Search for diboson resonances with boson-tagged jets in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration*

1. Introduction

A major goal of the physics programme at the Large Hadron Collider (LHC) is the search for new phenomena that may become visible in high-energy proton–proton ($pp$) collisions. One possible signature of such new phenomena is the production of a heavy resonance with the subsequent decay into a final state consisting of a pair of vector bosons ($WW$, $WZ$, $ZZ$). Many models of physics beyond the Standard Model (SM) predict such a signature. These include extensions to the SM scalar sector as in the two-Higgs-doublet model (2HDM) [1] that predict new spin-0 resonances, composite-Higgs models [2–4] and models motivated by Grand Unified Theories [5–7] that predict new $W'$ spin-1 resonances, and warped extra dimensions Randall–Sundrum (RS) models [8–10] that predict spin-2 Kaluza–Klein (KK) excitations of the graviton, $G_{KK}$. The heavy vector triplet (HVT) [11,12] phenomenological Lagrangian approach provides a more model-independent framework for interpretation of spin-1 diboson resonances.

The search presented here focuses on TeV-scale resonances that decay into pairs of high-momentum vector bosons which, in turn, decay hadronically. The decay products of each of those vector bosons are collimated due to the high Lorentz boost and are typically contained in a single jet with radius $R = 1.0$. While the use of hadronic decays of the vector bosons benefits from the largest branching ratio ($67\%$ for $W$ and $70\%$ for $Z$ bosons) amongst the possible final states, it suffers from a large background contamination from the production of multijet events. However, this contamination can be mitigated with jet substructure techniques that exploit the two-body nature of $V 	o qq$ decays (with $V = W$ or $Z$).

Previous searches for diboson resonances were carried out by the ATLAS and CMS collaborations with $pp$ collisions at $\sqrt{s} = 7$, 8 and 13 TeV. These include fully leptonic ($t\bar{t}e\bar{e}$, $t\bar{t}\mu\mu$) [13–16], semileptonic ($v\nuqq$, $\ell\nuqq$, $\ell\ellqq$) [17–19] and fully hadronic ($qqqq$) $VV$ [17,19] final states. By combining the results of searches in the $v\nuqq$, $\ell\nuqq$, $\ell\ellqq$ and $qqqq$ channels, the ATLAS Collaboration [17] set a lower bound of 2.60 TeV on the mass of a spin-1 resonance at the 95\% confidence level, in the context of the HVT model B with $gv = 3$ (described in Section 2). When interpreted in the context of the bulk RS model with a spin-2 KK graviton and $k/M_{PL} = 1$, this lower mass bound is 1.10 TeV. The results presented here benefit from an integrated luminosity of 36.7 fb$^{-1}$, which is an order of magnitude larger than was available for the previous search in the fully hadronic final state at $\sqrt{s} = 13$ TeV [17].

2. Signal models

The analysis results are interpreted in terms of different models that predict the production of heavy resonances with either spin 0, spin 1 or spin 2. In the case of the spin-0 interpretation, a heavy scalar is produced via gluon–gluon fusion with subsequent decay into a pair of vector bosons. For this empirical model, the width of the signal in the diboson mass distribution is assumed to be dominated by the experimental resolution. The width of a Gaussian distribution characterising the mass resolution after full event selection ranges from approximately 3\% to 2\% as the resonance mass
increases from 1.2 to 5.0 TeV. The spin-0 model is referred to as the heavy scalar model in the rest of this Letter.

In the HVT phenomenological Lagrangian model, a new heavy vector triplet \((W', Z')\) is introduced, with the new gauge bosons degenerate in mass (also denoted by \(V'\) in the following). The couplings between those bosons and SM particles are described in a general manner, thereby allowing a broad class of models to be encompassed by this approach. The new triplet field interacts with the Higgs field and thus with the longitudinally polarised \(W\) and \(Z\) bosons by virtue of the equivalence theorem [20–22].

The strength of the coupling to the Higgs field, and thus SM gauge bosons, is controlled by the parameter combination \(g_V c_H\), where \(c_H\) is a multiplicative constant used to parameterise potential deviations from the typical strength of triplet interactions to SM vector bosons, taken to be \(g_V\). Coupling of the triplet field to SM fermions is set by the expression \(g^2 c_F/g_V\), where \(g\) is the SM SU(2)\(_L\) gauge coupling and, for the coupling to the Higgs field, \(c_F\) is a multiplicative factor that modifies the typical coupling of the triplet field to fermions. The HVT model A with \(g_V = 1, c_H \sim -g^2/2g_V\) and \(c_F \sim 1\) [11] is used as a benchmark. In this model, the new triplet field couples weakly to SM particles and arises from an extension of the SM gauge group. Branching ratios for \(W' \rightarrow WZ\) and \(Z' \rightarrow WW\) are approximately 2.0% each. The intrinsic width \(\Gamma\) of the new bosons is approximately 2.5% of the mass, which results in observable mass peaks with a width dominated by the experimental resolution. In this model, the dominant decay modes are into fermion pairs and searches in the \(\ell\ell\) and \(\ell\nu\) final states [23, 24] provide the best sensitivity. The calculated production cross section times branching ratio \((\sigma \times B)\) values for \(W' \rightarrow WZ\) with \(W\) and \(Z\) bosons decaying hadronically are 8.3 and 0.75 fb for \(W'\) masses of 2 and 3 TeV, respectively. Corresponding values for \(Z' \rightarrow WW\) are 3.8 and 0.34 fb.

The HVT model B with \(g_V = 3\) and \(c_H \sim \sim 1.1\) [11] is used as another benchmark. This model describes scenarios in which strong dynamics give rise to the SM Higgs boson and naturally include a new heavy vector triplet field with electroweak quantum numbers. The constants \(c_H\) and \(c_F\) are approximately unity, and couplings to fermions are suppressed, giving rise to larger branching ratios (\(\sim 50\%\)) for either \(W' \rightarrow WZ\) or \(Z' \rightarrow WW\) decays than in model A. Resonance widths and experimental signatures are similar to those obtained for model A and the predicted \(\sigma \times B\) values for \(W' \rightarrow WZ\) with hadronic \(W\) and \(Z\) decays are 13 and 1.3 fb for \(W'\) masses of 2 and 3 TeV, respectively. Corresponding values for \(Z' \rightarrow WW\) are 6.0 and 0.55 fb.

The RS model with one warped extra dimension predicts the existence of spin-2 Kaluza–Klein excitations of the graviton, with the lowest mode being considered in this search. While the original RS model [8] (often referred to as RS1) is constructed with all SM fields confined to a four-dimensional brane (the “TeV brane”), the bulk RS model [8,9] employed here allows those fields to propagate in the extra-dimensional bulk between the TeV brane and the Planck brane. Although ruled out by precision electroweak and flavour measurements, the RS1 model is used as a benchmark model to interpret diphoton and dilepton resonance searches due to the sizeable \(G_{KK}\) couplings to light fermions in that model. In the bulk RS model, those couplings are suppressed and decays into final states involving heavy fermions, gauge bosons or Higgs bosons are favoured. The strength of the coupling depends on \(k/M_R\), where \(k\) corresponds to the curvature of the warped extra dimension, and the effective four-dimensional Planck scale \(M_R = 2.4 \times 10^{16}\) GeV. The cross section and intrinsic width scale as the square of \(k/M_R\). For the choice \(k/M_R = 1\) used in this search, the \(\sigma \times B\) values for \(G_{KK} \rightarrow WW\) with \(W\) decaying hadronically are 0.54 and 0.026 fb for \(G_{KK}\) masses of 2 and 3 TeV, respectively. Corresponding values for \(G_{KK} \rightarrow ZZ\) are 0.32 and 0.015 fb.

3. ATLAS detector

The ATLAS experiment [25,26] 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 detector for tracking surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadronic calorimeters, and a muon spectrometer. The inner detector covers the pseudorapidity range \(|\eta| < 2.5\). It consists of silicon pixel, silicon microstrip, and transition radiation tracking detectors. A new innermost pixel layer [26] inserted at a radius of 3.3 cm has been used since 2015. Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A hadronic (steel/scintillator-tille) calorimeter covers the central pseudorapidity range \(|\eta| < 1.7\). The end-cap and forward regions are instrumented with LAr calorimeters for both the EM and hadronic energy measurements up to \(|\eta| = 4.9\).

The muon spectrometer surrounds the calorimeters and features three large air-core toroidal superconducting magnet systems with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 Tm across most of the detector. The muon spectrometer includes a system of precision tracking chambers and fast detectors for triggering. A two-level trigger system [27] is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to reduce the accepted rate to at most 100 kHz. This is followed by a software-based trigger level that reduces the accepted event rate to 1 kHz on average.

4. Data and simulation

4.1. Data

The data for this analysis were collected during the LHC pp collision running at \(\sqrt{s} = 13\) TeV in 2015 and 2016. Events must pass a trigger-level requirement of having at least one large-radius jet with transverse energy \(E_T > 360\) GeV in 2015 and \(E_T > 420\) GeV in 2016, where the jet is reconstructed using the anti-\(k_T\) algorithm [28] with a radius parameter of 1.0. Those thresholds correspond to the lowest-\(E_T\), unresolved large-radius jet triggers for each of the two data-taking periods. After requiring that the data were collected during stable beam conditions and the detector components relevant to this analysis were functional, the integrated luminosity of the sample amounts to 3.2 fb\(^{-1}\) and 33.5 fb\(^{-1}\) of pp collisions in 2015 and 2016, respectively.

4.2. Simulation

The search presented here uses simulated Monte Carlo (MC) event samples to optimise the selection criteria, to estimate the
acceptance for different signal processes, and to validate the experimental procedure described below. However, it does not rely on MC event samples to estimate the background contribution from SM processes.

Signal events for the heavy scalar model [29] were produced at next-to-leading-order via the gluon–gluon fusion mechanism with Powheg-Box v1 [30,31] using the CT10 parton distribution function (PDF) set [32]. Events were interfaced with Pythia v8.186 [33] for parton showering and hadronisation using the CTEQ6L1 PDF set [34] and the A2NLO set of tuned parameters (later referred to as tune) [35]. The width of the heavy scalar is negligible compared to the experimental resolution.

In the case of the HVT and RS models, events were produced at leading order (LO) with the MadGraph5_aMC@NLO v2.2.2 [36] event generator using the NNPDF23LO PDF set [37]. To study the sensitivity of the spin–2 resonance search to production from quark–antiquark or gluon–gluon initial states as well as to different vector-boson polarisation states, events were generated with JHUGen v5.6.3 [38] and the NNPDF23LO PDF set. For these signal models, the event generator was interfaced with Pythia v8.186 for parton showering and hadronisation with the A14 tune [39]. The $G_{\text{KK}}$ samples are normalised according to calculations from Ref. [40]. In all signal samples, the $W$ and $Z$ bosons are longitudinally polarised.

Multijet background events were generated with Pythia v8.186 with the NNPDF23LO PDF set and the A14 tune. Samples of $W + j$ and $Z + j$ events were generated with Herwig++ v2.7.1 [41] using the CTEQ6L1 PDF set and the UEES5 tune [42].

For all MC samples, charm-hadron and bottom-hadron decays were handled by EvRGen v1.2.0 [43]. Minimum-bias events generated using Pythia 8 were added to the hard-scatter interaction in such a way as to reproduce the effects of additional pp interactions in each bunch crossing during data collection (pile-up). An average of 23 pile-up interactions are observed in the data in addition to the hard-scatter interaction. The detector response was simulated with GEANT 4 [44,45] and the events were processed with the same reconstruction software as for the data.

5. Event reconstruction and selection

5.1. Reconstruction

The selection of events relies on the identification and reconstruction of electrons, muons, jets, and missing transverse momentum. Although the analysis primarily relies on jets, other particle candidates are needed to reject events that are included in complementary searches for diboson resonances.

The trajectories of charged particles are reconstructed using measurements in the inner detector. Of the multiple $pp$ collision vertices reconstructed from the available tracks in a given event, a primary vertex is selected as the one with the largest $\sum p_T^2$, where the sum is over all tracks with transverse momentum $p_T > 0.4$ GeV that are associated with the vertex. Tracks that are consistent with the primary vertex may be identified as electron or muon candidates. Electron identification is based on matching tracks to energy clusters in the electromagnetic calorimeter and relying on the longitudinal and transverse shapes of the electromagnetic shower. Electron candidates are required to satisfy the “medium” identification criterion [46] and to pass the “loose” track-based isolation [46]. Muon identification relies on matching tracks in the inner detector to muon spectrometer tracks or track segments. Muon candidates must also satisfy the “medium” selection criterion [47] and the “loose” track isolation [47].

Large-radius jets (hereafter denoted large-$R$ jets) are reconstructed from locally calibrated clusters of energy deposits in calorimeter cells [48] with the anti-$k_T$ clustering algorithm using a radius parameter $R = 1.0$. Jets are trimmed [49] to minimise the impact of pile-up by reclustering the constituents of each jet with the $k_T$ algorithm [50] into smaller $R = 0.2$ subjets and removing those subjets with $p_T^{\text{subj}} / p_T^{\text{jet}} < 0.05$, where $p_T^{\text{subj}}$ and $p_T^{\text{jet}}$ are the transverse momenta of the subjet and original jet, respectively. The clustering and trimming algorithms use the FastJet package [51]. Calibration of the trimmed jet $p_T$ and mass is described in Ref. [52].

The large-$R$ jet mass is computed using measurements from the calorimeter and tracking systems [53] according to

$$m_j = w^{\text{cal}} m^{\text{cal}} + w^{\text{trk}} \frac{p_T^{\text{trk}} m^{\text{trk}}}{p_T^{\text{jet}}}$$

where $p_T^{\text{trk}}$ is the transverse momentum of the jet evaluated using only charged-particle tracks associated with the jet, $m^{\text{cal}}$ and $m^{\text{trk}}$ are the masses computed using calorimeter and tracker measurements, and $w^{\text{cal}}$ and $w^{\text{trk}}$ are weights inversely proportional to the square of the resolution of each of the corresponding mass terms. Ghost association [54] is performed to associate tracks to the jets before the trimming procedure is applied. In this method, tracks are added with an infinitesimally small momentum as additional constituents in the jet reconstruction. Tracks associated with the jets are required to have $p_T > 0.4$ GeV and satisfy a number of quality criteria based on the number of measurements in the silicon pixel and microstrip detectors; tracks must also be consistent with originating from the primary vertex [53]. Including information from the tracking system provides improved mass resolution, especially at high jet $p_T$, due to the relatively coarse angular resolution of the calorimeter.

The magnitude of the event’s missing transverse momentum ($E_T^{\text{miss}}$) is computed from the vectorial sum of calibrated electrons, muons, and jets in the event [55]. For this computation and the rejection of non-collision background discussed below, jets are reconstructed from topological clusters using the anti-$k_T$ algorithm with a radius parameter $R = 0.4$ and are required to satisfy $p_T > 20$ GeV and $|\eta| < 4.9$. Calibration of those jets is described in Ref. [56]. The $E_T^{\text{miss}}$ value is corrected using tracks associated with the primary vertex but not associated with electrons, muons or jets.

5.2. Selection

Events used in complementary searches for diboson resonances in different final states are removed, in anticipation of a future combination. Accordingly, events are rejected if they contain any electron or muon with $p_T > 25$ GeV and $|\eta| < 2.5$. Furthermore, events with $E_T^{\text{miss}} > 250$ GeV are rejected.

Events with jets that are likely to be due to non-collision sources, including calorimeter noise, beam halo and cosmic rays, are removed [57]. Events are required to contain at least two large-$R$ jets with $|\eta| < 2.0$ (to guarantee a good overlap with the tracking acceptance) and mass $m_j > 50$ GeV. The leading (highest $p_T$) large-$R$ jet must have $p_T > 450$ GeV and the subleading (second highest $p_T$) large-$R$ jet must have $p_T > 200$ GeV. The invariant mass of the dijet system formed by these two jets must be $m_{jj} > 1.1$ TeV to avoid inefficiencies due to the minimum jet-$p_T$ requirements and to guarantee that the trigger requirement is fully efficient. Only jets in this system are considered in the rest of this Letter. Events passing the above requirements are said to pass the event “preselection”.

Further kinematic requirements are imposed to suppress background from multijet production. The rapidity separation between the leading and subleading jets (identified with subscripts 1 and 2...
in the following) must be sufficiently small, $|\Delta y| = |y_1 - y_2| < 1.2$, which is particularly aimed at suppressing $t$-channel dijet production. The $p_T$ asymmetry between the two jets $A = (p_T1 - p_T2)/(p_T1 + p_T2)$ must be smaller than 0.15 to remove events where one jet is poorly reconstructed.

Jets must be consistent with originating from hadronic decays of $W$ or $Z$ bosons. Discrimination against background jets inside a mass window including the $W/Z$ mass is based on the variable $D_2$, which is defined as a ratio of two-point to three-point energy correlation functions that are based on the energies of and pairwise angular distances between the jet's constituents [58,59]. This variable is optimised with parameter $\beta = 1$ to distinguish between jets originating from a single parton and those coming from the two-body decay of a heavy particle. A detailed description of the optimisation can be found in Refs. [52,60]. The boson-tagging criteria—the jet-mass window size and maximum $D_2$ value—are simultaneously optimised to achieve the maximal background-jet rejection for a fixed $W$ or $Z$ signal-jet efficiency of 50%. The optimisation uses signal jets from simulated $W' \rightarrow W Z \rightarrow qqqq$ events and background jets from simulated multijet events, and depends on the jet $p_T$ to account for varying resolution as a function of jet $p_T$. The size of the $W$ ($Z$) mass window varies from 22 (28) GeV near $p_T = 600$ GeV to 40 (40) GeV at $p_T \geq 2500$ GeV and the maximum $D_2$ value varies from 1.0 to 2.0 as the jet $p_T$ increases. An event is tagged as a candidate $WW$ event if both jets are within the $W$ ($Z$) mass window. It can also be tagged as a candidate $WZ$ event if the lower- and higher-mass jets are within the $W$ and $Z$ mass windows, respectively. Because the mass windows are relatively wide and overlap, jets may pass both $W$- and $Z$-tagging requirements.

To specifically suppress gluon-initiated jets, the number of tracks associated with each jet must satisfy $n_{\text{track}} < 30$. The tracks used must have $p_T > 0.5$ GeV and $|\eta| < 2.5$, as well as originate from the primary vertex.

The above set of selection criteria constitutes the signal region (SR) definition. Fig. 1 illustrates the kinematic acceptance times selection efficiency ($A \times \varepsilon$) at different selection stages for simulated heavy scalar resonances, heavy gauge bosons and KK gravitons decaying to the $WW$ final state. Similar $A \times \varepsilon$ values are obtained in the $WZ$ final state for the HVT model and in the $ZZ$ final state for the heavy scalar and bulk RS models. Multijet background events are suppressed with a rejection factor of approximately $2 \times 10^7$, as determined from simulation. The figure shows that, among the different selection criteria described above, the boson tagging reduces the signal $A \times \varepsilon$ the most. However, this particular selection stage provides the most significant suppression of the dominant multijet background.

Table 1 summarises the $A \times \varepsilon$ values for a number of models at resonance mass values of 2 and 3 TeV for the $WW$ final state; similar results are obtained for the other diboson final states. In
the case of the bulk RS model, the KK gravitons are mostly produced via gluon-induced processes and decay into longitudinally polarised W bosons. The polarisation affects the angular separation and momentum sharing between the decay products in the \( W \rightarrow q \bar{q} \) decay and thus affects the boson-tagging efficiency. To test the impact of the polarisation, the \( A \times \varepsilon \) values are evaluated with dedicated signal MC samples initiated by gluons or quarks, and with W bosons either fully longitudinally polarised or transversely polarised. Significant differences in the signal \( A \times \varepsilon \) are observed, as can be seen in Table 1, and these may need to be taken into account in reinterpetations of the results presented in this Letter. Little dependence is observed on the resonance mass. Differences in \( A \times \varepsilon \) for gluon- and quark-initiated production arise primarily from differences in the acceptance for selection on the jet \( |y| \) of the two leading jets and their rapidity separation. The boson-tagging efficiency for transversely polarised W bosons is approximately half that for longitudinally polarised W bosons and does not depend appreciably on the heavy-resonance production mechanism. In the case of quark-initiated production, \( A \times \varepsilon \) is similar for longitudinally and transversely polarised W bosons, as the reduction in kinematic acceptance is approximately com-

Table 1

<table>
<thead>
<tr>
<th>Model/process</th>
<th>Acceptance (\times) efficiency</th>
<th>(m = 2) TeV</th>
<th>(m = 3) TeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy scalar</td>
<td>7.3%</td>
<td>7.2%</td>
<td></td>
</tr>
<tr>
<td>HVT model A, (g_V = 1)</td>
<td>13.8%</td>
<td>13.9%</td>
<td></td>
</tr>
<tr>
<td>Bulk RS, (k_B/M_R = 1)</td>
<td>12.7%</td>
<td>13.6%</td>
<td></td>
</tr>
<tr>
<td>(gg \rightarrow G_{KK} \rightarrow WW) (longitudinally polarised W)</td>
<td>12.3%</td>
<td>13.4%</td>
<td></td>
</tr>
<tr>
<td>(gg \rightarrow G_{KK} \rightarrow WW) (transversally polarised W)</td>
<td>1.8%</td>
<td>1.9%</td>
<td></td>
</tr>
<tr>
<td>(q\bar{q} \rightarrow G_{KK} \rightarrow WW) (longitudinally polarised W)</td>
<td>5.4%</td>
<td>5.4%</td>
<td></td>
</tr>
<tr>
<td>(q\bar{q} \rightarrow G_{KK} \rightarrow WW) (transversally polarised W)</td>
<td>5.2%</td>
<td>5.8%</td>
<td></td>
</tr>
</tbody>
</table>

densified by an increase in boson-tagging efficiency. In the case of gluon-initiated production, both kinematic acceptance and boson-tagging efficiency favour longitudinally polarised W bosons.

5.3. Validation

In addition to the nominal SR, several validation regions (VRs) are defined to check the analysis procedure and estimate some of the sources of systematic uncertainty.

The definitions of the signal and validation regions are summarised in Table 2. A check of the statistical approach described in Section 6 is performed in the three different sidband validation regions. These correspond to the same selection as for the signal region except for requiring the jet mass to be in one of two sidebands. Both jet masses must be below the W boson mass with 50 < \( m_1 < 60\)–72 GeV (low–low sideband), or above the Z boson mass with 106–110 < \( m_1 < 140\) GeV (high–high sideband), or with one jet mass belonging to the low-mass range and the other to the high-mass range (low–high sideband). These mass ranges are chosen to have no overlap with the \( p_T\)-dependent W and Z mass windows applied to define the signal regions. The \( p_T\)-dependent mass windows imply a range of 60–72 GeV for the upper edge of the lower sideband and 106–110 GeV for the lower edge of the higher sideband.

A \( V + \) jets validation region is defined primarily to compare the observed and simulated \( V + \) jets event yields as a function of the number of tracks associated with the large-\( R \) jets and thereby derive an uncertainty in the efficiency for the \( n_{\text{trk}} \) requirement. There is no attempt at using this validation region to constrain the \( V + \) jets contribution to the signal regions as the total background there is estimated from an empirical fit to the dijet mass distribution. The \( V + \) jets validation region requires the presence of at least two large-\( R \) jets with \( |y| < 2.0 \). The leading jet must satisfy \( p_T > 600\) GeV and the subleading jet \( p_T > 200\) GeV. A higher minimum \( p_T \) requirement is imposed on the leading jet than in the nominal event selection to obtain a sample with higher average leading jet \( p_T \) that better corresponds to the jet \( p_T \) values probed.

Table 2

<table>
<thead>
<tr>
<th>Signal region</th>
<th>Veto non-qqqq channels:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low–low sideband validation region</td>
<td>( m_1 = 50)–60 GeV and ( m_2 = 60)–72 GeV</td>
</tr>
<tr>
<td>High–high sideband validation region</td>
<td>( m_1 = 106)–110 GeV and ( m_2 = 140) GeV</td>
</tr>
<tr>
<td>Low–high sideband validation region</td>
<td>( m_1 = 50)–60 GeV and ( m_2 = 140) GeV, or ( m_1 = 106)–110 GeV and ( m_2 = 60)–72 GeV</td>
</tr>
<tr>
<td>( V + ) jets validation region</td>
<td>( p_T &gt; 600) GeV and ( p_T &gt; 200) GeV</td>
</tr>
</tbody>
</table>

For low-mass validation regions, a \( |y| > 1.2 \) requirement is applied to the leading jet.
in the search. Finally, the leading jet must pass the boson-tagging requirements based on the $D_2$ variable only (i.e. the jet mass is not included in the tagging); no boson tagging is applied to the subleading jet. The resulting event sample in this validation region is approximately an order of magnitude larger than the samples selected in the different signal regions. Fig. 2 shows the leading jet mass distribution in the range $50 < m_{l} < 150$ GeV for events in this $V +$ jets validation region for $n_{trk} < 30$ and $n_{trk} \geq 30$. A clear contribution of $W/Z$ events is visible for $n_{trk} < 30$ but it is much less apparent for $n_{trk} \geq 30$, supporting the use of an upper limit on the number of tracks in the signal region.

To establish the efficiency in data of the $n_{trk} < 30$ selection, the leading-jet mass distribution is analysed in eight multiplicity subsamples, covering $0 \leq n_{trk} \leq 39$ in groups of five tracks each. Events originating from $W +$ jets and $Z +$ jets processes are modelled using a double-Gaussian distribution with the shape parameters determined from simulation, while background events not originating from $V +$ jets processes are fit to data independently in each subsample using a fourth-order polynomial (denoted “Fit bkd.” in Fig. 2). The relative normalisation in each $n_{trk}$ bin is controlled by a function which has a scaling parameter, allowing a variation in the track efficiency. The relative $W$ and $Z$ boson event contributions are fixed to the prediction from the simulation but the total $W + Z$ event normalisation is determined in the fit. A small upward shift in the $W/Z$ boson peak position is observed as $n_{trk}$ increases, which is well modelled by the simulation. An overall data-to-simulation scale factor of $1.03 \pm 0.05$ is extracted for the $n_{trk}$ requirement per $V$ jet. As this factor is consistent with unity, no correction is applied.

6. Background parameterisation

The search for diboson resonances is performed by looking for narrow peaks above the smoothly falling $m_{jj}$ distribution expected in the SM. This smoothly falling background mostly consists of SM multijet events. Other SM processes, including diboson, $W/Z +$ jets and $t\bar{t}$ production, amount to about 15% of the total background. They are also expected to have smoothly falling invariant mass distributions, although not necessarily with the same slope. The background in this search is estimated empirically from a binned maximum-likelihood fit to the observed $m_{jj}$ spectrum in the signal region. The following parametric form is used:

$$
\frac{dn}{dx} = p_1 (1 - x)^{p_2} e^{-p_3 x^{p_3}}
$$

where $n$ is the number of events, $x = m_{jj}/\sqrt{s}$, $p_1$ is a normalisation factor, $p_2$ and $p_3$ are dimensionless shape parameters, and $\xi$ is a constant chosen to remove the correlation between $p_2$ and $p_3$ in the fit. The latter is determined by repeating the fit with different $\xi$ values. The observed $m_{jj}$ distribution in data is histogrammed with a constant bin size of 100 GeV and the parametric form above is fit in the range $1.1 < m_{jj} < 6.0$ TeV. Only $p_2$ and $p_3$ are allowed to vary in the fit since $p_1$ is fixed by the requirement that the integral of $dn/dx$ equals the number of events in the distribution. This function has been successfully used in previous iterations of this analysis [17]. Other functional forms were tested and no significant improvement in the fit quality was observed.

The ability of the parametric shape in Eq. (1) to model the expected background distribution is tested in the three background-enriched sideband validation regions defined in Table 2. The results of the fits to data are shown in Fig. 3 along with the $\chi^2$ per degree of freedom (DOF). Bins with fewer than five events are grouped with bins that contain at least five events to compute the number of degrees of freedom. The fit model is found to provide a good description of the data in all of the VRs.

A profile likelihood test following Wilks’ theorem [62] is used to determine if including an additional parameter in the background model is necessary. Using the simulated multijet background with the sample size expected for the 2015–2016 dataset, as well as large sets of pseudo-experiments, Eq. (1) is found to be sufficient to describe the data. Possible additional uncertainties due to the choice of background model are assessed by performing signal-plus-background fits (also called spurious-signal tests) to the data in the sideband validation regions, where a signal contribution is expected to be negligible. The background is modelled with Eq. (1) and the signal is modelled using resonance mass distributions from simulation. The signal magnitude obtained in these background-dominated regions is less than 25% of its statistical uncertainty at any of the resonance masses considered in this search. Therefore, no additional uncertainty is assigned.

7. Systematic uncertainties

Systematic uncertainties in the signal yield and $m_{jj}$ distribution are assessed, and expressed as additional nuisance parameters in the statistical analysis, as described in Section 8.2. The dominant sources of uncertainty in the signal modelling arise from uncertainties in the large-$R$ jet energy and mass calibrations, affecting the jet $p_T$, mass and $D_2$ values. The correlations between the uncertainties in these jet variables are investigated by calculating the resulting uncertainties in the yield at a variety of signal
mass points for three different configurations: “strong”, with all three variables fully correlated; “medium”, with $p_T$ and $m_T$ correlated, whilst the $D_z$ is uncorrelated; and “weak”, with all three variables fully uncorrelated. The “medium” configuration is chosen as it results in the most conservative (largest) uncertainty in the yield.

Uncertainties in the modelling of the jet energy scale (JES), jet mass scale (JMS) and $D_z$ scale are evaluated using track-to-calorimeter double ratios between data and MC simulation [63]. This method introduces additional uncertainties from tracking. Uncertainties associated with track reconstruction efficiency, impact parameter resolution, tracking in dense environments, rate for fake tracks and sagitta biases are included. The size of the total correlated JES (JMS) uncertainty varies with jet $p_T$ and is approximately 3% (5%) per jet for the full signal mass range. The uncorrelated scale uncertainty in $D_z$ also varies with jet $p_T$ and is approximately 3% per jet for the full signal mass range.

Uncertainties in the modelling of jet energy resolution (JER), jet mass resolution (JMR) and $D_z$ resolution are assessed by applying additional smearing of the jet observables according to the uncertainty in their resolution measurements [52,63]. For the JER a 2% absolute uncertainty is applied per jet, and to mass and $D_z$ relative uncertainties of 20% and 15% are applied per jet, respectively. The response of the $D_z$ requirement is not strictly Gaussian and therefore the RMS of the observed distribution is taken as an approximation of the nominal width. There are sufficient dijet data to derive jet-related uncertainties up to $p_T$ values of 3 TeV [64].

The efficiency of the $n_{trk} < 30$ requirement in data and MC simulation is evaluated in the $V +$ jets VR defined in Section 5.3. The $n_{trk}$ efficiency scale factor is predominantly extracted using jets with $p_T \approx 650$ GeV, whereas signal jets in the analysis extend to $p_T \geq 1$ TeV. Examining the distribution of the number of tracks associated with jets as a function of jet $p_T$ reveals similar increasing trends in data and MC simulation. However, the average track multiplicity in the simulation is 3% larger at high $p_T$. Combining the 5% track multiplicity scale uncertainty with the $n_{trk}$ modelling uncertainty leads to a total 6% uncertainty per tagged jet in the efficiency of the $n_{trk}$ requirement. The uncertainty from the trigger selection is found to be negligible, as the minimum requirement on the dijet invariant mass of 1.1 TeV guarantees that the trigger is fully efficient.

Uncertainties affecting the signal prediction are as follows. The uncertainty in the combined 2015 + 2016 integrated luminosity is 3.2%. It is derived, following a methodology similar to that detailed in Ref. [65], from a calibration of the luminosity scale using $x$–$y$ beam-separation scans performed in August 2015 and May 2016. Theoretical uncertainties in the signal prediction are accounted for via their impact on the signal acceptance. The uncertainty associated with PDFs at high $Q^2$ values is modelled by taking the envelope formed by the largest deviations produced by the error sets of three PDF sets, as set out by the PDF4LHC group [66]. For the HVT model, the uncertainty ranges from 0.5% to 6% depending on the mass being tested, while a constant 0.5% uncertainty is determined in the case of the heavy scalar and bulk RS models. Uncertainties arising from the choice of A14 tuning parameters are covered by producing samples with variations of the tuning parameters describing initial-state radiation, final-state radiation, and multi-parton interactions. The uncertainty in the signal acceptance is then evaluated at MC generator level, before boson tagging or $n_{trk}$ cuts, resulting in a constant uncertainty of 3% for the HVT model and 5% for the heavy scalar and bulk RS models.
The ATLAS Collaboration / Physics Letters B 777 (2018) 91–113

Fig. 4. Dijet mass distributions for data in the (a) WW, (b) WZ, and (c) ZZ signal regions, as well as in the combined (d) WW + WZ and (e) WW + ZZ signal regions. The solid lines correspond to the result of the fit and the shaded bands represent the uncertainty in the background expectation. The lower panels show the significance of the observed event yield relative to the background fits. Expected signals are shown for the HVT model B with $g_V = 3$ and the bulk RS model with $k/M_P = 1$. The predictions for $G_{KK}$ production are multiplied by a factor of 10. The lower panels show the significance of the observed event yield relative to the background fits taking their uncertainties into account as described in Ref. [61].

8. Results

8.1. Background fit

The fitting procedure described in Section 6 is applied to the data passing the $WW$, $WZ$ and $ZZ$ selections described in Section 5.2, and resulting dijet mass distributions are shown in Fig. 4. The mass spectra obtained in combined $WW + WZ$ and $WW + ZZ$ SRs are also shown. A total of 497, 904, 618, 980, and 904 events are found in the $WW$, $WZ$, $ZZ$, $WW + WZ$, and $WW + ZZ$ SRs. Approximately 20% of events are included in all three regions: $WW$, $WZ$ and $ZZ$. The requirements of the $WW$ ($ZZ$) SR are satisfied by 47% (57%) of the events in the $WZ$ SR. The fitted background functions shown, labelled “Fit”, are evaluated in bins between 1.1 and 6.0 TeV. No events are observed beyond 3.1 TeV.

The dijet mass distributions in all signal regions are described well by the background model over the whole range explored.

As a test of the background model, the fit is also performed on dijet mass distributions obtained with no boson tagging applied but with weights corresponding to the probability for each jet to satisfy the boson tagging requirements. This probability is derived from the data as a function of the jet $p_T$ and the resulting fits are consistent with the nominal background fits within uncertainties. The use of untagged data allows to validate the model with a sufficiently large number of data events up to dijet masses of 6 TeV.

8.2. Statistical analysis

The final results are interpreted using a frequentist statistical analysis. The parameter of interest is taken to be the signal
strength, $\mu$, defined as a scale factor to the number of signal events predicted by the new-physics model being tested. A test statistic $\lambda(\mu)$, based on a profile likelihood ratio [67] is used to extract information about $\mu$ from a maximum-likelihood fit of the signal-plus-background model to the data. The likelihood model is defined as

$$L = \prod_i P_{\text{pois}}(n_{i,\text{obs}}^i | n_{i,\text{exp}}^i) \times G(\alpha) \times \mathcal{N}(\theta),$$

where $P_{\text{pois}}(n_{i,\text{obs}}^i | n_{i,\text{exp}}^i)$ is the Poisson probability to observe $n_{i,\text{obs}}^i$ events in dijet mass bin $i$ if $n_{i,\text{exp}}^i$ events are expected, $G(\alpha)$ are a series of Gaussian probability density functions modelling the systematic uncertainties, $\alpha$, related to the shape of the signal, and $\mathcal{N}(\theta)$ is a log-normal distribution for the nuisance parameters, $\theta$, which model the systematic uncertainty in the signal normalisation. The expected number of events is the bin-wise sum of those expected for the signal and background: $n_{\text{exp}} = n_{\text{sig}} + n_{\text{bg}}$. The expected number of background events in bin $i$, $n_{\text{exp}}^i$, is obtained by integrating $dn/dx$ obtained from Eq. (1) over that bin. Thus, $n_{\text{exp}}$ is a function of the background parameters $p_1$, $p_2$, and $p_3$. The number of expected signal events, $n_{\text{sig}}$, is evaluated based on MC simulation assuming the cross section of the model under test multiplied by the signal strength $\mu$.

The significance of any deviation observed in the data with respect to the background-only expectation is quantified in terms of the local $p_0$ value. This is defined as the probability of fluctuations of the background-only expectation to produce an excess at least as large as the one observed. The largest deviation from the background model occurs in the $ZZ$ SR for a heavy scalar with mass of 2.4 TeV. The local significance of this deviation is 2.0 $\sigma$ and the corresponding global significance is less than 1 $\sigma$. No statistically significant excess is observed and upper exclusion limits are placed on the cross section times branching ratio for the production of heavy resonances decaying into diboson final states. A correction to account for the branching ratio of $V$ decays into hadronic final states is applied in the results below. The limits are set with the CLs method [68] using large sets of pseudo-experiments.

Limits on $\sigma \times B$ are set in each combined diboson channel as a function of the resonance mass. The HVT models A and B with degenerate $W’$ and $Z’$ are used as benchmarks for the combined $WW + WZ$ signal region, and the bulk RS or heavy scalar models are used for the $WW + ZZ$ signal region. Fig. 5(a) shows the observed limits on the production of a spin-1 vector triplet as a funct-
tion of resonance mass in the WW + ZZ signal region. A spin-1 vector triplet with couplings predicted by the HVT model A (B) with \( g_Y = 1 \) (\( g_Y = 3 \)) is excluded in the range 1.2 < \( m(V') < 3.1 \) (1.2 < \( m(V') < 3.5 \)) TeV, at the 95% confidence level (CL). Fig. 5(b) shows the observed limits on the production of a GKK as a function of \( m(GKK) \) in the WW + ZZ signal region. Production of a GKK in the bulk RS model with \( k/M_{Pl} = 1 \) is excluded in the range 1.3 < \( m(GKK) < 1.6 \) TeV, at the 95% CL. Fig. 5(c) shows the observed limits on the production of a heavy scalar as a function of \( m(\text{Scalar}) \) in the WW + ZZ signal region. Table 3 presents the resonance mass ranges excluded at the 95% CL in the various signal regions and model signals considered in the search. In the search for heavy scalar particles, upper limits are set on \( \sigma \times B \) at the 95% CL with values of 9.7 fb at \( m(\text{Scalar}) = 2 \) TeV and 3.5 fb at \( m(\text{Scalar}) = 3 \) TeV.

9. Conclusions

This Letter reports a search for massive resonances decaying via WW, WZ and ZZ into hadrons with 36.7 fb\(^{-1}\) of \( \sqrt{s} = 13 \) TeV pp collisions collected at the LHC with the ATLAS detector in 2015–2016. The search takes advantage of the high branching ratio of hadronic decays of the vector bosons and covers the resonance mass range between 1.2 and 5.0 TeV. In this kinematic range, the vector bosons are highly boosted and are reconstructed as single large-radius jets that are tagged by exploiting their two-body substructure. The invariant mass distribution of the two highest-\( p_T \) large-radius jets in each event is used to search for narrow resonance peaks over a smoothly falling background. No significant excess of data is observed and limits are set on the cross section times branching ratio for diboson resonances at the 95% confidence level. In the case of the phenomenological HVT model A (model B) with \( g_Y = 1 \) (\( g_Y = 3 \)), a spin-1 vector triplet is excluded for masses between 1.2 and 3.1 TeV (1.2 and 3.5 TeV). For the bulk RS model with \( k/M_{Pl} = 1 \), a spin-2 Kaluza–Klein graviton is excluded in the range between 1.3 and 1.6 TeV. Upper limits on the production cross section times branching ratio for new heavy scalar particles are set with values of 9.7 fb and 3.5 fb at scalar masses of 2 TeV and 3 TeV, respectively.

Acknowledgements

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; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMTR CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DRF/IRFU, France; SRNSF, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE 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, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, 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, the Canada Council, Canarie, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, ERDF, FPF, Horizon 2020 and Marie Sklodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BSF, GIF and Minerva, Israel; BRF, Norway; CERCA Programme Generalitat de Catalunya, Generalitat Valenciana, Spain; 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. [69].

References
