifier trained to distinguish pure samples of signal and background events—this is a specific application of classification without labels (CWOLA) [34]. To see why this is the case, suppose that it is possible to define an ideal sideband selection that contains only a background and no signal, and an ideal signal region that contains a background identical to that in the sideband but also a small signal that is distinct from the background. By the Neyman–Pearson lemma [35], the most powerful test statistic for discriminating signal (sig) events from background (bg) events using some observables  $Y$  is the likelihood ratio

$$L(Y) = \frac{p(Y|\text{sig})}{p(Y|\text{bg})}, \quad (1)$$

and a fully supervised classifier is trained to approximate any monotonic rescaling of this function. A classifier that is trained to discriminate signal region events (sig + bg) from sideband region events (bg) will instead ideally learn to approximate a monotonic rescaling of the function

$$\hat{L}(Y) = \frac{p(Y|\text{sig} + \text{bg})}{p(Y|\text{bg})} = f_{\text{sig}} \frac{p(Y|\text{sig})}{p(Y|\text{bg})} + f_{\text{bg}}, \quad (2)$$

where  $f_{\text{sig}}$  and  $f_{\text{bg}}$  are the proportions of signal and background events in the signal region. The fact that

Eq. (2) is itself a monotonic rescaling of  $L(Y)$  from Eq. (1) shows that there is no fundamental obstruction for the CWOLA-based classifier to identify the ideal decision boundaries for signal selection. The above argument also holds if the sideband region has a small amount of signal, as long as the signal proportion is less than the signal region [34]. Practical limits will arise from limited statistics (particularly for the signal) and other technical difficulties that may obstruct a trainable classifier from reaching the performance achievable with labeled simulations, and also from the small differences in the background characteristics between signal and sideband regions.

Our extended bump hunt procedure also has some features in common with the sPlot [36] technique. In particular, sPlot provides a method for determining the distribution of multiple event classes for a resonant feature (“control variable” in the language of Ref. [36]) using a set of uncorrelated auxiliary features (“discriminating variables” in Ref. [36]). The key differences between sPlot and the extended bump hunt are (1) we are interested in using machine learning to isolate a signal-rich region of phase space, and (2) we do not take the probability distribution for the auxiliary features as input—the classification procedure learns useful information directly from the data.

A danger that is present when training and testing a classifier on the same data set is that it may overfit the training data and learn the specific statistical fluctuations in that data set rather than the true underlying distribution. Classifiers used in this way will preferentially select signal-region events based on their statistical fluctuations, and they will create a fake bump in the resonance-variable distribution even when no real signal is present. A simple way to mitigate the background sculpting is to split the underlying data set randomly into a training set and a test set that will have uncorrelated statistical fluctuations. This would, however, result in an effective loss of luminosity available both for training and for testing. Instead we advocate for an  $n$ -fold cross-validation procedure, in which the data are randomly partitioned into  $n$  sets of equal size (stratified by  $m_{\text{res}}$  bin). The selection on each of the  $n$  partitions is performed using the output of a classifier trained and validated on the remaining  $n - 1$  partitions, resulting in a total of  $n$  classifiers. Any statistical fluctuations learned by a classifier from its training data will be uncorrelated with those in the data on which it is used for event selection. The effects of overtraining on the performance of the classifiers can be mitigated by a nested cross-validation procedure, as is described in detail in Ref. [37].

In using the cross-validation procedure, there is a danger that the bin counts become non-Poissonian due to correlations between the selections, which would need to be accounted for with computationally expensive test statistic calibration based on a large number of simulated toys. If this were found to be prohibitive in a specific application, a simple test-train split remains a possibility to avoid this

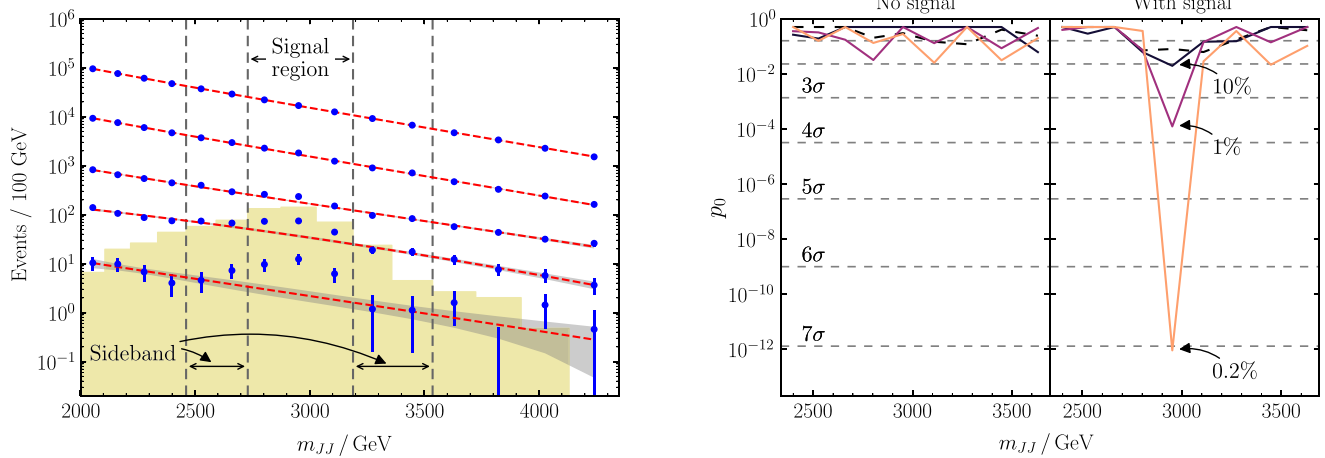


FIG. 1. Left:  $m_{JJ}$  distribution of dijet events (including injected signal, indicated by the filled histogram) before and after applying jet substructure cuts using the NN classifier output for the  $m_{JJ} \simeq 3$  TeV mass hypothesis. The dashed red lines indicate the fit to the data points outside of the signal region, with the gray bands representing the fit uncertainties. The top set of markers represent the raw dijet distribution with no cut applied, while the subsequent sets of markers have cuts applied at thresholds with efficiency of  $10^{-1}$ ,  $10^{-2}$ ,  $2 \times 10^{-3}$ , and  $2 \times 10^{-4}$ . Right: Local  $p_0$ -values for a range of signal mass hypotheses in the case that no signal has been injected (left), and in the case that a 3 TeV resonance signal has been injected (right). The dashed lines correspond to the case where no substructure cut is applied, and the various solid lines correspond to cuts on the classifier output with efficiencies of  $10^{-1}$ ,  $10^{-2}$ , and  $2 \times 10^{-3}$ .

difficulty. However, we find in our tests that this does not distort the test statistic distributions in our examples in Ref. [37], and we find that asymptotic formulas [38] or throwing toys, with counts based on the merged selected events, provide accurate  $p$  values.

To summarize, the extended bump hunt algorithm proceeds as follows, for a single resonance mass hypothesis  $\hat{m}_{\text{res}}$ : (1) Identify an observable  $m_{\text{res}}$ , in which a signal is expected to be resonant, and a set of auxiliary variables  $Y$  that are to be used for signal selection. The variables  $Y$  must be independent of  $m_{\text{res}}$ . There are a number of methods for correcting this if not inherently true [39–46]. A background model  $f(m_{\text{res}})$  is needed for  $m_{\text{res}}$ . Typically (and in the example below) this is done with a parametric fit, though nonparametric methods are also possible [47]. (2) Define a signal region in a window around  $\hat{m}_{\text{res}}$ . (3) Define sideband regions that are disjoint from the signal region but still sufficiently close that the background distribution in  $Y$  is expected to be nearly identical. (4) Use a cross-validation procedure to separate training samples from test samples. For each test subsample: (a) Train a classifier to discriminate training events drawn from the sideband regions from those drawn from the signal region, using variables  $Y$ . (b) Select a fraction  $\epsilon$  of the most signal-like test events as determined by the classifiers. (5) Merge selected event samples. (6) Perform a statistical test for the presence of an excess in the signal region of the  $m_{\text{res}}$  distribution after the cut has been applied, using the data outside of the signal region for a background determination using the background model  $f(m_{\text{res}})$ . The statistical analysis can be performed using pseudoexperiments generated by subsampling from the data itself. This procedure is repeated

starting from step (2) for a series of resonance mass hypotheses, as in a usual bump hunt. This entails the usual trials factor associated with the scan over the resonance variable, but it does not invoke any additional trials factor associated with the space of auxiliary variables. Using asymptotic formulas [38] or throwing toys, with counts based on the merged selected events, provides accurate  $p$  values [37].

As a concrete example of the new bump hunting strategy, suppose there is a new resonance that decays into unusual jets. We do not know *a priori* how to look for the new resonance, but we can consider the substructure of each jet to look for an anomalous radiation pattern. The left plot of Fig. 1 shows the invariant mass of two jet four-vectors in simulated QCD dijet events [48]. To illustrate the power of the technique, we have also injected events from the decay of a  $W'$  particle with a mass of 3 TeV. This  $W'$  is constructed to decay to a  $W$  boson ( $m_W \approx 80$  GeV) and a new  $X$  particle ( $m_X \approx 400$  GeV), which itself decays into two  $W$  bosons, as described in Ref. [52–54]. We consider the all-hadronic channel in which each  $W$  boson decays into quark pairs. The signal is thus characterized as having two large jets, one with a two-prong substructure and one with a four-prong substructure. The shaded histogram in the left plot of Fig. 1 peaks at the resonance mass of 3 TeV with a broad width due to jet fragmentation and clustering effects. Without any selection on the jets' substructure, there is no significant indication of the signal hiding under the smooth background from generic quarks and gluons.

To enhance the sensitivity of this search using the extended bump hunt method described above, a suite of classifiers are trained to distinguish a sliding signal region

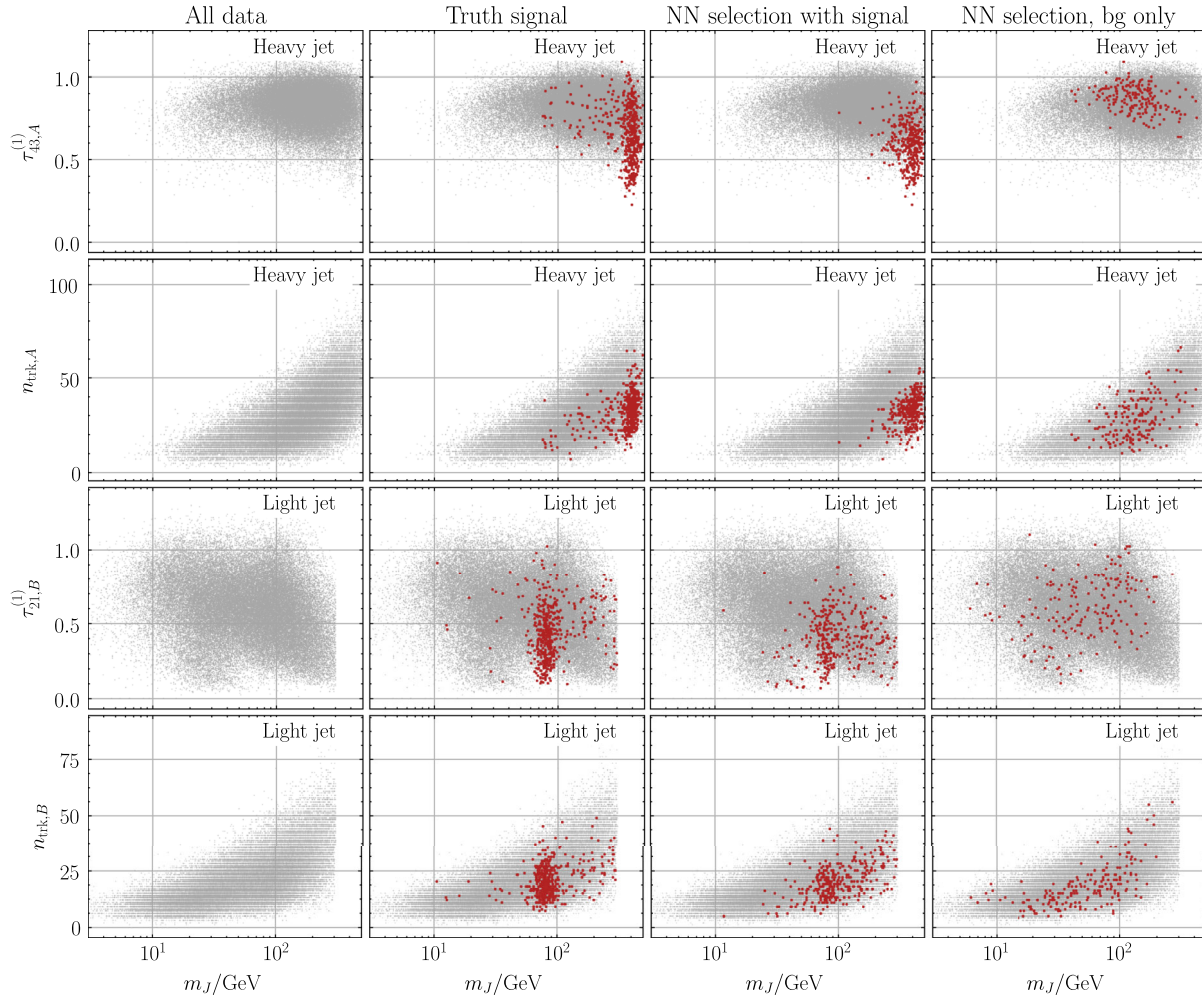


FIG. 2. Two-dimensional projections of the twelve-dimensional feature space of the signal region data set. First column: all signal region events. This is repeated in the other columns to aid comparisons. Second column: truth-level simulated signal events highlighted in red. Third column: The red dots are the 0.2% most signal-like events selected by the classifier described in the text. Fourth column: The red dots are the 0.2% most signal-like events selected by a classifier trained on the same sample but with true-signal events removed.

from sideband regions. For each jet, the following substructure information ( $Y$ ) is used:

$$m_J, \sqrt{\tau_1^{(2)}/\tau_1^{(1)}}, \tau_{21}, \tau_{32}, \tau_{43}, n_{\text{trk}}, \quad (3)$$

where  $m_J$  is the jet mass,  $n_{\text{trk}}$  is the number of charged particles (tracks) in the ungroomed jet, the  $N$ -subjettiness ratios are defined by  $\tau_{MN} = \tau_M^{(1)}/\tau_N^{(1)}$ , and the observables  $\tau_N^{(\beta)}$  are defined in Ref. [55].

The output of the classifiers are then used to select signal-like events over the full range of the  $m_{JJ}$  distribution. The resulting distributions are shown in Fig. 1 (left) after applying thresholds on the NN output with overall efficiencies 10%, 1%, 0.2%, and 0.02%, respectively, in descending order. Prior to applying any threshold, the resonant signal has  $S/B = 6.4 \times 10^{-3}$  and a significance  $S/\sqrt{B} = 1.8$  in the signal region, and the  $m_{JJ}$  distribution has no discernible resonant feature.

However, after applying the threshold determined by the classifier, a clear bump develops in the signal region with local significance of  $7\sigma$  at the 0.2% threshold. Of course, in the event that the resonance mass is not known in advance, then a scan must be performed over possible resonance masses. It is important that the procedure does not create fake bumps in the background when no signal is present. We show in Fig. 1 (right) the  $p$  values obtained in the mass scan over this distribution in the case that (a) no signal is present, and (b) the case that the signal has been injected. We find that no significant bumps are created in the signal-free test. Furthermore, we find that traditional searches aimed at finding diboson resonances using jet substructure-based supervised learning algorithms (but for SM bosons) are not able to enhance the significance of this signal for a wide range of  $S/B$  and classifier working points [37].

In order to characterize the signal that the classifier has found, we can study the distribution of selected signal-like events, as illustrated in Fig. 2. We see that the classifier

trained in the presence of a true signal has identified a population of events with a heavier jet with mass  $m_{JA} \simeq 400$  GeV, a small number of tracks, and a small  $\tau_{43}$ , as well as a lighter jet with mass  $m_{JB}$ , a small number of tracks, and a small  $\tau_{21}$ .

In conclusion, we have presented a new technique to search for physics beyond the SM that requires very little prior knowledge of the signal. The method was demonstrated in simulation on an all-hadronic resonance search at the LHC, where an uninteresting excess was enhanced to a level of discovery. There are many other possibilities for applying this technique directly to data, in any case where the signal is expected to be localized in one dimension. By naturally exploiting the power of modern machine learning, we hope that this extended bump hunt will help to expose new distance scales in nature on the quest for BSM physics at the LHC and beyond.

The data sets and code used for the case study can be found at Refs. [56,57].

We appreciate helpful discussions with and useful feedback on the manuscript from Timothy Cohen, Aviv Cukierman, Patrick Fox, Jack Kearney, Zhen Liu, Eric Metodiev, Brian Nord, Bryan Ostdiek, Francesco Rubbo, and Jesse Thaler. We would also like to thank Peizhi Du for providing the UFO file for the benchmark signal model. The work of J. H. C. is supported by NSF under Grant No. PHY-1620074 and by the Maryland Center for Fundamental Physics (MCFP). The work of B. N. is supported by the DOE under Contract No. DE-AC02-05CH11231. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

\*jhc296@umd.edu

†khowe@fnal.gov

‡bpnachman@lbl.gov

- [1] G. Aad *et al.* (ATLAS Collaboration), Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC, *Phys. Lett. B* **716**, 1 (2012).
- [2] S. Chatrchyan *et al.* (CMS Collaboration), Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, *Phys. Lett. B* **716**, 30 (2012).
- [3] ATLAS Collaboration, Supersymmetry searches (2018), <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults>.
- [4] ATLAS Collaboration, Exotic physics searches (2018), <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ExoticsPublicResults>.
- [5] CMS Collaboration, CMS exotica public physics results (2018), <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsEXO>.
- [6] CMS Collaboration, CMS supersymmetry physics results (2018), <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>.
- [7] CMS Collaboration, CMS beyond-two-generations (b2g) public physics results (2018), <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsB2G>.
- [8] See e.g., J.R. Ellis, Beyond the standard model with the LHC, *Nature (London)* **448**, 297 (2007).
- [9] ATLAS Collaboration, A model independent general search for new phenomena with the ATLAS detector at  $\sqrt{s} = 13$  TeV, Technical Report No. ATLAS-CONF-2017-001 (CERN, Geneva, 2017).
- [10] CMS Collaboration, Model unspecific search for new physics in pp collisions at  $\sqrt{s} = 7$  TeV, Technical Report No. CMS-PAS-EXO-10-021 (CERN, Geneva, 2011).
- [11] ATLAS Collaboration, A general search for new phenomena with the ATLAS detector in pp collisions at  $\sqrt{s} = 7$  TeV, Technical Report No. ATLAS-CONF-2012-107 (CERN, Geneva, 2012).
- [12] ATLAS Collaboration, A general search for new phenomena with the ATLAS detector in pp collisions at  $\sqrt{s} = 8$  TeV, Technical Report No. ATLAS-CONF-2014-006 (CERN, Geneva, 2014).
- [13] A. Aktas *et al.* (H1 Collaboration), A general search for new phenomena in ep scattering at HERA, *Phys. Lett. B* **602**, 14 (2004).
- [14] F. D. Aaron *et al.* (H1 Collaboration), A general search for new phenomena at HERA, *Phys. Lett. B* **674**, 257 (2009).
- [15] B. Abbott *et al.* (D0 Collaboration), Search for new physics in  $e\mu X$  data at DØ using Sherlock: A quasi model independent search strategy for new physics, *Phys. Rev. D* **62**, 092004 (2000).
- [16] V. M. Abazov *et al.* (D0 Collaboration), A quasi model independent search for new physics at large transverse momentum, *Phys. Rev. D* **64**, 012004 (2001).
- [17] T. Aaltonen *et al.* (CDF Collaboration), Model-independent and quasi-model-independent search for new physics at CDF, *Phys. Rev. D* **78**, 012002 (2008).
- [18] T. Aaltonen *et al.* (CDF Collaboration), Global search for new physics with  $2.0 \text{ fb}^{-1}$  at CDF, *Phys. Rev. D* **79**, 011101 (2009).
- [19] B. Knuteson, Ph.D. thesis, University of California, Berkeley, 2000.
- [20] B. Knuteson, Systematic analysis of high-energy collider data, *Nucl. Instrum. Methods A* **534**, 7 (2004).
- [21] See e.g., analyses that cite Choudalakis, Georgios, On hypothesis testing, trials factor, hypertests and the Bump-Hunter, in Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland, 2011 (2011).
- [22] The first use of jet substructure for new particle searches were proposed in Refs. [23–25]; for recent theoretical and experimental reviews, see Refs. [26,27], respectively.

- [23] M. H. Seymour, Searches for new particles using cone and cluster jet algorithms: A comparative study, *Z. Phys. C* **62**, 127 (1994).
- [24] J. M. Butterworth, B. E. Cox, and J. R. Forshaw, *WW* scattering at the CERN LHC, *Phys. Rev. D* **65**, 096014 (2002).
- [25] 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).
- [26] 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, [arXiv:1709.04464](https://arxiv.org/abs/1709.04464).
- [27] L. Asquith *et al.*, Jet substructure at the large hadron collider: Experimental review, [arXiv:1803.06991](https://arxiv.org/abs/1803.06991).
- [28] ATLAS Collaboration, Optimisation and performance studies of the ATLAS *b*-tagging algorithms for the 2017-18 LHC run, Technical Report No. ATL-PHYS-PUB-2017-013 (CERN, Geneva, 2017).
- [29] CMS Collaboration, Identification of *b* quark jets at the CMS Experiment in the LHC Run 2, Technical Report CMS-PAS-BTV-15-001 (CERN, Geneva, 2016).
- [30] ATLAS Collaboration, Identification of hadronically-decaying W Bosons and top quarks using high-level features as input to boosted decision trees and deep neural networks in ATLAS at  $\sqrt{s} = 13$  TeV, Technical Report No. ATL-PHYS-PUB-2017-004 (CERN, Geneva, 2017).
- [31] CMS Collaboration, W and top tagging scale factors, CERN, Report No. CMS-DP-2017-026, 2017.
- [32] ATLAS Collaboration, Boosted Higgs ( $\rightarrow b\bar{b}$ ) Boson identification with the ATLAS detector at  $\sqrt{s} = 13$  TeV, Technical Report No. ATLAS-CONF-2016-039 (CERN, Geneva, 2016).
- [33] CMS Collaboration, Identification of double-*b* quark jets in boosted event topologies, Technical Report No. CMS-PAS-BTV-15-002 (CERN, Geneva, 2016).
- [34] E. M. Metodiev, B. Nachman, and J. Thaler, Classification without labels: Learning from mixed samples in high energy physics, *J. High Energy Phys.* **10** (2017) 174.
- [35] J. Neyman, IX. On the problem of the most efficient tests of statistical hypotheses, *Phil. Trans. R. Soc. A* **231**, 289 (1933).
- [36] M. Pivk and F. R. Le Diberder, SPlot: A Statistical tool to unfold data distributions, *Nucl. Instrum. Methods A* **555**, 356 (2005).
- [37] B. Nachman, J. Collins, and K. Howe, CWoLa hunting: Extending the bump hunt with machine learning, in preparation.
- [38] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, *Eur. Phys. J. C* **71**, 1554 (2011); Erratum, *Eur. Phys. J. C* **73**, 2501 (2013).
- [39] G. Louppe, M. Kagan, and K. Cranmer, Learning to pivot with adversarial networks, [arXiv:1611.01046](https://arxiv.org/abs/1611.01046).
- [40] I. Moulton, B. Nachman, and D. Neill, Convolved substructure: Analytically decorrelating jet substructure observables, *J. High Energy Phys.* **05** (2018) 002.
- [41] J. Dolen, P. Harris, S. Marzani, S. Rappoccio, and N. Tran, Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure, *J. High Energy Phys.* **05** (2016) 156.
- [42] C. Shimmin, P. Sadowski, P. Baldi, E. Weik, D. Whiteson, E. Goul, and A. Sogaard, Decorrelated jet substructure tagging using adversarial neural networks, *Phys. Rev. D* **96**, 074034 (2017).
- [43] J. A. Aguilar-Saavedra, J. H. Collins, and R. K. Mishra, A generic anti-QCD jet tagger, *J. High Energy Phys.* **11** (2017) 163.
- [44] ATLAS Collaboration, Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS, Technical Report No. ATL-PHYS-PUB-2018-014 (CERN, Geneva, 2018).
- [45] J. Stevens and M. Williams, uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers, *J. Instrum.* **8**, P12013 (2013).
- [46] A. M. Sirunyan *et al.* (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).
- [47] M. Frate, K. Cranmer, S. Kalia, A. Vandenberg-Rodes, and D. Whiteson, Modeling smooth backgrounds and generic localized signals with gaussian processes, [arXiv:1709.05681](https://arxiv.org/abs/1709.05681).
- [48] Events are generated with Madgraph5\_aMC@NLO v2.5.5 [49] + Pythia 8.226 [50] + Delphes 3.4.1 [51]. For further details, see Ref. [37].
- [49] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *J. High Energy Phys.* **07** (2014) 079.
- [50] T. Sjostrand, S. Mrenna, and P. Z. Skands, A brief introduction to PYTHIA 8.1, *Comput. Phys. Commun.* **178**, 852 (2008).
- [51] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast simulation of a generic collider experiment, *J. High Energy Phys.* **02** (2014) 057.
- [52] K. Agashe, P. Du, S. Hong, and R. Sundrum, Flavor universal resonances and warped gravity, *J. High Energy Phys.* **01** (2017) 016.
- [53] K. Agashe, J. H. Collins, P. Du, S. Hong, D. Kim, and R. K. Mishra, Dedicated strategies for triboson signals from cascade decays of vector resonances, [arXiv:1711.09920](https://arxiv.org/abs/1711.09920).
- [54] K. Agashe, J. H. Collins, P. Du, S. Hong, D. Kim, and R. K. Mishra, Detecting a boosted diboson resonance, *J. High Energy Phys.* **11** (2018) 027..
- [55] J. Thaler and K. Van Tilburg, Identifying boosted objects with N-subjettiness, *J. High Energy Phys.* **03** (2011) 015.
- [56] B. Nachman, J. Collins, and K. Howe, CWoLa hunting: Data sample, Mendeley data, DOI: [10.17632/57dh4b7f9m.1](https://doi.org/10.17632/57dh4b7f9m.1) (2018).
- [57] B. Nachman, J. Collins, and K. Howe, CWoLa Hunting: Code (2018), <https://github.com/Jackadsa/CWoLa-Hunting>.