Dijet resonance search with weak supervision using $\sqrt{s} = 13$ TeV pp collisions in the ATLAS detector

ATLAS Collaboration

DOI
10.1103/PhysRevLett.125.131801

Publication date
2020

Document Version
Final published version

Published in
Physical Review Letters

License
CC BY

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Dijet Resonance Search with Weak Supervision Using $\sqrt{s}=13$ TeV $pp$ Collisions in the ATLAS Detector

G. Aad et al.*
(ATLAS Collaboration)

(Received 7 May 2020; revised 12 July 2020; accepted 4 August 2020; published 21 September 2020)

This Letter describes a search for narrowly resonant new physics using a machine-learning anomaly detection procedure that does not rely on signal simulations for developing the analysis selection. Weakly supervised learning is used to train classifiers directly on data to enhance potential signals. The targeted topology is dijet events and the features used for machine learning are the masses of the two jets. The resulting analysis is essentially a three-dimensional search $A \rightarrow BC$, for $m_A \sim \mathcal{O}$(TeV), $m_B, m_C \sim \mathcal{O}(100$ GeV) and $B, C$ are reconstructed as large-radius jets, without paying a penalty associated with a large trials factor in the scan of the masses of the two jets. The full run 2 $\sqrt{s}=13$ TeV $pp$ collision dataset of 139 fb$^{-1}$ recorded by the ATLAS detector at the Large Hadron Collider is used for the search. There is no significant evidence of a localized excess in the dijet invariant mass spectrum between 1.8 and 8.2 TeV. Cross-section limits for narrow-width $A$, $B$, and $C$ particles vary with $m_A, m_B, m_C$. For example, when $m_A = 3$ TeV and $m_B \gtrsim 200$ GeV, a production cross section between 1 and 5 fb is excluded at 95% confidence level, depending on $m_C$. For certain masses, these limits are up to 10 times more sensitive than those obtained by the inclusive dijet search. These results are complementary to the dedicated searches for the case that $B$ and $C$ are standard model bosons.

DOI: 10.1103/PhysRevLett.125.131801

A search for dijet resonances is one of the first analyses performed when a hadron collider reaches a new center-of-mass energy [1–8]. While such searches are sensitive to nearly all resonance decays $A \rightarrow BC$, dedicated searches for particular decays will always be more sensitive. This is the motivation for dedicated resonance searches for the case where $B$ and $C$ are $\tau$ leptons [9,10], $b$ quarks [11–13], top quarks [14,15], vector bosons [16,17], Higgs bosons [18–23], and more, including asymmetric combinations. In all cases, a selection on the structure of the energy flow from each side of the decay is used to enhance events with the targeted topology. Searches for any combination of standard model (SM) particles can be well-motivated by one or more theory frameworks beyond the SM (BSM), but not all combinations are currently covered by dedicated searches [24]. Furthermore, there are only a small number of searches [25–39] that cover the vast set of possibilities where at least one of $B$ or $C$ is itself a BSM particle [40]. There is no previous search where all of $A$, $B$, and $C$ are BSM particles and can have different masses.

While it is crucial to continue searching for particular dijet topologies, the fact that not all SM and BSM possibilities are covered suggests that a complementary generic search effort is required. What is needed is a method for searching for many topologies all at once that ideally does not pay a large statistical trials factor. A variety of existing and proposed model-agnostic searches range from nearly signal model independent but fully background model dependent [41–56] (because they compare data with SM simulation) to varying degrees of partial signal-model and background-model independence [57–72]. The method used for this analysis employs a machine-learning-based anomaly detection procedure to perform a dijet search in which the jets from a potential signal have a nontrivial but unknown structure [70,71]. Simply stated, classifiers are trained to distinguish particular dijet invariant mass bins from their neighbors. Localized resonances will be enhanced with a selection based on the classifier.

This Letter presents a search for a generic $A \rightarrow BC$ resonance, in which all of $A$, $B$, and $C$ could be BSM particles and $m_B, m_C \ll m_A$ so that the decay products of $B$ and $C$ can be contained within single large-radius jets. A search by the CMS Collaboration [17] involves a three-dimensional fit over jet and dijet masses, but results are reported only for the case that $B$ and $C$ are $W$ or $Z$ bosons. The analysis presented here achieves sensitivity to BSM cases without performing a scan in masses other than from the $A$ particle. In particular, the search uses events collected by the ATLAS detector [73,74] using the full 139 fb$^{-1}$ run.
$2 \sqrt{s} = 13$ TeV $pp$ collision dataset. Weakly supervised classifiers are used to enhance potential signals without using simulations of any particular signal models.

Events with at least two jets are considered, and the invariant mass distribution of the two leading jets is used to perform a “bump hunt.” Jets are formed [75,76] from locally calibrated calorimeter cell clusters [77] using the anti-$k_t$ algorithm [78] with a radius parameter of $R = 1.0$. These jets are trimmed [79] by reclustering the jet constituents with the $k_t$ algorithm using $R = 0.2$ and removing the constituents with transverse momentum ($p_T$) less than 5% of the original jet $p_T$. The jet four-vectors are then calibrated as detailed in Ref. [80]. The two jets are required to each have $p_T > 200$ GeV and pseudorapidity $|\eta| < 2.0$. In order to be broadly sensitive to hadronically decaying narrowly resonant particles, events are required to have at least one jet with $p_T > 500$ GeV and two leading jets with a rapidity difference of $|\Delta y_{JJ}| < 1.2$. The $p_T$ threshold is chosen so that the online trigger system is fully efficient [82,83]. Furthermore, both jets must have jet mass 30 GeV < $m_J$ < 500 GeV for stability of the neural network (NN) training described below. The upper threshold reduces the $m_{JJ}$ dependence of the $m_J$ distribution. The bump hunt is performed for dijet invariant masses in the range 2.28 TeV < $m_{JJ}$ < 6.81 TeV.

The masses of the two leading jets are used for classification. As the first application of fully data-driven machine-learning anomaly detection, this restricted feature set is used to establish the procedure and is already sensitive to a wide range of BSM possibilities. Weakly supervised classifiers for high-energy physics [84–87] aim to distinguish signal from background without having labeled examples for training. In particular, the classification without labels method [84] calls for two mixed samples that are statistically identical aside from different class proportions. For this search, the two samples are constructed using signal regions in $m_{JJ}$ with width 20% × $m_{JJ}$, chosen to correspond to the detector resolution for a narrow resonance. The signal regions are labeled 0–7 and have boundaries [1.90,2.28,2.74,3.28,3.94,4.73,5.68,6.81,8.17] TeV. The last bin edge is chosen so that there are sufficiently many events for training targeting the penultimate region.

The jet mass probability density varies slowly with $m_{JJ}$, so neighboring regions in $m_{JJ}$ can be used to construct the mixed event samples required for weak supervision. In particular, a network is trained to distinguish between a given $m_{JJ}$ signal region and the two neighboring sideband regions. For the case in which some signal is present in the signal region, the network will learn to tag that signal and enhance a bump in the $m_{JJ}$ spectrum, while for the case in which there is no signal in the signal region, the tagging of the network will be essentially random, and the $m_{JJ}$ spectrum will remain smooth after tagging. Since every signal region requires two neighboring sidebands, the $m_{JJ}$ regions 1–6 are chosen, and the entire process outlined below is repeated for each signal region. In order to reduce existing correlations between jet mass and $m_{JJ}$ [88], the jet masses ($m_1, m_2$), with $m_1 \geq m_2$, in each $m_{JJ}$ region are each mapped to be between 0 and 1. This mapping is accomplished with the empirical cumulative distribution function of the marginal distribution over both jets, each of mass $m_{JJ}$: $m_{\text{jet}} \mapsto R = \frac{1}{n_{\text{jets}}} \sum_{i=1}^{n_{\text{jets}}} [m_{\text{jet}} > m_{\text{jet},i}]$. The indicator function $I[\cdot]$ is unity if its argument is true and zero otherwise, and the resulting marginal distribution is uniform in each $m_{JJ}$ bin (signal sensitivity is contained in the correlations). Additional decorrelation is achieved by assigning the same total weight to each sideband. Note that full decorrelation between the classifier and $m_{JJ}$ is sufficient but not necessary: sculpting an artificial bump requires that the classifier has at least a quadratic dependence on the dijet mass so that the efficiency is lower on the left side band, higher in the signal region, and then lower again in the right sideband. The NNs performing the weakly supervised classification using $m_1$ and $m_2$ are fully connected networks built from four hidden layers with sizes 64, 32, 8, and 1. Rectified linear units connect each intermediate hidden layer and the final activation function is sigmoidal. Networks are implemented in Keras [89] with the Tensorflow back end [90] and minimize the binary cross entropy using the Adam optimizer [91].

In order to eliminate a trial factor associated with $(m_1, m_2)$, the NN identifies a region of interest, and no event is used to train the NN that is applied to it. A $k$-fold cross-validation procedure is employed in which the full dataset is divided randomly into $k$ parts of equal size. Among these, $k – 2$ parts are used for training $n$ classifiers (the training set) with different initializations, and the $(k – 1)$th part is used to decide, based on the loss, which of these $n$ networks to select (the validation set). The selected network is then mapped to an efficiency $\epsilon$ in the $k$th part (the test set) so that the meaning of the network output can be compared across datasets and trainings. The efficiency $\epsilon$ is defined as the fraction of events with a given NN value or higher. This output is averaged across the $k – 1$ other permutations of the training and validation parts. The entire procedure is then repeated $k$ times, where each part is a test set exactly once. For this analysis, $k = 5$ and $n = 3$, so there are $3 \times 4 \times 5 = 60$ NNs trained for each signal region. Two event selections from thresholds imposed on the NN outputs are used: one that keeps the 10% most signal-region-like events ($\epsilon = 0.1$) and one that keeps the 1% most signal-region-like events ($\epsilon = 0.01$).

As the classifier-based event selection depends on the data, and in particular on the possible presence of true signals, it is not possible to directly define control regions to validate the method. The entire procedure was validated using simulated events as well as a validation region with $|\Delta y_{JJ}| > 1.2$. For $s$-channel resonances, it is expected that this inverted rapidity difference requirement reduces the
signal efficiency while enhancing the dijet background by over an order of magnitude. In these validation tests, the learning works effectively and there is no evidence for selection-induced excesses. The expected limits are comparable to the ones that will be reported for the unblinded data in Fig. 3.

Following the validation, first the performance of the NNs on data is studied with and without injected signals. Since the NNs are two-dimensional functions, they can be visualized directly as images. Figure 1 presents the network output from a representative signal region in the absence of signal and also in the presence of injected signals. By construction, there must be a region of low efficiency and the data are the same in all four plots. In the absence of a signal, regions of low efficiency are located randomly throughout the \((m_1, m_2)\) plane. The signals are \(W^0 \rightarrow WZ\), for a new vector boson \(W^0\) \cite{92}, and the \(W\) and \(Z\) boson masses are varied, with widths set close to zero. These signals were simulated using \textsc{pythia 8.2} \cite{93–95} with the A14 set of tuned parameters \cite{96} and NNPDF 2.3 parton distribution function \cite{97}. All samples of simulated data were processed using the full ATLAS detector simulation \cite{98} based on \textsc{geant4} \cite{99}. The amount of signal injected in all cases is about the same as, or less than, the level already excluded by the all-inclusive dijet search \cite{100}. In all cases, the low-efficiency (signal-like) regions of the NN are localized near the injected signal. Some signals are easier to find than others; the difficulty is set both by the relative size of the signal and by the total number of events available for training in the signal vicinity.

After applying an event selection based on the NN trained on a particular signal region, the \(m_{JJ}\) spectra are fit with a parametric function. The entire \(m_{JJ}\) spectrum between 1.8 and 8.2 TeV is fit with a binning of 100 GeV; however, a fit signal region and fit sideband region are defined for evaluating the quality of the fit. The

![Figure 1](https://via.placeholder.com/150)

**FIG. 1.** The efficiency mapped output of the NN versus the input variables for the events in signal region 2 for four cases: (a) there is no injected signal; (b) there is an injected signal of \(m_A = 3\) TeV, and \(m_B = 400\) GeV and \(m_C = 80\) GeV, (c) there is an injected signal of \(m_A = 3\) TeV, and \(m_B = 200\) GeV and \(m_C = 200\) GeV, and (d) there is an injected signal of \(m_A = 3\) TeV, and \(m_B = 400\) GeV and \(m_C = 400\) GeV. The location of \((m_B, m_C)\) for the given injected signal is marked with a green \(\times\). The injected cross section is just below the limit at low \(m_B\) and \(m_C\) from the inclusive dijet search \cite{100}. Additional signal region plots in the absence of an injected signal can be found in Ref. \cite{101}.
fit signal regions are defined as the $m_{JJ}$ signal regions the NN used for training, combined with the adjacent halves of the left and right neighboring regions; the fit sidebands are defined as the complement of the fit signal regions. An iterative procedure is applied until the $p$ value from the fit sideband $\chi^2$ is greater than 0.05. Since the NN is trained to distinguish the signal region from its neighboring regions, it is expected that the $m_{JJ}$ spectrum is smooth in the fit sideband region in the presence or absence of a true signal. First, the data are fit to $dn/dx = p_1(1-x)^{p_2-\xi_1}x^{-p_3}$, where $x = m_{JJ}/\sqrt{s}$, $p_i$ are fit parameters, and the $\xi_i$ are chosen to ensure that the $p_i$ are uncorrelated. If the fit quality is insufficient, an extended function is used instead [100]: $dn/dx = p_1(1-x)^{p_2-\xi_1}x^{-p_3} + (p_4-\xi_2)p_3^{-\xi_3}x^{-\xi_4}\log(x)$. If the fit quality remains insufficient, a variation of the.

![Graphs](image-url)  

**FIG. 2.** A comparison of the fitted background and the data in all six signal regions, indicated by vertical dashed lines, and for (a),(c) $\epsilon = 0.1$ and (b),(d) $\epsilon = 0.01$. Dashed histograms represent the fit uncertainty. The lower panel is the Gaussian-equivalent significance of the deviation between the fit and data. The fits are performed including the sidebands, but only the signal region predictions and observations in each region are shown. As the NN is different for each signal region, the presented spectrum is not necessarily smooth. The top plots (a),(b) show the result without injected signal, and the bottom plots (c),(d) present the same results but with signals injected only for the NN training at $m_A = 3$ TeV (signal 1) and $m_A = 5$ TeV (signal 2), each with $m_B = m_C = 200$ GeV. The injected cross section for each signal is just below the limit from the inclusive dijet search [100].
The UA2 [2] fit function is tested: 
\[ \frac{dn}{dx} = p_1 x^p_1 e^{-p_2 x} \left( 1 + \frac{x^p_1 - x^p_2}{x^p_2} \right) \]. If the fit quality is still insufficient, the fit sidebands are reduced by 400 GeV on both sides and the three functions are tried again in order. This procedure is then iterated until the fit is successful. The fit results in the signal regions for the \( \epsilon = 0.1 \) and \( \epsilon = 0.01 \) NN efficiency selections are presented in Fig. 2. The largest positive deviation from the fit model is 3.0\( \sigma \) in signal region 1, around 2500 GeV, at \( \epsilon = 0.1 \) (the corresponding NN output does not show any significant features [101]). Globally, the positive tail of the signal region significance distribution is consistent with a standard normal distribution at the 1.5\( \sigma \) level.

The W' signal models can be used to set limits on the production cross section of specific new particles. To illustrate the sensitivity of the analysis to the full three-dimensional parameter space \( (m_A, m_B, m_C) \), two \( m_A \) points and multiple \( (m_B, m_C) \) points are selected. As the NN performance depends on the data, the entire learning procedure has to be repeated every time a new signal model and signal cross section are injected into the data. In order to reduce statistical fluctuations related to the shape of the signal, for each signal cross section the network is retrained with five random samplings from the signal simulation, and the network with the median performance is chosen. A profile-likelihood-ratio test is used to determine 95\% confidence intervals for the excluded signal cross section. When the number of expected events is much larger than one, asymptotic formulas [102] are used for this test, otherwise, the test is performed numerically [only Fig. 3(d)]. The excluded cross section is reported as \( \max(\sigma_{CL}, \sigma_{injected}) \), where \( \sigma_{CL} \) is the cross section determined from the profile-likelihood-ratio test and \( \sigma_{injected} \) is the injected cross section. This procedure is chosen because the network’s performance may not be as good if there were truly less signal than was injected. The resulting exclusion limits are presented in Fig. 3. As the background expectation is determined entirely from data, the only systematic

![Graphs and tables showing results](image-url)
uncertainty associated with the background is the statistical uncertainty from the fit. The only other relevant uncertainties are those related to the signal $m_{jj}$ and $m_J$ modeling; experimental uncertainties in the reconstructed jet kinematics account for about a 10% uncertainty in the excluded cross section.

The limits on $W^*$ production vary with $m_A$, $m_B$, and $m_C$. For $m_B = m_C = 400$ GeV, the excluded cross section is about 1 fb, a significant improvement over existing limits. Lower $m_B$ and $m_C$ result in weaker limits because of the larger SM background in those regions; it is therefore difficult for the NN to learn to tag these signals. The NN is most powerful when the local signal-to-background ratio is high and there are enough events for it to learn effectively. For some models, such as $(m_A, m_B, m_C) = (5000, 80, 80)$ GeV, the NN is not able to identify the signal effectively, resulting in limits weaker than those from previous searches. For comparison, the sensitivities of the ATLAS inclusive dijet search (recast with signals from this Letter) [103] and the all-hadronic diboson resonance search [100] are also shown in Fig. 3. The inclusive dijet search sensitivity decreases for high $m_B$ and $m_C$ masses due to the use of small-radius jets that do not capture all of the $B$ and $C$ decay products. The diboson resonance search has greater sensitivity when $m_B, m_C \approx m_W, m_Z$, but it has no sensitivity away from these points. In this case, the diboson search uses more information than the weakly supervised one, but the trend is expected: assuming that the simulations used for developing the analysis selection are reliable, a fully supervised approach should outperform the weakly supervised one for any particular signal model. Direct searches for $B$ and $C$ that trigger on initial-state radiation are also sensitive to these signal models [34–39], but the sensitivity is much weaker than 10 fb.

While the regions are chosen with definite boundaries, the analysis is sensitive to signals across the entire range. In particular, it is found that for nearly 75% of the range, the efficiency for a signal is unaffected by a shifted peak location. This efficiency is everywhere above half of the nominal efficiency.

In conclusion, this Letter presents a model-agnostic resonance search in the all-hadronic final state using the full LHC run 2 $pp$ dataset of the ATLAS experiment. Weakly supervised classification NNs are used to identify the presence of potential signals without training on simulations of any particular signal models. For jets produced from Lorentz-boosted heavy-particle decays, this search is more sensitive than the inclusive dijet search and extends the coverage of the all-hadronic diboson search to regions away from the SM boson masses. This is the first search that covers $A \rightarrow BC$ production where all of $A$, $B$, and $C$ are BSM particles that can have different masses. The feature space used by the NNs is only two dimensional, so there is great potential to extend this method to include additional features and more final states in order to ensure broad coverage of unanticipated scenarios.

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently. We acknowledge the support of ANPCyT, Argentina; YerPhi, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; STh, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC, and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST, and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR, and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russia Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; MSTNF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF, and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; ERC, ERDF, Horizon 2020, Marie Skłodowska-Curie Actions and COST, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales, and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; CERCA Programme Generalitat de Catalunya and PROMETEO Programme Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK), and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [105].


[20] CMS Collaboration, Search for resonances decaying to a pair of Higgs bosons in the $b\bar{b}q\bar{q}l^+l^-$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 10 (2019) 125.


[23] CMS Collaboration, Search for heavy resonances decaying into two Higgs bosons or into a Higgs boson and a W or Z boson in proton-proton collisions at 13 TeV, J. High Energy Phys. 01 (2019) 051.


[38] CMS Collaboration, Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at \( \sqrt{s} = 13 \) TeV, J. High Energy Phys. 01 (2018) 097.


[81] Pseudorapidity is defined in terms of the angle $\theta$ relative to the beam line as $\eta = -\ln \tan(\theta/2)$.
[88] This is similar to the method used in Ref. [106]; additional decorrelation techniques are described in Refs. [107–116].


PHYSICAL REVIEW LETTERS 125, 131801 (2020)