Search for boosted resonances decaying to two $b$-quarks and produced in association with a jet at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

A search for new resonances decaying into a pair of bottom quarks in the merged regime is reported using an integrated luminosity of 80.5 fb$^{-1}$ of proton-proton collisions at $\sqrt{s} = 13$ TeV recorded by the ATLAS detector at the Large Hadron Collider. For the resonance to be merged, an extra jet with high transverse momentum is required. The mass range from 70 to 230 GeV is inspected. For the Standard Model Higgs boson, the observed signal strength is $\mu_H = 5.8 \pm 3.1$ (stat.) $\pm 1.9$ (syst.) $\pm 1.7$ (th.), consistent with the background-only hypothesis at 1.6 standard deviations. No evidence of a significant excess of events beyond the expected background is found and limits on leptophobic $Z'$ bosons with democratic axial couplings to all quark generations are set in the probed mass range.

© 2018 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.
1 Introduction

Analyses searching for dijet resonances have been thoroughly explored by the LHC experiments [1–3]. The main focus of these analyses are resonances at high masses, with the lowest mass boundary being dictated by the trigger and data-acquisition systems of the experiments. Searching for resonances below the TeV mass range is well motivated and alternative approaches employing more sophisticated trigger or analysis strategies has flourished within the ATLAS and CMS collaborations when analysing the $pp$ collision data provided by the LHC [4–11].

An analysis searching for boosted resonances decaying to a pair of bottom quarks using an integrated luminosity of 80.5 fb$^{-1}$ of proton-proton collisions at $\sqrt{s} = 13$ TeV is presented. For the resonance to be boosted, an extra jet with high transverse momentum ($p_T$) is required to be present in the event. The decay products of a hadronically-decaying resonance can be reconstructed in a single jet with a large radius parameter (large-$R$ jet), vastly reducing the amount of background from Standard Model (SM) processes. The final state of interest therefore consists of a signal large-$R$ jet plus an additional jet. The additional jet is also taken with a large-radius cone to simplify the analysis and the propagation of the related uncertainties.

Many physics models predict low-mass dijet resonances decaying to bottom quarks. Yukawa couplings being proportional to the mass of the decay particle favour decays to $b$-quarks if the resonance mass is lower than two-times the top-quark mass, as is true for the Higgs boson. Although the Higgs boson has been discovered in Refs. [12, 13] and its coupling with bottom quarks recently observed in Refs. [14, 15], an inclusive analysis also targeting the gluon-fusion production mode and the decay to $b$-quarks was conducted only recently in Ref. [16]. At high $p_T$, the Higgs boson production starts being sensitive to the top-quark loop in the gluon-fusion production mode [17], including possible contributions from new resonances, and to the presence of anomalous couplings that give rise to an effective gluon-gluon-Higgs interaction. In the presence of new physics contributions, increases in the high-$p_T$ Higgs boson production are predicted to be as high as 50% [18]. A leptonophobic $Z'$ decaying to two quarks with a mass ranging between 100 and 200 GeV is also searched for.

2 The ATLAS detector

The ATLAS experiment [19] at the LHC is a multi-purpose particle detector with a forward-backward symmetric cylindrical geometry and a near 4$\pi$ coverage in solid angle. It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadronic calorimeters, and a muon spectrometer. The inner tracking detector covers the pseudorapidity range $|\eta| < 2.5$. It consists of, in ascending order of radius from the beam-line, silicon pixel, silicon microstrip (SCT), and transition radiation tracking detectors. The pixel detectors are crucial for $b$-jet identification. For the second LHC data-taking period at $\sqrt{s} = 13$ TeV a new inner pixel layer, the Insertable B-Layer (IBL) [20, 21], was added at a mean sensor radius of 3.2 cm from the beam-line. The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) electromagnetic calorimeters. An

---

1 ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z axis along the beam pipe. The x-axis points to the centre of the LHC ring, and the y axis points upwards. Cylindrical coordinates ($r, \phi$) are used in the transverse plane, $\phi$ being the azimuthal angle around the z-axis. The pseudorapidity $\eta$ is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$.
additional thin LAr presampler, covering $|\eta| < 1.8$, is used to correct for energy loss in material upstream of the calorimeters. Hadronic calorimetry is provided by a steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadronic endcap calorimeters. The solid-angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimised for electromagnetic and hadronic measurements, respectively. The muon spectrometer is composed of separate trigger and high-precision tracking chambers, measuring the deflection of muons in a magnetic field generated by superconducting air-core toroids. The precision chamber system covers the region $|\eta| < 2.7$ with three layers of monitored drift tubes, complemented by cathode strip chambers in the forward region, where the particle flux is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive plate chambers in the barrel, and thin-gap chambers in the endcap regions. A two-level trigger system, using custom hardware followed by a software-based level, is used to reduce the event rate to a maximum of around 1 kHz for offline storage [22].

3 Data and simulated event samples

The data were collected by the ATLAS detector in $pp$ collisions at the LHC with a centre-of-mass energy of 13 TeV during the years 2015, 2016 and 2017. The data used correspond to an integrated luminosity of 80.5 fb$^{-1}$. Events used for this search were recorded using a trigger requiring the presence of at least one reconstructed anti-$k_t$ jet with radius parameter $R = 1.0$ [23]. The trigger threshold ranged between $p_T > 360$ GeV and 460 GeV across the years and was adapted following the different instantaneous luminosity profiles and the inclusion of pile-up suppressing techniques at the trigger level [24].

Monte Carlo (MC) samples of the expected backgrounds are used for developing the non-resonant background estimation procedure and for modelling the resonant backgrounds themselves, as detailed in Section 6.

Simulated QCD dijet events were generated using the Pythia 8.186 [25] generator with the NNPDF 2.3 at next-to-leading order (NLO) parton distribution function [26] and the A14 set of tuned parameters (tune) [27]. To maintain a constant statistical power over a large energy range, the weighted events were generated with a flat jet $p_T$ spectrum and split into several samples. The hadronically decaying $W$ and $Z$ events, with up to 4 additional partons, were generated using Sherpa 2.1.1 [28] with the CT10 [29] parton distribution function. Matrix elements were calculated at leading order (LO) and the cross sections were corrected to NLO using $K$-factors derived from corresponding samples with leptonic vector-boson decays generated at NLO. The procedure, described in Ref. [5], results in a $K$-factor of 1.28 for $W$ + jets and 1.38 for $Z$ + jets with relative uncertainties of 10% based on the analysis described in Ref. [30]. To evaluate the uncertainty due to the choice of MC generator, an alternative set of hadronically decaying $W$ and $Z$ samples were generated using Herwig++ 2.7 [31] with the CTEQ6L1 [32] parton distribution function.

The simulated $t\overline{t}$ events were generated at tree-level using Powheg-Box 2 [33–35] and the NNPDF 2.3 NLO parton distribution function. The hadronization was performed using Pythia 8.230 [36] with the A14 tune and the NNPDF 2.3 LO parton distribution function. To evaluate the uncertainty due to the choice of MC generator, a second set of $t\overline{t}$ events was generated by Sherpa 2.2.1 using the NNPDF 3.0 NNLO parton distribution functions [37]. A sample containing single-top events in the $Wt$ channel was generated at tree-level using Powheg-Box 2 with the NNPDF 3.0 parton distribution function. The process is showered using Pythia 8.230 configured with the A14 tune and the NNPDF 2.3 LO parton distribution function.

3
The signal samples considered are the Standard Model Higgs boson decaying to two $b$-quarks and a leptophobic $Z'$ with democratic axial coupling ($g_q$) to all quark generations. Several Higgs boson samples were simulated corresponding to the three main production mechanisms; gluon-gluon fusion ($ggF$) and Higgs-strahlung ($W/\bar{Z}$ associated production). All contribute in a non-negligible manner to the signal region. The $ggF$ and $VBF$ Higgs boson events were generated at NLO with top-mass-effect corrections following the MiNLO procedure [38, 39] using Powheg-Box 2 [40, 41] with the NNPDF 3.0 NNLO parton distribution function. They were showered using Pythia 8.212 with the AZNLO [42] tune and the CTEQ6L1 parton distribution function. The Higgs boson events produced in association with a $W$ or $Z$ boson and initiated by quarks were generated using Pythia 8.212 with the AZNLO tune and the CTEQ6L1 parton distribution function. The cross-sections are then scaled to the LHCHXSWG recommendation [43], including extra factors to account for $ZH$ production initiated by gluons, which were not generated in the utilized samples. All decay modes of the vector boson were enabled. The MiNLO procedure, with finite top-mass effects, is known to be a good approximation of the $ggF$ Higgs production at NNLO with parton shower matching and at high transverse momentum, with an inclusive cross section of $29^{+24}_{-21}\%$ fb for matrix-element level Higgs bosons with $p_T > 400$ GeV, which is consistent with fixed-order calculations at NNLO with top-mass effects corrections [39].

The events containing a leptophobic $Z'$ decaying to two quarks and produced in association with a jet were generated using a Simplified Dark Matter (DM) model [44] implemented in MadGraph5_aMC@NLO [45] with the NNPDF 3.0 LO parton distribution function. The showering was performed using Pythia 8.212 with the A14 tune and the NNPDF 2.3 LO parton distribution function. Signal samples were generated at several $Z'$ masses spanning the search region. The values of $m_{Z'}$ are between 100 and 200 GeV in steps of 25 GeV. The simplified model defines an absolute axial coupling, $g_q$, to a quark-antiquark pair that is taken to be the same for all generations. The signal events were simulated with $g_q = 0.25$. The specific value does not have a significant effect on the kinematics of the reconstructed events because the natural width of the $Z'$ is smaller than the large-$R$ jet mass resolution. The cross-section at other values can be obtained by scaling as $g_q^2$. The calculation was implemented using the DM width calculator in Ref. [46]. The DM particle mass is set to 10 TeV to not alter the $Z'$ branching ratios.

All simulated events were created using the full detector simulation in Geant 4 [47, 48] and the same event reconstruction algorithms were used as for the data. The decay of $b$- and $c$-hadrons, with the exception of the events generated by Sherpa, is consistently performed using EvtGen [49]. To account for the presence of multiple interactions in the same and neighbouring bunch crossings (pile-up), all samples were overlayed with minimum bias events simulated using Pythia 8.1 with the A3 tune [50] and the NNPDF 2.3 LO parton distribution function. All simulated events are weighted to match the pile-up distribution observed in the data.

### 4 Event reconstruction

The analysis makes use of the reconstruction of large-$R$ jets with $b$-tagging applied to the associated small-$R$ jets. The associated small-$R$ jets are track-jets with the radius dependent on their $p_T$ (variable-radius track-jets). Muons are used to define a $t\bar{t}$ control region as detailed in Section 6.

Collision vertices are reconstructed from at least two reconstructed tracks with transverse momentum $p_T > 0.5$ GeV. The primary vertex is selected as the one with the highest $\sum p_T^2$, calculated considering all associated tracks with $p_T > 0.5$ GeV.
Large-radius jets are reconstructed from topological clusters of energy deposited in the calorimeters using the anti-\(k_T\) algorithm with radius parameter of \(R = 1.0\) and are trimmed [51] to improve the mass resolution and reduce dependence on pile-up. The input clusters are calibrated using the local calibration method [52]. The trimming is achieved by re-clustering the constituents of the initial jet, using the \(k_T\) algorithm [53, 54], into smaller \(R = 0.2\) subjets and then by removing any subjet that has a \(p_T\) less than 5% of the \(p_T\) of the parent jet [55]. The jet energy and pseudorapidity is corrected using \(p_T\)- and \(\eta\)-dependent calibrations derived from simulation [56]. The jet mass is calibrated by deriving \(p_T\)- and \(\eta\)-dependent calibration factors and by combining calorimeter and tracking information as described in Ref. [57]. A smaller mass resolution and a better estimate of the median mass value is obtained than using only calorimeter energy clusters. Only trimmed large-\(R\) jets with \(p_T > 250\) GeV and \(|\eta| < 2\) are considered in the final selection.

Track-jets are clustered using the variable-radius jet algorithm [58, 59] with \(\rho = 30\) GeV, \(R_{\text{min}} = 0.02\) and \(R_{\text{max}} = 0.4\) on a subset of the reconstructed tracks. The \(R_{\text{min}}\) and \(R_{\text{max}}\) parameters, respectively, control the minimum and maximum cone size, preventing the jets from becoming too large at low \(p_T\) and going beyond the detector resolution at high \(p_T\). The \(\rho\) parameter instead determines how fast the jet size decreases with the jet \(p_T\). The considered tracks are required to have \(p_T > 0.5\) GeV and \(|\eta| < 2.5\), at least 7 hits in the SCT and pixel detector layers, no more than one hit shared by multiple tracks in the pixel detector, no more than one missing hit in the pixel detector and no more than two missing hits in the SCT detector. Fake tracks and tracks originating from pile-up vertices are suppressed by requiring the longitudinal impact parameter (\(z_0\)) to be within \(|z_0 \cdot \sin \theta| < 3\) mm of the primary vertex. Only track-jets with \(p_T > 10\) GeV, \(|\eta| < 2.5\), at least two track constituents and satisfying \(\Delta R / \min R > 1\), with \(\min R\) the smallest radius of the track-jets and \(\Delta R\) the distance among any pair of them, are considered. The last requirement discards collinear track-jets to prevent anomalous \(b\)-tagging response. The signal efficiency is not affected by imposing this requirement. The reconstructed track-jets are ghost-associated [60] to the large-\(R\) jets before trimming.

Track-jets containing \(b\)-hadrons are identified using the multivariate \(b\)-tagging algorithm MV2c10 [61], which makes use of observables with discriminating information provided by an impact parameter algorithm, an inclusive secondary vertex finding algorithm and a multi-vertex finding algorithm. The MV2c10 algorithm is applied to a set of charged-particle tracks that satisfy quality and impact-parameter criteria and are matched to reconstructed track-jets. Two \(b\)-tagging requirements are employed in the analysis for defining the signal and control regions that, respectively, correspond to efficiencies of 77% and 85% for \(b\)-jets with \(p_T > 20\) GeV when evaluated in a sample of simulated \(\tau\bar{\tau}\) events. These two requirements are referred to as ‘tight’ and ‘loose’ working points in the following. Scale factors to account for efficiency differences observed in data and simulated events are applied.

Muons are reconstructed by combining tracks in the inner detector with tracks in the muon spectrometer and are required to have \(p_T > 10\) GeV, \(|\eta| < 2.5\) and pass the ‘medium’ identification and the ‘loose’ isolation working points [62] to suppress fake reconstructed muons.

### 5 Event selection

Events containing jets induced by calorimeter noise bursts, beam-induced background or cosmic rays are removed using the same criteria as in Ref. [63]. At least one reconstructed vertex and at least two large-\(R\) jets are then required. Events are further preselected by requiring the leading large-\(R\) jet to satisfy \(p_T > 480\) GeV and the sub-leading one \(p_T > 250\) GeV. The requirement on the leading large-\(R\) jet ensures full trigger efficiency across years.
Figure 1: (a) Predicted flavour composition of the dijet background in the SR based on the truth-matched hadron content of the two leading-$p_T$ track-jets associated to the signal candidate large-$R$ jet, with the B/C labels indicating the presence of a $b/c$-quark and L indicating the presence of a light quark or a gluon. (b) The expected shape of the dijet background in the SR and CR$_{QCD}$ normalised to the same event count between 70 GeV $< m_J <$ 230 GeV.

Next the large-$R$ jet assumed to contain the decay products of the $Z'$ or Higgs boson, referred to as the signal jet candidate in the following, is chosen as the highest-$p_T$ large-$R$ jet among those passing the following requirements: large-$R$ jet $p_T > 480$ GeV, at least two associated track-jets and $2m_J/p_T < 1$ with $m_J$ being the invariant mass of the large-$R$ jet. The last requirement ensures that boosted decays are selected.

The events are further classified based on the number of $b$-tagged jets in the two leading-$p_T$ track-jets associated to the signal jet candidate. Events with exactly 0 loose $b$-tagged track-jets form the control region for the non-resonant background estimation studies (CR$_{QCD}$). Events with exactly 2 tight $b$-tagged track-jets form the signal region (SR). The 77% $b$-tagging working point used in the SR was optimised for best signal significance. The mass range considered in the analysis is from 70 GeV to 230 GeV. The lower bound is selected to avoid turn-on effects while the upper bound excludes high masses not resulting in merged decays.

Figure 1 (left) shows how the dijet background in the SR is dominated by double $b$-quarks plus combinations of ($b$- or $c$-) quarks. Events with no heavy ($b$- or $c$-) quarks make up only 5%. Figure 1 (right) shows the simulated distribution of the signal candidate large-$R$ jet mass of the dijet background in the SR and CR$_{QCD}$ and how the shapes are close to each other in the mass range of interest. The signal region has a higher turn-on curve ending roughly at 70 GeV which is not included in the analysis.

The efficiencies and yields in the CR$_{QCD}$ and SR for the considered resonant backgrounds and the Higgs boson and $Z'$ signals are shown in Table 1. The composition of the vector boson, $t\bar{t}$ and $H \rightarrow b\bar{b}$ resonant components in the SR and CR$_{QCD}$ is shown in Table 2. The vector boson component is dominated by the $W$ boson in the CR$_{QCD}$. However, as expected by imposing the $b$-tagging requirement, $Z \rightarrow b\bar{b}$ is the main contribution in the SR. The $t\bar{t}$ contribution is divided equally in the two regions with $\sim 60\%$ coming from hadronic decays, $\sim 40\%$ from semi-leptonic events and only a small percentage from di-leptonic decays of the top quarks. The $H \rightarrow b\bar{b}$ component is mainly produced via $ggF$ in all regions, contributing 53% of the Higgs boson signal in the SR. The alternative production mechanisms also contribute in a non-negligible manner; VBF production contributes 25% and Higgs-strahlung the remaining 22%. This
<table>
<thead>
<tr>
<th>Process</th>
<th>CRQCD Eff. (%)</th>
<th>CRQCD Yield 80.5 fb^{-1}</th>
<th>CRQCD Yield 1.4 fb^{-1}</th>
<th>SR Eff. (%)</th>
<th>SR Yield 80.5 fb^{-1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z \to q\bar{q} + jets</td>
<td>46.2</td>
<td>84400</td>
<td>1470</td>
<td>3.4</td>
<td>6200</td>
</tr>
<tr>
<td>W \to q\bar{q} + jets</td>
<td>51.3</td>
<td>219000</td>
<td>3810</td>
<td>0.4</td>
<td>1500</td>
</tr>
<tr>
<td>t\bar{t}</td>
<td>25.9</td>
<td>110900</td>
<td>1929</td>
<td>2.5</td>
<td>10550</td>
</tr>
<tr>
<td>H \to bb (ggF)</td>
<td>23.6</td>
<td>140</td>
<td>2</td>
<td>19.4</td>
<td>115</td>
</tr>
<tr>
<td>H \to bb (VBF)</td>
<td>15.8</td>
<td>41</td>
<td>1</td>
<td>20.7</td>
<td>53</td>
</tr>
<tr>
<td>H \to bb (WH)</td>
<td>32.4</td>
<td>71</td>
<td>1</td>
<td>12.0</td>
<td>26</td>
</tr>
<tr>
<td>H \to bb (ZH)</td>
<td>30.5</td>
<td>40</td>
<td>1</td>
<td>15.8</td>
<td>21</td>
</tr>
<tr>
<td>H \to bb (Total)</td>
<td>24.3</td>
<td>292</td>
<td>5</td>
<td>17.9</td>
<td>216</td>
</tr>
<tr>
<td>Z' (m = 100 GeV)</td>
<td>43.9</td>
<td>87700</td>
<td>1530</td>
<td>4.1</td>
<td>8200</td>
</tr>
<tr>
<td>Z' (m = 125 GeV)</td>
<td>43.6</td>
<td>82800</td>
<td>1440</td>
<td>3.8</td>
<td>7300</td>
</tr>
<tr>
<td>Z' (m = 150 GeV)</td>
<td>43.5</td>
<td>78300</td>
<td>1360</td>
<td>3.7</td>
<td>6700</td>
</tr>
<tr>
<td>Z' (m = 175 GeV)</td>
<td>42.8</td>
<td>71100</td>
<td>1240</td>
<td>3.3</td>
<td>5550</td>
</tr>
<tr>
<td>Z' (m = 200 GeV)</td>
<td>41.8</td>
<td>64800</td>
<td>1127</td>
<td>3.2</td>
<td>4910</td>
</tr>
<tr>
<td>Data</td>
<td>38.7</td>
<td>29883000</td>
<td>519710</td>
<td>0.6</td>
<td>484600</td>
</tr>
</tbody>
</table>

Table 1: The efficiencies and yields in the CRQCD and SR for the non-QCD background, the Higgs boson and Z' boson signals and data. The yields in the CRQCD are given for the full luminosity 80.5 fb^{-1} and for the luminosity employed, as explained in Section 6, for the background estimate of the non-resonant dijet process. The efficiencies are relative to leading large-R jet p_T > 480 GeV requirement.

<table>
<thead>
<tr>
<th>Process</th>
<th>CRQCD</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>V + jets</td>
<td>Z + jets</td>
<td>0.28</td>
</tr>
<tr>
<td>W + jets</td>
<td>0.72</td>
<td>0.20</td>
</tr>
<tr>
<td>t\bar{t}</td>
<td>All hadronic</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Semi-leptonic</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Dileptonic</td>
<td>0.04</td>
</tr>
<tr>
<td>H \to bb</td>
<td>ggF</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>VBF</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>WH</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>ZH</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Table 2: The fractional composition of the different resonant contributions in the defined regions. The fraction is evaluated using the given contribution type as the total.

indicates that a dedicated Higgs boson analysis with selection optimized for each of the production modes could improve the sensitivity of the search.

### 6 Background modelling

The dominant background contribution in the SR is the non-resonant dijet process. Its estimate through MC is not reliable due to the statistical precision and the underlying accuracy of the event generation. A data-driven estimate is therefore employed by fitting the mass distribution m_J in the SR with a parametric function and by validating the procedure using the data in the CRQCD. The shapes of the Z + jets and W + jets contributions are taken from MC while their normalisation is fitted to data. The t\bar{t} contribution is
also estimated from MC with extra scale factors taken from a specific control region, denoted as CR_{t\bar{t}} in the following.

The background estimation procedure is tested on the CR_{QCD}. It is found that the statistical precision in \( \sim 1 \text{ fb}^{-1} \) of CR_{QCD} data is comparable to the SR when using the luminosity of 80.5 fb\(^{-1}\). The CR_{QCD} is therefore sliced into independent datasets used as probes for fitting the dijet background in the SR. The CR_{QCD} slices are constructed by grouping adjacent data runs. A total of about 60 slices is formed, with an average of 1.4 fb\(^{-1}\) of data per slice.

To find a parametric function that fits the dijet background well, the following family of polynomial exponential functions is considered:

\[
f_n(x; \vec{\theta}) = \theta_0 \exp \left( \sum_{i=1}^{n} \theta_i x^i \right),
\]

where \( x = (m_J - 150 \text{ GeV})/80 \text{ GeV} \) and \( \theta_i \) are the parameters of the fit. The parameterization of the polynomial exponential function maps the independent variable to \( x \in [-1, 1] \) for the fit range of [70, 230] GeV, which is empirically seen to provide improved numerical stability.

A slice is chosen randomly for performing studies to select the appropriate fit function. The Z + jets, W + jets and \( t\bar{t} \) contributions, scaled by their cross sections times the luminosity, are subtracted. The \( t\bar{t} \) contribution is corrected using the scale factor derived from the CR_{t\bar{t}}. A likelihood ratio test, in combination with Wilks’ theorem [64], and the F-test [65] are both used to determine the minimum number of model parameters to describe the shape of the distribution. With the former preferring a five-parameter model and latter favouring a four-parameter model, the five-parameter model is chosen to be conservative.

The fit is performed by adding the templates for the Z + jets, W + jets, \( t\bar{t} \) and single top contributions. The fit function is validated using the data slices previously described. Within the statistical precision given by the different data slices, the \( \chi^2/ndf \) from the individual fits follows the expected distribution of a good fit.

Boosted \( t\bar{t} \) events represent a significant background contribution in the SR. The \( t\bar{t} \) yield in the SR is corrected with a data-driven normalization scale factor obtained by fitting the \( t\bar{t} \) MC template in the CR_{t\bar{t}}. The CR_{t\bar{t}} is achieved by requiring a muon from the semi-leptonic decay from the top quark in the opposite hemisphere of the top selected as the signal jet candidate.

The \( t\bar{t} \) events show the expected back-to-back topology of the muon and signal jet candidate, hence a \( \Delta \phi > 2\pi /3 \) requirement between the muon and the signal jet candidate greatly reduces the contribution from muons coming from dijet and V + jets processes. Additional requirements on muon \( p_T > 40 \text{ GeV} \) and \( \Delta R < 1.5 \) between the muon and the closest jet, plus a b-tagged jet at the loose working point are required to eliminate soft muons from dijet events. The resulting CR_{t\bar{t}} is 90% pure in \( t\bar{t} \) events. The \( t\bar{t} \) MC template is fitted to the CR_{t\bar{t}} data in the large-\( R \) mass region between 100 and 200 GeV. A normalisation factor of 0.84 is extracted and used to constrain the \( t\bar{t} \) contribution in the final fit to the SR.

7 Systematic uncertainties

Sources of systematic uncertainty that could potentially result in biases in the search are related to experimental calibrations and to MC modeling of signal and background processes. These uncertainties
contribute both to the uncertainties in the overall yield (“normalisation”) and differential shape of the large-$R$ jet mass observable (“shape”) that is used in the statistical procedure for the final search.

Systematic uncertainties associated with the QCD background are estimated with pseudoexperiments. Poissonian toys are generated from the QCD component of the fit in the SR containing all components (QCD, $V+\text{jets}$, $t\bar{t}$ and exotic signal or Higgs boson components) and no nuisance parameters. The toys are refitted with the nominal QCD function, and an alternative parametric function (formal Laurent series [66]). These two sets of fits to toys, nominal and alternative, provide a measure of the statistical uncertainty on the QCD parametrization (the spread of fit parameters) and of the systematic uncertainty from the choice of fitting function (the difference between the two fitted shapes).

Uncertainties affecting all MC templates are related to the large-$R$ jet energy and mass calibrations [67] and the calibration of the MV2c10 algorithm [61], which impacts different jet flavors with different magnitudes. Large-$R$ jet energy and mass calibration uncertainties affect both the shape and normalisation of the utilized MC templates, therefore, their impact on the analysis is assessed by varying the jet energy and mass within their uncertainties and propagating those variations through the analysis chain. The effect of the jet energy resolution uncertainty is also tested and found to be negligible. The impact of uncertainties on the calibration of MV2c10 have been found to be independent of the large-$R$ jet mass for both the signal and background samples, and hence only affecting the signal normalisation.

Extra shape uncertainties are applied to the $V+\text{jets}$ and $t\bar{t}$ MC templates related to modeling uncertainties. To assess these uncertainties, the large-$R$ jet mass shapes from two different MC generators are compared in both cases. For the $V+\text{jets}$ component, the nominal shape, generated using Sherpa 2.1.1, is compared to an alternative shape, generated using Herwig++ 2.7. Similarly, for the $t\bar{t}$ component, the comparison is performed between the nominal Powheg-Box 2 shape, and the Sherpa 2.2.1 alternative shape.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>V+jets</th>
<th>Higgs</th>
<th>Z' (100 GeV)</th>
<th>Z' (175 GeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jet energy and mass scale</td>
<td>Norm. &amp; Shape</td>
<td>15%</td>
<td>14%</td>
<td>23%</td>
<td>18%</td>
</tr>
<tr>
<td>Jet mass resolution</td>
<td>Norm. &amp; Shape</td>
<td>20%</td>
<td>17%</td>
<td>30%</td>
<td>20%</td>
</tr>
<tr>
<td>$V+\text{jets}$ modeling</td>
<td>Shape</td>
<td>9%</td>
<td>4%</td>
<td>4%</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>$t\bar{t}$ modeling</td>
<td>Shape</td>
<td>&lt; 1%</td>
<td>1%</td>
<td>&lt; 1%</td>
<td>11%</td>
</tr>
<tr>
<td>$b$-tagging ($b$)</td>
<td>Normalisation</td>
<td>11%</td>
<td>12%</td>
<td>11%</td>
<td>15%</td>
</tr>
<tr>
<td>$b$-tagging ($c$)</td>
<td>Normalisation</td>
<td>3%</td>
<td>1%</td>
<td>3%</td>
<td>5%</td>
</tr>
<tr>
<td>$b$-tagging ($l$)</td>
<td>Normalisation</td>
<td>4%</td>
<td>1%</td>
<td>4%</td>
<td>7%</td>
</tr>
<tr>
<td>$t\bar{t}$ scale factor</td>
<td>Normalisation</td>
<td>2%</td>
<td>3%</td>
<td>2%</td>
<td>58%</td>
</tr>
<tr>
<td>Luminosity</td>
<td>Normalisation</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>Alternative QCD function</td>
<td>Norm. &amp; Shape</td>
<td>4%</td>
<td>4%</td>
<td>3%</td>
<td>17%</td>
</tr>
<tr>
<td>W/Z and QCD (Theory)</td>
<td>Normalisation</td>
<td>14%</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Higgs (Theory)</td>
<td>Normalisation</td>
<td>–</td>
<td>30%</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 3: Summary of the impact of the main systematic uncertainties on the uncertainty $\sigma$ on the measurement of the signal strength $\mu$ for the $V+\text{jets}$, Higgs boson and $Z'$ signals.

As previously mentioned, the $t\bar{t}$ normalisation in the signal region is constrained with the scale factor obtained in the $t\bar{t}$ enriched control region. The $t\bar{t}$ scale factor has been measured with an uncertainty of 13%, which is treated as a systematic uncertainty on the $t\bar{t}$ normalisation.
The uncertainty in the integrated luminosity is 2.1%. It is derived following a methodology similar to that detailed in Ref. [68] and it is applied to all processes modelled using MC simulations.

In order to compare the observed number of events from the $V + \text{jets}$ and Higgs components of the signal region fit, we apply theory uncertainties to their normalisations. Theory uncertainties on the $V + \text{jets}$ processes come from the impact of higher order electroweak and QCD corrections to their differential cross sections [69]. Among the Higgs boson production theory uncertainties, the gluon-gluon fusion uncertainty is dominant. It is taken as 30%, consistent with the cross section uncertainty as calculated with the MiNLO procedure and including top-mass effects for Higgs bosons with $p_T > 400 \text{ GeV}$. The same uncertainty is also applied to the other production mechanisms at high $p_T$. This results in a total uncertainty on the Higgs cross-section of 30%.

The effects of the main systematic uncertainties on the measurement of the signal strength $\mu$, defined as the scale factor multiplying the cross section times branching fraction predicted by the signal hypothesis, for the $V + \text{jets}$, Higgs boson and $Z'$ signals are displayed in Table 3. The impact of a systematic uncertainty is defined as the difference in quadrature between the uncertainty $\sigma$ in $\mu$ computed when all other uncertainties are considered and when are fixed to their pre-fit values. The total systematic uncertainty is then defined as the difference in quadrature between the total uncertainty in $\mu$ and the total statistical uncertainty, denoted as $\sqrt{\Delta \sigma^2}$ in the table.

### 8 Results

To search for and quantify the significance of $V + \text{jets}$ production and to set limits on the Higgs boson and $Z'$ boson models, a Bayesian method is used to calculate a posterior likelihood as a function of number of signal events for the signal hypotheses under consideration [70]. Systematic uncertainties enter the likelihood as nuisance parameters. In this method, the Bayes' theorem is utilised to build the final conditional probability of the parameters of interest given the observed data by integrating over the nuisance parameters (“marginalisation”) with a Markov Chain Monte Carlo procedure. The posterior probability distribution is used to gauge the fitted signal statistical significance or to set 95% credibility-level upper limits on the cross-section times acceptance times efficiency.

The first result obtained is the combined SM fit of the $V + \text{jets}$, $H + \text{jets}$ and $t\bar{t}$ plus QCD components. With this fit, the signal strengths of the $V + \text{jets}$ and $H + \text{jets}$ processes, which are not constrained to positive values, are obtained simultaneously. The normalisation of the $t\bar{t}$ component is constrained with the scale factor obtained in the dedicated CR$_{t\bar{t}}$. The comparison to data after the marginalisation of the nuisance parameters is shown in Figure 2. The observed signal strength for the $V + \text{jets}$ process is $\mu_V = 1.5 \pm 0.22 \text{ (stat.)} \pm 0.29 \text{ (syst.)} \pm 0.18 \text{ (th.),}$ corresponding to a significance of 5 standard deviations. For the $H + \text{jets}$ process, the observed signal strength is $\mu_H = 5.8 \pm 3.1 \text{ (stat.)} \pm 1.9 \text{ (syst.)} \pm 1.7 \text{ (th.),}$ consistent with the background-only hypothesis at 1.6 $\sigma$. The combined likelihood of $\mu_V$ and $\mu_H$ is shown in Figure 3.

The BumpHunter search procedure [71, 72], assuming the $V + \text{jets}$ and $t\bar{t}$ templates, is then performed for looking for possible extra exotic signals in the large-$R$ jet mass distribution. The effect of the Higgs boson with a SM strength is smaller than the expected uncertainty on the $Z'$ limit and is therefore neglected. The first step to this procedure is to perform a fit with the full set of systematic uncertainties and obtain the best fit values for the nuisance parameters that are utilized in the analysis. Given these postfit shapes, the BumpHunter algorithm looks for significant deviations from the background model. The largest deviation
Figure 2: Postfit plot of the SM Higgs boson, $V + \text{jets}$, $t\bar{t}$ and QCD fit comparison to data. The middle panel shows the postfit and data distributions with the QCD and $t\bar{t}$ components subtracted. The lower panel shows the same distributions when also the $V + \text{jets}$ component is subtracted.
from the background found by the BumpHunter algorithm, as depicted in Figure 4, is found for a large-\( R \) mass value around 125 GeV, although with a small statistical significance. The corresponding global \( p \)-value is found to be 0.54.

This analysis also sets 95\% confidence level limits on signals from dark matter mediators that decay democratically to quarks, with masses between 100 and 200 GeV. These limits are shown in Figure 5, in terms of cross section times branching ratio, acceptance and efficiency (left), and in terms of the \( g_q \) parameter that controls the coupling of the DM mediator to quarks and thus determines the cross-section (right).

9 Conclusion

A search for boosted resonances decaying to two \( b \)-quarks and produced in association with an extra jet with high transverse momentum is performed. The analysis uses an integrated luminosity of 80.5 fb\(^{-1}\) of proton-proton collisions with a center-of-mass energy of \( \sqrt{s} = 13 \) TeV recorded by the ATLAS detector at the Large Hadron Collider. The search probes the mass range between 70 and 230 GeV. The observed signal strength for the \( V \) + jets process is \( \mu_V = 1.5 \pm 0.22 \) (stat.) \( ^{+0.29}_{-0.25} \) (syst.) \( \pm 0.18 \) (th.), corresponding to a significance of 5 standard deviations. For the SM Higgs boson process, the observed signal strength is \( \mu_H = 5.8 \pm 3.1 \) (stat.) \( \pm 1.9 \) (syst.) \( \pm 1.7 \) (th.), consistent with the background-only hypothesis at 1.6 \( \sigma \).

No evidence of a significant excess of events above the expected SM processes is found and limits on leptophobic \( Z' \) with democratic axial couplings to all quark generations (\( g_q \)) are set in the probed mass range.
Figure 4: The reconstructed mass distribution $m_J$ with the event reconstruction and selection as described in the text. The solid red line depicts the background prediction, consisting of the non-resonant dijet, $V +$ jets and $t\bar{t}$ processes. The vertical blue lines indicate the most discrepant interval identified by the BumpHunter algorithm. Without including systematic uncertainties, the probability that fluctuations of the background model would produce an excess at least as significant as the one observed in the data anywhere in the distribution, the BumpHunter probability, is 0.54. The low panel shows the bin-by-bin significances of the differences between the data and the fit, considering only statistical fluctuations.

Figure 5: The 95% credibility-level upper limits obtained from the invariant mass distribution on (a) the cross-section times acceptance times branching ratio times efficiency for the $Z'$ model described in the text and on (b) the $g_q$ parameter that controls the decay width of the DM mediator into SM particles.
A Data/MC comparisons in the signal region

Comparisons between some key distributions in the data and simulated background samples in the SR, for illustration only as the non-resonant dijet MC is not used for background estimation in the analysis, are shown in Figure 6. The considered backgrounds are the non-resonant dijet, Z + jets, W + jets and t¯t processes. Good modelling of the dijet background is not is expected. The normalization is corrected by fitting, in the range 70 GeV < m_J < 230 GeV, the signal candidate large-R jet mass distribution with an overall normalisation scale factor which is found to be 0.903 for SR. This factor is applied in the distributions in Figure 6. Uncertainties in the ratio plots are only statistical. The gray band is the uncertainty from Monte Carlo statistics.

Figure 6: The data vs MC comparison in the SR (a) of the number of large-R jets with p_T > 250 GeV in the event, (b) of the leading signal candidate jet p_T and (c) large-R jet mass, and (d) the number of track-jets associated to the signal candidate are shown. All the distributions are extracted by looking at the SR. The simulated dijet contribution is scaled by a constant factor of 0.90 in all the distributions, as determined when fitting the candidate large-R jet mass distribution. Uncertainties in the ratio plots are only statistical. The gray band is the uncertainty from Monte Carlo statistics.
References


[14] ATLAS Collaboration, 
Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector, 

[15] CMS Collaboration, 
Observation of Higgs boson decay to bottom quarks, 

[16] CMS Collaboration, 
Inclusive search for a highly boosted Higgs Boson decaying to a bottom quark–antiquark pair, 

Boosted Higgs shapes, 

[18] C. Grojean, E. Salvioni, M. Schlaffer and A. Weiler, 
Very boosted Higgs in gluon fusion, 

[19] ATLAS Collaboration, 
The ATLAS Experiment at the CERN Large Hadron Collider, 

[20] ATLAS Collaboration, 
ATLAS Insertable B-Layer Technical Design Report, 
(2010), CERN-LHCC-2010-013, ATLAS-TDR-019, 

[21] ATLAS Collaboration, 
ATLAS Insertable B-Layer Technical Design Report Addendum, 
(2012), CERN-LHCC-2012-009. ATLAS-TDR-19-ADD-1, 
url: http://cds.cern.ch/record/1451888.

[22] ATLAS Collaboration, 
Performance of the ATLAS trigger system in 2015, 

[23] M. Cacciari, G. P. Salam and G. Soyez, 
The anti-$k_t$ jet clustering algorithm, 

[24] ATLAS Collaboration, 
Performance of pile-up mitigation techniques for jets in pp collisions at \(\sqrt{s} = 8\) TeV using the ATLAS detector, 

[25] T. Sjöstrand, S. Mrenna and P. Z. Skands, 
A brief introduction to PYTHIA 8.1, 

[26] R. D. Ball et al., 
Parton distributions with LHC data, 

[27] ATLAS Collaboration, 
ATLAS Pythia 8 tunes to 7 TeV data, 
ATL-PHYS-PUB-2014-021, 2014, 
url: https://cds.cern.ch/record/1966419.

Event generation with SHERPA 1.1, 

[29] H.-L. Lai et al., 

[30] ATLAS Collaboration, 
Measurements of the production cross section of a Z boson in association with jets in pp collisions at \(\sqrt{s} = 13\) TeV with the ATLAS detector, 

[31] M. Bahr et al., 
Herwig++ physics and manual, 
[32] J. Pumplin et al.,
New generation of parton distributions with uncertainties from global QCD analysis,

[33] P. Nason, A New method for combining NLO QCD with shower Monte Carlo algorithms,

[34] S. Frixione, P. Nason and C. Oleari,
Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,

shower Monte Carlo programs: the POWHEG BOX, JHEP 06 (2010) 043,
arXiv: 1002.2581 [hep-ph].


[37] R. D. Ball et al., Parton distributions for the LHC Run II, JHEP 04 (2015) 040,

[38] K. Hamilton, P. Nason and G. Zanderighi, MINLO: Multi-Scale Improved NLO,

[39] K. Hamilton, P. Nason and G. Zanderighi,
Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator,

[40] P. Nason and C. Oleari,
NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG,

[41] J. M. Campbell et al., NLO Higgs boson production plus one and two jets using the POWHEG BOX,

[42] ATLAS Collaboration, Measurement of the $Z/\gamma^*$ boson transverse momentum distribution in $pp$
collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, JHEP 09 (2014) 145,

[43] B. Mellado Garcia, P. Musella, M. Grazzini and R. Harlander,
url: https://cds.cern.ch/record/2150771.

[44] D. Abercrombie et al., Dark matter benchmark models for early LHC Run-2 searches: report of the
ATLAS/CMS dark matter forum, (2015), ed. by A. Boveia, C. Doglioni, S. Lowette, S. Malik and S. Mrenna,

The automated computation of tree-level and next-to-leading order differential cross sections, and


[64] S. S. Wilks, 


