Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS Collaboration

Abstract

A search for Higgs-boson pair production in the $b\bar{b}b\bar{b}$ final state is carried out with 13.3 fb$^{-1}$ of LHC proton-proton collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector in 2015 and 2016. The data are consistent with the estimated background and are used to set upper limits on the production cross section of Higgs-boson pairs times branching ratio to $b\bar{b}b\bar{b}$ for both non-resonant and resonant production. In the context of the Randall–Sundrum model, upper limits in the 1000 fb to 2 fb range are obtained for masses between 300 and 3000 GeV, at the 95% confidence level. The cross section times branching ratio for non-resonant Standard Model Higgs-boson pair production is constrained to be less than 330 fb, 29 times the Standard Model predicted cross section of $11.3^{+0.9}_{-1.0}$ fb.

Note: the current version of this note was updated on 17th March 2017 to add the non-resonant expected limit, which was missing due to an oversight.
1 Introduction

The discovery of a Higgs boson ($h$) [1, 2] at the Large Hadron Collider (LHC) motivates searches for physics beyond the Standard Model (SM) in channels involving coupling to the Higgs boson. In particular, many models of new physics predict cross sections for Higgs boson pair production that are significantly greater than the SM prediction. Resonant Higgs boson pair production is predicted in models such as the bulk Randall-Sundrum (RS) model [3, 4], which features spin-2 Kaluza-Klein gravitons, $G_{KK}^*$, that are produced via gluon-fusion and then may subsequently decay to a pair of Higgs bosons. Similarly, 2HDM models [5] posit the existence of a heavy spin-0 scalar, $H$, that may decay to $h$ pairs. Enhanced, non-resonant Higgs boson pair production is predicted by other models, for example those featuring light coloured scalars [6] or direct $t\bar{t}hh$ vertices [7, 8].

The analysis presented in this paper makes use of the dominant $h \to b\bar{b}$ decay mode to search for Higgs boson pair production using the $b\bar{b}b\bar{b}$ final state. It is based on the previous ATLAS searches at $\sqrt{s} = 8$ TeV [9] and $\sqrt{s} = 13$ TeV [10], which set limits on both resonant and non-resonant Higgs boson pair production. For the Bulk RS model, these previous analyses excluded $G_{KK}^*$ masses between 480 GeV up to $770–990$ GeV, depending on the model parameters. For non-resonant production, the $\sqrt{s} = 13$ TeV analysis [10] set a cross section upper limit of 1.22 pb [11]. CMS has also searched for narrow resonances in the $b\bar{b}b\bar{b}$ channel using $\sqrt{s} = 8$ TeV data [12]. Other final states have been analysed for evidence of Higgs boson pair production using $\sqrt{s} = 8$ TeV data: ATLAS has searched in $b\bar{b}\gamma\gamma$ [13], $b\bar{b}t^+\tau^-$ and $W^+W^-\gamma\gamma$, which were combined with the $b\bar{b}b\bar{b}$ results in [11]. CMS has performed analyses of the $b\bar{b}\gamma\gamma$ [14] and $b\bar{b}t^+\tau^-$ [15] final states in addition to a search in multilepton and multilepton+photons final states [16]. None of these analyses found any evidence of Higgs boson pair production.

Two techniques to reconstruct the Higgs boson, which are complementary in their acceptance, are presented in this note. Both use anti-$k_t$ jets [17] to reconstruct the kinematics of the Higgs boson candidates, but with different values of the jet radius parameter, $R$. The "resolved" technique is used to analyse $hh$ systems that have sufficiently low mass ($m_{hh}$) to enable the reconstruction of each $b$-quark as a distinct $b$-jet with $R = 0.4$. These jets are identified as containing $b$-hadrons using a multivariate $b$-tagging algorithm [18, 19]. The "boosted" technique is used to analyse those $hh$ systems with higher mass, in which the higher Lorentz boosts cause an inability to resolve the decay products of each Higgs boson as separate $R = 0.4$ $b$-jets. In these cases, the Higgs boson candidates consist of a single anti-$k_t$ jet with $R = 1.0$, with $b$-tagging performed using the same multivariate algorithm as in the resolved analysis, but operating on $R = 0.2$ track-jets matched to the $R = 1.0$ jet. The final results are obtained using the resolved analysis up to a resonance mass of 1000 GeV, with the boosted analysis used for higher masses.

The search reported in this note has been improved relative to the previous analysis [10] and yields better sensitivity to non-resonant Higgs boson pairs, resonances with masses below 500 GeV, and resonances with masses above 2500 GeV. The improvement for non-resonant and low-mass resonant signals has been gained by loosening the topological selection requirements in order to increase signal acceptance, while the improvement at high mass was achieved by removing the requirement that at least two $R = 0.2$ track-jets are associated with the boosted Higgs candidates, again increasing signal acceptance. The overall analysis sensitivity also benefits from a factor of four increase in integrated luminosity by including the dataset recorded by ATLAS during March-July 2016.
2 ATLAS detector

The ATLAS experiment [20] 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 (ID) surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic (EM) and hadronic calorimeters, and a muon spectrometer (MS). The ID covers the pseudorapidity range $|\eta| < 2.5$. It consists of silicon pixel, silicon micro-strip, and straw tube transition radiation tracking detectors. An additional pixel detector layer [21], inserted at a mean radius of 3.3 cm, is used in the Run-2 data-taking. Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic energy measurements. A steel/scintillator-tile hadronic calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The end-cap and forward regions are instrumented with LAr calorimeters for both EM and hadronic energy measurements up to $|\eta| = 4.9$. The muon spectrometer (MS) surrounds the calorimeters and includes three large superconducting air-core toroids. The field integral of the toroids ranges between 2.0 and 6.0 Tm for most of the detector. The MS includes a system of precision tracking chambers and fast detectors for triggering. A dedicated trigger system is used to select events [22]. The first-level trigger is implemented in hardware and uses the calorimeter and muon detectors to reduce the accepted event rate to 100 kHz. This is followed by a software-based high-level trigger (HLT) that reduces the accepted event rate to 1 kHz on average.

3 Data and simulated samples

3.1 Data

This analysis is performed on two LHC $pp$ collision datasets at $\sqrt{s} = 13$ TeV. The first dataset was collected during 2015. After requiring that the data were collected during stable beam conditions and that all relevant detector systems were functional, the integrated luminosity was measured to be 3.2 $fb^{-1}$. The uncertainty in the integrated luminosity is found to be $\pm 2.1\%$, from a calibration of the luminosity scale using $x - y$ beam-separation scans performed in August 2015, following a methodology similar to that detailed in Ref. [23, 24]. The second dataset was collected during March-July 2016 and, after the LHC beam and detector data-quality requirements, corresponds to an integrated luminosity of 10.1 $fb^{-1}$, with a preliminary uncertainty of $\pm 3.7\%$ found using $x - y$ beam-separation scans performed in May 2016.

Events were selected from these datasets for the resolved analysis using two sets of four triggers: one for the 2015 dataset and one for the 2016 dataset. All of the triggers required the presence of either one or two $b$-tagged jets, with different kinematic and $b$-tag quality requirements. Some of the triggers required additional jets. Multiple triggers were needed in order to maintain high signal efficiency: greater than 65% for signals with $m_{hh} = 300$ GeV, increasing to more than 95% for $m_{hh} \geq 500$ GeV. Different trigger sets were used in 2015 and 2016 due to the different LHC conditions and to benefit from improved online $b$-tagging. For the boosted analysis, events were selected from the 2015 dataset using a trigger that required a single anti-$k_t$ jet with $R = 1.0$ and with $p_T > 360$ GeV. In 2016, a similar trigger was used.
but with a higher threshold of $p_T > 420$ GeV. These triggers were 100% efficient for signals passing the offline large-$R$ jet kinematic requirements.

### 3.2 Signal models and simulation

Simulated Monte Carlo (MC) event samples are used in this analysis to model signal production and the background from $t\bar{t}$ and $Z$+jets events. The dominant multi-jet background is modelled using data-driven techniques.

Signal $G^*_KK \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events were generated at leading order (LO) with **MG5_aMC@NLO v2.2.2** [25] using the NNPDF2.3 LO parton distribution function (PDF) set [26]. The matrix elements produced were passed to **Pythia 8.186** [27] for parton showering, hadronization and simulation of the underlying event. The A14 set of tuned underlying-event parameters [28] was used. The signal samples were generated with $k/M_{Pl} = 1$ ($k$ is the curvature of the warped extra dimension and $M_{Pl} = 2.4 \times 10^{18}$ GeV is the effective four-dimensional Planck scale).

Non-resonant SM $pp \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ signal events were generated via the gluon-fusion process with the next-to-leading-order (NLO) **MG5_aMC@NLO** using form factors for the top-quark loop from HPAIR [29, 30]. The cross section times branching ratio to the $b\bar{b}b\bar{b}$ final state, evaluated at next-to-next-to-leading order (NNLO) with the summation of logarithms at next-to-next-leading-logarithm accuracy and including top-quark mass effects at NLO [31] is $11.3^{+0.9}_{-1.0}$ fb [32]. The uncertainty includes the effects due to renormalization and factorization scales, PDF set, $\alpha_S$, and the $h \rightarrow b\bar{b}$ branching ratio.

In all signal samples, the mass of the Higgs boson ($m_h$) was set to 125.0 GeV.

Generation of $t\bar{t}$ events was performed with **Powheg-box v1** using the CT10 PDF set. The parton shower, hadronization, and the underlying event were simulated using **Pythia 6.428** [33] with the CTEQ6L1 PDF set and the corresponding Perugia 2012 set of tuned underlying-event parameters [34]. The top-quark mass was set to 172.5 GeV. Higher-order corrections to $t\bar{t}$ cross sections were computed with **Top++ 2.0** [35]. These incorporate next-to-next-to-leading-order corrections in QCD, including resummation of next-to-next-to-leading logarithmic soft gluon terms. The $Z$+jets sample was generated using **Pythia 8.186** with the NNPDF2.3 LO PDF set.

For all simulated samples, charm-hadron and bottom-hadron decays were handled by **EvtGen 1.2.0** [36]. To simulate the impact of multiple $pp$ interactions that occur within the same or nearby bunch crossings (pile-up), minimum-bias events generated with Pythia 8 are overlaid on top of the hard scatter event, with the instantaneous luminosity profile and accelerator bunch structure present in 2015 and 2016 data. The detector response was simulated with **Geant 4** [37, 38] and the events were processed with the same reconstruction software as that used for the data.

### 4 Event reconstruction

Higgs boson kinematic properties are reconstructed using anti-$k_t$ jets built from topological clusters of energy deposits in calorimeters cells [39], using a four-momentum reconstruction scheme with massless clusters as input. The directions of jets are corrected to point back to the identified hard-scatter proton–proton collision vertex, the vertex with the highest $\Sigma p_T^2$ of constituent tracks.
The $R = 0.4$ jets used in the resolved analysis are reconstructed from clusters calibrated at the electromagnetic scale. They are corrected for additional energy deposited from the pile-up interactions using an area-based correction [40], with a further, small correction based on the expected average number of $pp$ collisions in the bunch-crossing that removes any residual energy offset. Jets are then calibrated using $p_T$- and $\eta$-dependent calibration factors derived from MC simulation, before global sequential calibration [41] is applied, which reduces flavour-dependent differences in calorimeter response. Final calibration is performed based on in situ measurements in real collision data [42]. Jets with a majority of their energy arising from these pile-up interactions are suppressed using tracking information as detailed in Ref. [43].

The $R = 1.0$ jets used in the boosted analysis are built from locally-calibrated [41] topological clusters. They are trimmed [44] to minimize the impact of energy deposits from pile-up interactions. Trimming proceeds by reclustering the jet with the $k_t$ algorithm [45] into smaller $R = 0.2$ subjets and then removing those subjets with $p_{T,\text{subjet}}/p_{T,\text{jet}} < 0.05$, where $p_{T,\text{subjet}}$ is the transverse momentum of the subjet and $p_{T,\text{jet}}$ that of the original jet. The energy and mass scales of the trimmed jets are then calibrated using $p_T$- and $\eta$-dependent calibration factors derived from MC simulations [46].

Jets containing $b$-hadrons are identified using a multivariate $b$-tagging algorithm [18]. This algorithm is applied to a set of charged-particle tracks associated to each jet that satisfy quality and impact parameter criteria. Tracks are associated with $R = 0.4$ jets using a $p_T$-dependent $\Delta R$ requirement. For $R = 1.0$ jets, $b$-tagging is performed on $R = 0.2$ track-jets [47] matched to the $R = 1.0$ jets using ghost association [48]. The smaller radius parameter of the track-jets enables two close-by $b$–hadrons to be identified within a single jet, which is beneficial when reconstructing high-$p_T$ Higgs boson candidates. The $b$-tagging requirements of the resolved (boosted) analysis result in an efficiency of 70% (77%) for jets containing $b$-hadrons in a sample of simulated $t\bar{t}$ events. The corresponding efficiencies for $c$-hadron jets and light-quark or gluon jets are 7.8% (24%) and 0.4% (1.7%), respectively. Different $b$-tagging operating points are chosen in the two analyses to maximise their respective sensitivities.

Muons are reconstructed by combining tracks in the ID and MS, and are required to satisfy medium muon identification criteria [49]. The four-momentum of a muon with $p_T > 4$ GeV and $|\eta| < 2.5$, that is within $\Delta R$ of 0.4 (0.2) of a jet used for $b$-tagging in the resolved (boosted) analysis, is added to the calorimeter-jet’s four-momentum to partially account for the energy lost in semi-leptonic $b$-hadron decays.

5 Resolved analysis

The resolved analysis is optimised to discover signals that result in either non-resonant or low-mass resonant Higgs boson pairs. The strategy is to select two Higgs boson candidates, each composed of two $b$-tagged anti-$k_t$, $R = 0.4$ jets, with masses near $m_h$.

The mass of the two Higgs boson candidate system ($m_{4j}$) is used as the final discriminant between Higgs boson pair production and the backgrounds. Resonant signals would lead to a localised excess, while non-resonant production would result in an excess in the tail of the $m_{4j}$ spectrum.

The 2015 and 2016 datasets are analysed separately because LHC conditions such as instantaneous luminosity were different in these two years, necessitating the use of different trigger schemes. The results are combined statistically as described in Section 7.
5.1 Selection

The combined acceptance times efficiency at different stages of the event selection is shown in Figure 1 for the Bulk RS and SM non-resonant signal models.

The selection begins with the requirement that the event contains at least four \( b \)-tagged anti-\( k_t \) jets with \( p_T > 30 \text{ GeV} \) and \( |\eta| < 2.5 \). The acceptance times efficiency of this requirement is shown in Figure 1 as the curve labelled “4 \( b \)-tagged jets”. The four jets with the highest \( b \)-tagging score are paired to construct two Higgs boson candidates. The angle between the decay products of the Higgs boson in the lab frame is dependent on the Lorentz boost of the Higgs boson and thus the invariant mass of the four-jet system, \( m_{4j} \). Accordingly, pairings of jets into Higgs boson candidates are only accepted if they satisfy the following requirements, where \( m_{4j} \) is expressed in GeV:

\[
\begin{align*}
\text{if } m_{4j} < 1250 \text{ GeV} & : \\
\frac{360}{m_{4j}} - 0.5 \leq & \Delta R_{j,j,\text{lead}} \leq \frac{655}{m_{4j}} + 0.475 \\
\frac{235}{m_{4j}} < & \Delta R_{j,j,\text{subl}} \leq \frac{875}{m_{4j}} + 0.35
\end{align*}
\]

\[
\begin{align*}
0 < & \Delta R_{j,j,\text{lead}} < 1 \\
0 < & \Delta R_{j,j,\text{subl}} < 1
\end{align*}
\]

In these expressions, \( \Delta R_{j,j,\text{lead}} \) is the angular distance between jets in the leading Higgs boson candidate and \( \Delta R_{j,j,\text{subl}} \) for the sub-leading candidate. The leading Higgs boson candidate is defined to be the
Figure 2: $D_{hh}$ is used to choose the best pairing of jets into Higgs boson candidates in the resolved analysis. It is the distance of the pairing’s \( (m_{2j}^{lead}, m_{2j}^{subl}) \) from the line connecting \((0 \text{ GeV}, 0 \text{ GeV})\) and the centre of the signal region, \((120 \text{ GeV}, 115 \text{ GeV})\).

candidate with the highest scalar sum of jet $p_T$. Figure 1 shows the efficiency after requiring events to contain two Higgs boson candidates satisfying these criteria as the curve labelled “$\Delta R_{jj}$”. Most of the efficiency loss shown is the rejection of events where one of the $b$-tagged jets did not come from a Higgs boson decay.

In events with $m_{4j} < 500$ GeV, there are likely to be multiple Higgs boson candidates satisfying these requirements, necessitating an algorithm to choose the correct pairs. In the absence of energy loss through semi-leptonic decays, the optimal choice would be the combination most consistent with the decays of two particles of equal mass. To account for energy loss, the choice is modified to the pairing that minimizes the distance, $D_{hh}$, from a line connecting \((0 \text{ GeV}, 0 \text{ GeV})\) and \((120 \text{ GeV}, 115 \text{ GeV})\):

$$D_{hh} = \sqrt{(m_{2j}^{lead})^2 + (m_{2j}^{subl})^2} \sin^{-1}\left(\frac{m_{2j}^{subl}}{m_{2j}^{lead}} \tan^{-1}\left(\frac{115}{120}\right)\right)$$

(1)

where $m_{2j}^{lead}$ is the mass of the leading Higgs boson candidate and $m_{2j}^{subl}$ the mass of the sub-leading candidate (see Figure 2). The values of 120 GeV and 115 GeV are the centre of the signal region in $m_{2j}^{lead}$ and $m_{2j}^{subl}$ respectively, chosen because they correspond to the median values of the narrowest intervals that contain 90% of the signal in simulations.

The resolved analysis searches for resonances with a wide range of masses, $300$ GeV $< m < 1200$ GeV, as well as non-resonant signals. Event selection criteria that vary as a function of the reconstructed resonance
mass are used to enhance the analysis sensitivity across this range. Mass-dependent requirements are made on the leading Higgs boson candidate $p_T$, $p_T^{lead}$, the sub-leading Higgs boson $p_T$, $p_T^{subl}$, and the pseudorapidity difference between the two Higgs boson candidates, $|\Delta \eta_{hh}|$:

\[
p_T^{lead} > 0.5m_{4j} - 90 \text{ GeV}
\]
\[
p_T^{subl} > 0.33m_{4j} - 70 \text{ GeV}
\]
\[
|\Delta \eta_{hh}| < \begin{cases} 
1.1 & \text{if } m_{4j} < 850 \text{ GeV} \\
2 \times 10^{-3}m_{4j} - 0.6 & \text{if } m_{4j} > 850 \text{ GeV} 
\end{cases}
\]

where $m_{4j}$ is again expressed in GeV.

Following this mass-dependent selection, a further requirement is made that the two Higgs boson candidates must be separated in $\Delta R$, demanding $\Delta R (h, h) > 1.5$. Figure 1 shows the efficiency of the mass-dependent selection and Higgs separation requirements as the curve labelled “$m_{4j}$ Dependent Cuts, $\Delta R_{hh}$”.

A requirement on the Higgs boson candidates’ masses is used to define the signal region:

\[
X_{hh} = \sqrt{\left(\frac{m_{2j}^{lead} - 120 \text{ GeV}}{0.1m_{2j}^{lead}}\right)^2 + \left(\frac{m_{2j}^{subl} - 115 \text{ GeV}}{0.1m_{2j}^{subl}}\right)^2} < 1.6
\]  

(2)

where the $0.1m_{2j}$ terms represent the widths of the leading and sub-leading Higgs boson candidate mass distributions. The signal region is shown as the inner region of Figure 3. The efficiency of the full event selection, including this signal region requirement, is shown in Figure 1 as “$X_{hh}$”.

Finally, the effect of the trigger requirement is seen in the curve labelled "trigger", which shows the total efficiency of the event selection criteria and the trigger.

The final analysis discriminant, $m_{4j}$, is simply the invariant mass of the selected four-jet system. No correction is made based on the known Higgs boson mass.

5.2 Background estimation

After the full event selection, about 95% of the background consists of multijet events, which are modelled using data. The remaining 5% is $t\bar{t}$, which is modelled using MC simulation. There is negligible background from all other sources including processes involving Higgs bosons.

5.2.1 Multijet background

The multijet background is modelled with an independent data sample selected using the same trigger and selection requirements as described above, except for the $b$-tagging requirement: exactly two jets must pass the $b$-tagging requirement. This “2-tag” selection yields a data sample that consists of 98% multijet events and 2% $t\bar{t}$ events. The predicted signal contamination is negligible.

The 2-tag sample used to model the full-selection “4-tag” sample is a signal-free sideband region of the $m_{2j}^{lead}$-$m_{2j}^{subl}$ plane, as illustrated in Figure 3. The modelling is verified in a control region. The sideband
region is defined as $48 \text{ GeV} < \sqrt{(m_{\text{lead}2j} - 120 \text{ GeV})^2 + (m_{\text{subl}2j} - 115 \text{ GeV})^2} < 88 \text{ GeV}$, while the control region is defined as the region in the $m_{\text{lead}2j} - m_{\text{subl}2j}$ plane between the signal and sideband regions. These definitions are chosen to be orthogonal to the signal region and to give approximately equal event yields in both the sideband and control regions. The outer limit on the sideband region ensures that the kinematic properties of the sideband region are representative of the signal region.

Events in the 2-tag data sample are reweighted to correct for differences introduced by the additional $b$-tagging requirement on the 4-tag sample. These differences can arise for the following reasons: events with more jets are more likely to satisfy the requirement of at least four $b$-tagged jets; the $b$-tagging efficiency varies as a function of jet $p_T$ and $\eta$; the various multijet processes contribute in different fractions in each sample; and the fractions of events passed by each trigger path changes. The weights are derived in the sideband region from linear fits to the ratio of the total background model to data for four distributions that are found to have the largest disagreement between 2-tag and 4-tag: the number of additional jets in the event, the leading Higgs boson candidate $p_T$, and the energies of the jets in the sub-leading Higgs boson candidate. The reweighting is performed using one-dimensional distributions but is iterated so that correlations between the four variables are taken into account.

After reweighting, the normalization of the multijet background prediction is set by scaling the number of events in each region of the 2-tag sample by the following factor $\mu_{\text{multijet}}$ calculated in the sideband region:

$$\mu_{\text{multijet}} = \frac{N_{\text{4-tag Multijet}}}{N_{\text{2-tag Multijet}}} = \frac{N_{\text{4-tag data}}}{N_{\text{2-tag data}}} - \frac{N_{\text{4-tag t\bar{t}}}}{N_{\text{2-tag t\bar{t}}}}.$$  \hspace{1cm} (3)
Table 1: The number of events in data and the predicted number of background events for the 2015 and 2016 datasets, after applying the mass-dependent requirements in the sideband and control regions for the resolved analysis. The uncertainties are purely statistical. The data and prediction in the sideband agree by construction.

<table>
<thead>
<tr>
<th>Sample</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sideband</td>
<td>Control Region</td>
</tr>
<tr>
<td>Multijet</td>
<td>4196 ± 65</td>
<td>3861 ± 86</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>103 ± 6</td>
<td>159 ± 15</td>
</tr>
<tr>
<td>Total</td>
<td>4300 ± 66</td>
<td>4021 ± 88</td>
</tr>
<tr>
<td>Data</td>
<td>4300</td>
<td>3995</td>
</tr>
</tbody>
</table>

where $N_{\text{data}}^{2/4\text{-tag}}$ is the number of events observed in the sideband region in the 2- or 4-tag data sample, respectively. The yields $N_{t\bar{t}}^{2/4\text{-tag}}$ are the estimated number of $t\bar{t}$ events in the 2-/4-tag selected sideband region estimated from MC simulation. In the 2015 dataset $\mu_{\text{multijet}} = (1.45 ± 0.02) \times 10^{-2}$ and in the 2016 dataset $\mu_{\text{multijet}} = (1.12 ± 0.01) \times 10^{-2}$. The two different values of $\mu_{\text{multijet}}$ are a result of the different LHC conditions and different triggers used in each year.

After the correction process, there is agreement between the background model and sideband region data, including variables that were not directly corrected.

5.2.2 $t\bar{t}$ background

The $t\bar{t}$ background is modelled using MC simulation. However, due to the limited statistical precision of the simulated sample after the full selection, the $t\bar{t} m_{4j}$ distribution of simulated events passing the “2-tag” selection is used. A systematic uncertainty is assigned to cover the differences between the 2-tag and 4-tag $m_{4j}$ distributions.

5.2.3 Background validation

Control region data is used to validate the background model, particularly the multijet model, as well as to assign systematic uncertainties as described in Section 5.3. Table 1 compares the observed data yield in the control region with the corresponding background estimate. The modelling of the $m_{4j}$ distribution in the control region is shown in Figure 4. The 4-tag events in the control region are well-described by the background model in both normalization and $m_{4j}$ shape.

5.3 Systematic uncertainties

Systematic uncertainties on the modelling of signal and background processes are evaluated. The modelling of signals and $t\bar{t}$ backgrounds share common sources of systematic uncertainty because they depend on MC simulation: detector modelling uncertainties and theoretical uncertainty in the acceptance.

The theoretical uncertainties are fully correlated between the 2015 and 2016 datasets. The uncertainties on the acceptance comprise: missing higher-order terms in the matrix elements and PDF set, as well as
modelling of the underlying event, hadronic showers, initial- and final-state radiation. The total theoretical uncertainty is dominated by the uncertainties associated with the modelling of the initial- and final-state radiation.

The following detector modelling uncertainties are evaluated: uncertainties in the jet energy scale (JES) and resolution (JER), and uncertainties in the $b$-tagging efficiency. These uncertainties are fully correlated between 2015 and 2016. The MC simulated samples also share common luminosity uncertainties described in Section 3, which are completely uncorrelated between 2015 and 2016. The jet energy uncertainties are derived from the 2015 dataset, using in situ measurement techniques described in Refs. [50–52]. The JES systematic uncertainty is evaluated following the prescription outlined in Ref. [53]. The JER uncertainty is evaluated by smearing jet energies according to the systematic uncertainties of the resolution measurement [53]. The uncertainty in the $b$-tagging efficiency is evaluated by propagating the systematic uncertainty in the measured tagging efficiency for $b$-jets [54]. The efficiencies are measured as a function of $b$-jet $p_T$ and $\eta$. For $b$-jets with $p_T > 300 \text{ GeV}$, systematic uncertainties in the tagging efficiencies are extrapolated with MC simulation and are consequently larger [19].

Systematic uncertainties in the normalization and shape of the multijet background model are treated as uncorrelated in the 2015 and 2016 datasets. Both uncertainties are assessed for each year in the control region. For 2015, the background prediction in the control region agrees with the observed data to within 0.2%, with good agreement in the 2016 dataset as well, with the yields agreeing within 2%. To further test the robustness of the background estimation, the background models for the 2015 and 2016 datasets are re-evaluated using different sideband and control region definitions and different $b$-tagging requirements on the 2-tag sample. These variations comprise testing dummy signal regions centred on (90 GeV, 86 GeV) and (160 GeV, 153 GeV), where no observable signal is expected; using only the inner or outer half of

Figure 4: The $m_{4j}$ distribution in the control region of the resolved analysis for the data and the predicted background (top panels) for (a) the 2015 dataset and (b) the 2016 dataset. The data to background ratio (bottom panels) is shown with the statistical uncertainty on the background prediction as the grey band. The first-order polynomial fit to the data-to-background ratio above 500 GeV — used to define the multijet shape systematic uncertainty in this mass range — is shown as the solid red line, with the associated uncertainties “Tail Up” and “Tail Down” as dashed lines.
the sideband region; adding an inner radius to the control region to move it further from the signal region; shifting the centre of the control region; and tightening and loosening the $b$-tagging vetoes applied in the 2-tag sample. The largest discrepancy observed in these variations in both 2015 and 2016 is ±5%, which is used as the normalisation uncertainty.

These changes affect the kinematic and flavour compositions of the various regions used in the background prediction. The control region agreement and signal region predictions of all variations considered are all consistent to within the assigned uncertainties in the multijet background prediction.

The uncertainties in the description of the multijet $m_{4j}$ distributions is determined by comparing the background predictions to the data in the control region and dummy signal regions. The description of the kinematic turn-on and peak of the distribution are evaluated by fitting Landau functions to both 2-tag and 4-tag data in the dummy signal regions. The differences between the four-tag and two-tag fit parameters are then propagated to the signal region prediction. To evaluate the level of agreement in the high mass tail, linear fits are performed on the ratios of the distributions as shown in Figure 4. These fits, along with their uncertainties, are shown in the bottom panels of Figure 4. They both have slopes consistent with zero. The uncertainties in the multijet background shapes are defined using the uncertainties in the fitted slopes. These uncertainties are also tested further using variations in the definitions of the control and sideband regions and the $b$-tagging requirements used for the 2-tag sample. The shape in each of these variations is consistent within the assigned uncertainties.

Table 2 summarizes the relative impact of the uncertainties in the event yields.

Table 2: Summary of systematic uncertainties (expressed in percentage yield) in the total background and signal event yields in the signal region of the resolved analysis. Uncertainties are provided for both the 2015 and 2016 analyses for background, for non-resonant SM Higgs pair production and for a $G_{kk}^{*}$ resonance in the Bulk RS model with $k/M_{Pl} = 1$ and $m = 800$ GeV.

<table>
<thead>
<tr>
<th>Source</th>
<th>2015 Background</th>
<th>2015 SM $hh$</th>
<th>2015 $G_{kk}^{*}$ (800 GeV)</th>
<th>2016 Background</th>
<th>2016 SM $hh$</th>
<th>2016 $G_{kk}^{*}$ (800 GeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminosity</td>
<td>–</td>
<td>2.1</td>
<td>2.1</td>
<td>–</td>
<td>3.7</td>
<td>3.7</td>
</tr>
<tr>
<td>JER</td>
<td>–</td>
<td>5.7</td>
<td>3.3</td>
<td>–</td>
<td>5.4</td>
<td>3.5</td>
</tr>
<tr>
<td>JES</td>
<td>–</td>
<td>6.4</td>
<td>1.3</td>
<td>–</td>
<td>6.6</td>
<td>1.3</td>
</tr>
<tr>
<td>$b$-tagging</td>
<td>–</td>
<td>23</td>
<td>35</td>
<td>–</td>
<td>23</td>
<td>35</td>
</tr>
<tr>
<td>Theoretical</td>
<td>–</td>
<td>9.7</td>
<td>4.2</td>
<td>–</td>
<td>9.7</td>
<td>4.2</td>
</tr>
<tr>
<td>Multijet</td>
<td>5</td>
<td>–</td>
<td>–</td>
<td>5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$tar{t}$</td>
<td>58</td>
<td>–</td>
<td>–</td>
<td>58</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total</td>
<td>5.5</td>
<td>26</td>
<td>35</td>
<td>5.5</td>
<td>27</td>
<td>36</td>
</tr>
</tbody>
</table>

5.4 Signal region event yields

The predicted number of background events in the signal region, the number of events observed in the data, and the predicted yield for two potential signals are presented in Table 3 for both the 2015 and 2016 datasets. The numbers of predicted background events and observed events in both years are in agreement.
Table 3: The number of predicted background events in the signal region for the resolved analysis compared to the data, for the 2015 and 2016 datasets. The yields for two potential signals, SM non-resonant Higgs pair production and an 800 GeV $G^*_{KK}$ resonance with $k/\bar{M}_{Pl} = 1$ are shown. The quoted errors include both the statistical and systematic uncertainties.

<table>
<thead>
<tr>
<th>Sample</th>
<th>2015 Signal Region</th>
<th>2016 Signal Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multijet</td>
<td>$1131 \pm 68$</td>
<td>$3670 \pm 200$</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>$57 \pm 34$</td>
<td>$190 \pm 110$</td>
</tr>
<tr>
<td>Total</td>
<td>$1189 \pm 76$</td>
<td>$3860 \pm 230$</td>
</tr>
<tr>
<td>Data</td>
<td>$1231$</td>
<td>$3990$</td>
</tr>
<tr>
<td>SM $hh$</td>
<td>$0.47 \pm 0.12$</td>
<td>$1.5 \pm 0.4$</td>
</tr>
<tr>
<td>$G^*<em>{KK}$ (800 GeV), $k/\bar{M}</em>{Pl} = 1$</td>
<td>$8 \pm 3$</td>
<td>$24 \pm 8$</td>
</tr>
</tbody>
</table>

Figure 5 shows comparisons of the predicted $m_{4j}$ background distributions to that observed in the 2015 and 2016 datasets. The predicted background and observed distributions are in agreement, with no significant local excess.

Figure 5: Distributions of $m_{4j}$ in the signal region of the resolved analysis for (a) 2015 data and (b) 2016 data, compared to the predicted backgrounds. The hatched bands shown in the data/background ratio in the bottom panels represent the combined statistical and systematic uncertainties in the total background estimates. The expected signal distributions for SM non-resonant $hh$ production and $G^*_{KK}$ resonances with masses of 300 and 800 GeV are also shown.
6 Boosted analysis

The boosted analysis is optimised to discover signals that result in high-mass resonant Higgs boson pairs. The strategy is to select two Higgs candidates with mass near $m_h$ each composed of a single anti-$k_t$ jet with at least one $b$-tagged track-jet ghost-associated [48] to it. Different signal samples are defined based on the total number of $b$-tagged track-jets associated with the Higgs boson candidates. In contrast to the resolved analysis, the differences observed between the 2015 and 2016 datasets are minimal due to the higher $p_T$ thresholds and so the two datasets are combined at the event level.

As in the resolved analysis, the mass of the two Higgs boson candidates system, $m_{2J}$, is used as the final discriminant between Higgs boson pair production and the SM backgrounds.

6.1 Selection

The boosted analysis selects events with at least two anti-$k_t$ jets with $p_T > 250$ GeV, $|\eta| < 2.0$, and mass $m_J > 50$ GeV. Only the two jets with highest $p_T$ are retained for further selection. As a result of the trigger thresholds, the leading jet is required to have $p_T > 450$ GeV. This high threshold also suppresses $t\bar{t}$ events, since all decay products from a top quark decay would be contained in a single jet, with mass close to that of the top quark. The efficiency for signal events to have two jets satisfying these requirements is shown in Figure 6(a) as the curve labelled “large-$R$ jets $\geq 2$”.

Since high-mass resonances tend to produce jets that are more central than multijet background processes, the two large-$R$ jets are required to have a separation $|\Delta\eta| < 1.7$. The efficiency of this requirement is shown in Figure 6(a) as “$|\Delta\eta_{hh}| < 1.7$”.

To be considered as a Higgs boson candidate, each large-$R$ jet must have at least one $b$-tagged jet. These track-jets are required to be consistent with the primary vertex of the event as well as to satisfy $p_T > 10$ GeV and $|\eta| < 2.5$. The $b$-tagging requirement has been evaluated as 77% efficient for those jets containing $b$-hadrons in $t\bar{t}$ events. The efficiency of this requirement is shown in Figure 6(a) as “$b$-tagged track-jets $\geq 2$”.

Three separate samples of events are selected, defined by the total number of $b$-tagged track-jets associated to the Higgs boson candidates. The “two-tag-split” selection, requires each Higgs boson candidate to have exactly one associated $b$-tagged track-jet. The “three-tag” and “four-tag” samples require that there are exactly three or exactly four $b$-tagged track-jets associated to Higgs boson candidates in the event, with two $b$-tagged jets associated with one candidate and one or two associated with the other candidate. Increasing the required number of associated $b$-tagged track-jets in the event increases signal purity, but at the expense of lower signal efficiency. As illustrated in Figure 6(b), this loss of efficiency is particularly pronounced for the highest resonance masses. It is difficult to identify four distinct track-jets containing $b$-hadrons in these high mass events not only due to the inefficiency in $b$-tagging high-$p_T$ track-jets, but also because it is difficult to resolve each $b$-quark as a separate track-jet due to the extremely high Lorentz boosts.

Signal event candidates are selected if each of the large-$R$ jets has a mass consistent with that of the Higgs boson. This is defined analogously to the resolved analysis in Eq. (2):

$$X_{hh} = \sqrt{\frac{(m_{\text{lead}} - 124 \text{ GeV})^2}{0.1m_{\text{lead}}^2} + \frac{(m_{\text{subl}} - 115 \text{ GeV})^2}{0.1m_{\text{subl}}^2}} < 1.6$$

(4)
Figure 6: (a) The absolute selection efficiency of the boosted analysis as a function of the generated $G_{KK}^*$ mass at each stage of the event selection. The trigger efficiency is 100% after the requirement for 2 large-$R$ jets, so is not shown. (b) The selection efficiency of various $b$-tagged track-jet requirements after requiring $X_{hh} < 1.6$, as a function of the generated $G_{KK}^*$ mass.

with the $R = 0.4$ dijet masses replaced by the large-$R$ jet masses, $m_1^{\text{lead}}$ and $m_1^{\text{subl}}$. The requirement of $X_{hh} < 1.6$ is retained to define the signal region in the leading–subleading Higgs boson candidate mass plane, as illustrated in Figure 7. The efficiency of this requirement is shown in Figure 6(a) as “$X_{hh} < 1.6$”.

### 6.2 Background estimation

As in the resolved analysis, the principle background is multijet events, which comprise 83-87% of the total background (depending on the number of $b$-tagged track-jets required), with the remainder being dominated by $t\bar{t}$ events. All other sources of background, including processes involving Higgs bosons, are negligible. The multijet events are modelled entirely using data. In contrast, while the $t\bar{t}$ yield is also estimated using a data-driven technique, the $m_{23}$ shape is taken from MC.

The multijet background is modelled using independent data samples selected using the same trigger and selection requirements as described above, except for the $b$-tagging requirements: while the Higgs boson candidates must have the requisite number of track-jets associated to them to model each sample, it is required that all these track-jets must fail the $b$-tagging requirement. These three samples will be referred to collectively as “zero-tag” samples and have a multijet fraction of over 99%.

In a similar manner to the estimation of multijet yields in the resolved analysis, the normalisation of the multijet background is derived using a signal-free sideband region, in this case defined as:

$$36 \text{ GeV} < \sqrt{(m_1^{\text{lead}} - 124 \text{ GeV})^2 + (m_1^{\text{subl}} - 115 \text{ GeV})^2} < 63 \text{ GeV}.$$
In contrast to the resolved analysis however, this sideband region is also used to derive the normalisation of the $t\bar{t}$ background. A control region between the signal and sideband regions is again used to verify the background models and to assign systematic uncertainties. The regions of the $m_{\text{lead}}^{\text{J}}$, $m_{\text{subl}}^{\text{J}}$ plane are depicted in Figure 7.

Figure 7: The $m_{\text{lead}}^{\text{subl}}$ vs. $m_{\text{lead}}^{\text{lead}}$ distribution for the boosted analysis background model. The signal region is the area surrounded by the dashed red contour line, centred on ($m_{\text{lead}}^{\text{lead}}=124\text{ GeV}$, $m_{\text{subl}}^{\text{subl}}=115\text{ GeV}$). The control region is the area between the signal region and the orange contour line. The sideband region is the area between the control region and the yellow contour line.

The background yield in each of the three signal samples, $N_{\text{bckgrd}}^{n\text{-tag}}$ (where $n$-tag represents 2-, 3- and 4-tag), is evaluated using the following expression:

$$N_{\text{bckgrd}}^{n\text{-tag}} = \mu_{\text{multijet}}^{n\text{-tag}} N_{\text{multijet}}^{0\text{-tag}} + \alpha_{t\bar{t}}^{n\text{-tag}} N_{t\bar{t}}^{n\text{-tag}}$$

(5)

where $N_{\text{multijet}}^{0\text{-tag}}$ is the number of multijet events in the 0-tag sample and $N_{t\bar{t}}^{n\text{-tag}}$ are the numbers of events predicted by the $n$-tag $t\bar{t}$ simulation. The parameter $\mu_{\text{multijet}}^{n\text{-tag}}$ corresponds to the ratio of multijet event yields in the $n$-tag to 0-tag samples. Finally, the parameter $\alpha_{t\bar{t}}^{n\text{-tag}}$ is a scale factor designed to correct the $t\bar{t}$ event yield estimated from the MC simulation.

The values of $\mu_{\text{multijet}}^{n\text{-tag}}$ and $\alpha_{t\bar{t}}^{n\text{-tag}}$ are extracted using Equation 5 from binned likelihood fits to the leading $R=1.0$ jet mass distribution, as illustrated in Figure 8. The leading jet mass distributions (both normalisation and shape) for multijet events are extracted from the relevant 0-tag sample, after subtraction of the $t\bar{t}$ contributions predicted by MC simulation. The $t\bar{t}$ distributions are taken directly from the MC simulation. The fitted values of $\mu_{\text{multijet}}$ and $\alpha_{t\bar{t}}$ are given in Table 4.

Similar to the resolved analysis, corrections are made for differences between the 0-tag and 2-tag-split, 3-tag and 4-tag samples by reweighting events in the 0-tag sample. These weights are again derived in the sideband regions, from fits to the ratio of the total background model to data for those distributions.
Figure 8: The leading $R = 1.0$ jet mass distributions in the sideband regions of (a) the 2-tag-split, (b) the 3-tag and (c) the 4-tag samples. The normalisation of the multijet and $\bar{t}t$ background predictions is estimated using a binned likelihood fit to the data. The grey hatched bands represent the statistical uncertainties on the background templates.
Table 4: The fitted values of $\mu_{\text{multijet}}$ and $\alpha_{t\bar{t}}$ and their correlation for the 2-tag-split, 3-tag and 4-tag samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\mu_{\text{multijet}}$</th>
<th>$\alpha_{t\bar{t}}$</th>
<th>$\rho(\mu_{\text{multijet}}, \alpha_{t\bar{t}})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-tag-split</td>
<td>$(3.63 \pm 0.04) \times 10^{-2}$</td>
<td>$1.44 \pm 0.06$</td>
<td>$-0.65$</td>
</tr>
<tr>
<td>3-tag</td>
<td>$(8.38 \pm 0.17) \times 10^{-3}$</td>
<td>$1.77 \pm 0.16$</td>
<td>$-0.61$</td>
</tr>
<tr>
<td>4-tag</td>
<td>$(8.48 \pm 0.73) \times 10^{-5}$</td>
<td>$2.5 \pm 1.1$</td>
<td>$-0.63$</td>
</tr>
</tbody>
</table>

that exhibit the largest discrepancies: the $p_T$ of the leading Higgs boson candidate and the $p_T$ of the leading track-jets associated to each of the Higgs boson candidate. The reweighting is performed using up to three-dimensional distributions but is iterated so that correlations between the three variables are taken into account. After the correction process, there is agreement between the background model and sideband region data.

The impact of limited statistical precision for $m_{2J} > 1.2$ TeV in the multijet and $t\bar{t}$ models is ameliorated using a background smoothing procedure, which fits the background distributions with the following function:

$$f(m_{2J}) = p_1 \left(1 - \frac{m_{2J}}{\sqrt{s}}\right)^{p_2} \left(\frac{m_{2J}}{\sqrt{s}}\right)^{p_3}$$

where $p_i$ are fit parameters. The multijet background is fitted in the range $1.2$ TeV < $m_{2J}$ < $3.0$ TeV, while the $t\bar{t}$ background is fitted in the range $1.1$ TeV < $m_{2J}$ < $3.0$ TeV. These fitted functional forms are then used to generate smooth background templates for the statistical analysis. Since very few simulated $t\bar{t}$ events pass the full four-tag selection, the $t\bar{t}$ distribution in the four-tag sample is taken from the three-tag distribution. A systematic uncertainty is assigned to account for this approximation.

The modelling of the background yields and kinematics is validated in the control regions of the 2-tag-split, 3-tag and 4-tag samples. Good agreement is observed between the yield in data and the yield of predicted backgrounds in the sideband and control regions of each of the samples, as shown in Table 5. Figure 9 compares the predicted background dijet mass distributions to data in the control regions of the three samples. The systematic uncertainties derived in Section 6.3 are shown. Good agreement is obtained in all cases.

Table 5: The number of events in data and predicted background events in the sideband and control regions of the 2-tag-split, 3-tag and 4-tag samples for the boosted analysis. The number of multijet and $t\bar{t}$ background events in the sideband regions are constrained by the number of observed events, as explained in the text. The uncertainties are statistical, with fit uncertainties included for backgrounds. The anti-correlation between the multijet and $t\bar{t}$ yields is accounted for in the uncertainty on the total background yield.

<table>
<thead>
<tr>
<th>Sample</th>
<th>2-tag-split</th>
<th>3-tag</th>
<th>4-tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multijet</td>
<td>$19,400 \pm 200$</td>
<td>$5,917 \pm 62$</td>
<td>$4,294 \pm 89$</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>$3,860 \pm 160$</td>
<td>$1,038 \pm 45$</td>
<td>$720 \pm 68$</td>
</tr>
<tr>
<td>Total</td>
<td>$23,260 \pm 160$</td>
<td>$6,954 \pm 52$</td>
<td>$5,014 \pm 73$</td>
</tr>
<tr>
<td>Data</td>
<td>$23,277$</td>
<td>$7,200$</td>
<td>$5,007$</td>
</tr>
</tbody>
</table>
Figure 9: The $m_{4j}$ distribution in the control region of the boosted analysis for the data and the predicted background (top panels) for (a) the 2-tag-split, (b) the 3-tag and (c) the 4-tag samples. The data to background ratio (bottom panels) is shown with the statistical uncertainty plus the control region normalisation and shape uncertainties on the background prediction shown as the grey hatched band.
6.3 Systematic uncertainties

Evaluation of systematic uncertainties in the boosted analysis generally follows the same approach as that described in Section 5.3. Differences are highlighted here.

The large-\(R\) jet energy and mass uncertainties are derived \textit{in situ} from 13 TeV \(pp\) collisions, using techniques described in [55]. The uncertainty in the \(b\)-tagging efficiency for track-jets is evaluated with the same method used for \(R = 0.4\) calorimeter jets.

Systematic uncertainties in the normalization and shape of the background model are assessed in the control region. The background predictions in each of the 2-tag-split, 3-tag and 4-tag control regions agree with the observed data to within statistical uncertainties. Further tests are made using variations of the sideband and control region definitions similar to those performed for the resolved analysis, and the largest deviation in these control region or signal region yields are assigned as systematic uncertainties in the multijet background normalizations. The multijet background shape uncertainties are assessed in the control regions: the data is smoothed like the prediction to reduce the impact of statistical fluctuations and then the ratio to the prediction is used to define the shape uncertainty shown in Figure 9. The \(t\bar{t}\) shape for the four-tag sample has an additional uncertainty due to the use of the three-tag distribution. The impact of requiring the extra \(b\)-tagged track-jet is estimated using the ratio of the three-tag and two-tag \(t\bar{t}\) shapes. An additional source of shape uncertainty for both the multijet and \(t\bar{t}\) models is the use of the fit-based smoothing procedure at high \(m_{2J}\). This uncertainty is evaluated by considering alternative functional forms and different fit ranges and taking the largest differences observed.

Relative systematic uncertainties in both the background and signal event yields are summarized in Table 6 for the three samples. For the background, the entries labeled “Statistical” correspond to the statistical uncertainty from the fit to the leading large-\(R\) jet mass (see Section 6.2) used to extract the multijet and \(t\bar{t}\) background yields, taking the correlation between these yields into account. It also includes the uncertainties associated with modelling the distribution of leading large-\(R\) jet mass in \(t\bar{t}\) and the statistical uncertainty associated with the data yield in the 0-tag sample. Uncertainties in the \(m_{2J}\) shape of the multijet and \(t\bar{t}\) backgrounds are not listed in Table 6, as they do not affect the event yields, but are accounted for in the statistical analysis.

6.4 Signal region event yields

The predicted number of background events in the signal region, the number of events observed in the data, and the predicted yield for two potential signals are presented in Table 7 for the 2-tag-split, 3-tag and 4-tag samples. The numbers of predicted background events and observed events are in agreement.

Figure 10 shows comparisons of the predicted \(m_{2J}\) background distributions to that observed in the data. The predicted background and observed distributions are in agreement, with no significant local excess.
Table 6: Summary of systematic uncertainties (expressed in percentage yield) in the total background and signal event yields in the signal region of the boosted analysis. Uncertainties are provided for each of the three samples for background and for a $G^{\ast}_{KK}$ resonance in the Bulk RS model with $k/\overline{M}_{Pl} = 1$ and $m = 2.0\,\text{TeV}$.

<table>
<thead>
<tr>
<th>Source</th>
<th>2-tag-split</th>
<th>3-tag</th>
<th>4-tag</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Background</td>
<td>$G^{\ast}_{KK}$ (2 TeV)</td>
<td>Background</td>
</tr>
<tr>
<td>Luminosity</td>
<td>-</td>
<td>2.9</td>
<td>-</td>
</tr>
<tr>
<td>JER</td>
<td>-</td>
<td>0.1</td>
<td>-</td>
</tr>
<tr>
<td>JMR</td>
<td>-</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>JES/JMS</td>
<td>-</td>
<td>4.5</td>
<td>-</td>
</tr>
<tr>
<td>$b$-tagging</td>
<td>-</td>
<td>58</td>
<td>-</td>
</tr>
<tr>
<td>Theoretical</td>
<td>-</td>
<td>2.7</td>
<td>-</td>
</tr>
<tr>
<td>Bkg Estimate</td>
<td>4.4</td>
<td>-</td>
<td>4.6</td>
</tr>
<tr>
<td>Statistical</td>
<td>0.5</td>
<td>1.4</td>
<td>1.1</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>1.6</td>
<td>-</td>
<td>4.7</td>
</tr>
<tr>
<td>Total Sys</td>
<td>4.7</td>
<td>59</td>
<td>6.6</td>
</tr>
</tbody>
</table>

Table 7: The number of predicted background events in the signal region for the boosted analysis compared to the data, for the 2-tag-split, 3-tag and 4-tag datasets. The yields for a potential signal of a 2.0 TeV $G^{\ast}_{KK}$ resonance with $k/\overline{M}_{Pl} = 1$ is shown. The quoted errors include both the statistical and systematic uncertainties. The anti-correlation between the multijet and $t\bar{t}$ yields is accounted for in the uncertainty on the total background yield.

<table>
<thead>
<tr>
<th>Sample</th>
<th>2-tag-split</th>
<th>3-tag</th>
<th>4-tag</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Background</td>
<td>$G^{\ast}_{KK}$ (2 TeV)</td>
<td>Background</td>
</tr>
<tr>
<td>Multijet</td>
<td>2.310 ± 240</td>
<td>515 ± 41</td>
<td>32.6 ± 7.6</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>460 ± 170</td>
<td>81 ± 37</td>
<td>5.7 ± 5.2</td>
</tr>
<tr>
<td>Total</td>
<td>2.770 ± 130</td>
<td>596 ± 39</td>
<td>38.3 ± 9.0</td>
</tr>
<tr>
<td>Data</td>
<td>2.813</td>
<td>671</td>
<td>32</td>
</tr>
</tbody>
</table>

$G^{\ast}_{KK}$ (2 TeV), $k/\overline{M}_{Pl} = 1$ 0.17 ± 0.10 0.31 ± 0.06 0.15 ± 0.06
Figure 10: Distributions of $m_{2J}$ in the signal regions of the boosted analysis for: (a) the two-tag-split sample; (b) the three-tag sample; and (c) the four-tag sample, compared to the predicted backgrounds. The grey hatched bands represent the combined statistical and systematic uncertainties in the total background estimates. The expected signal shape for a $G_{kk}^*$ resonance of mass 2.0 TeV is also shown.
7 Statistical analysis

The statistical interpretation of results from the resolved and boosted analyses are made separately using the statistical procedure described in Ref. [1] and references therein. A test statistic based on the profile likelihood ratio [56] is used to test hypothesized values of $\mu = \sigma / \sigma_{\text{theory}}$, the global signal strength factor, separately for each model tested. The statistical analysis described below is performed using the data observed in the signal regions. For the resolved analysis, the $m_{4j}$ distribution is used as the final search discriminant, while for the boosted analysis the $m_{2j}$ distribution is used.

Systematic uncertainties are treated within each signal region using Gaussian or log-normal constraint terms in the definition of the likelihood function. For the resolved analysis, the 2015 and 2016 datasets are fitted simultaneously. In the boosted analysis, the data from the 2-tag-split, 3-tag and 4-tag signal regions are fitted simultaneously.

7.1 Exclusion limits

The data are used to set upper limits on the cross sections for the different benchmark signal processes. Exclusion limits are based on the value of the statistic $CL_s$ [57], with a value of $\mu$ regarded as excluded at the 95% confidence level (CL) when $CL_s$ is less than 5%.

The nonresonant search is performed using the resolved analysis, since it has better sensitivity than the boosted analysis. Using the SM $hh$ nonresonant production as the signal model, the observed 95% CL upper limit is $\sigma(pp \to hh \to b\bar{b}b\bar{b}) < 330$ fb, a value to be compared with the inclusive SM prediction (as defined in Section 3) of $\sigma(pp \to hh \to b\bar{b}b\bar{b}) = 11.3^{+0.9}_{-1.0}$ fb. The observed exclusion is within one standard deviation of the expected 95% C.L. upper limit of $\sigma(pp \to hh \to b\bar{b}b\bar{b}) < 430$ fb.

For the resonant Higgs-boson pair production search, the resolved and boosted analyses offer their best sensitivity in complementary resonance mass regions. The resolved analysis gives a more stringent expected exclusion limit for resonance masses up to (and including) 1000 GeV, while the boosted analysis offers better sensitivity beyond that mass. A simple combination of the separate exclusion limits from the resolved and boosted analyses is used. This is achieved by taking the limit from the analysis with the more stringent expected exclusion at each mass point for each of the signal models.

Figure 11 shows the combined 95% CL upper limits for a spin-2 $G^*_KK$ in the bulk RS model with $k/\tilde{M}_{Pl} = 1$. Limits on $\sigma(pp \to G^*_KK \to hh \to b\bar{b}b\bar{b})$ are set in the range between 1000 fb and 2 fb for masses between 300 and 3000 GeV. These cross section upper limits translate into observed (expected) excluded mass range of $360$ GeV $\leq m_{G^*_KK} \leq 860$ GeV ($380$ GeV $\leq m_{G^*_KK} \leq 910$ GeV).

The search sensitivity of this analysis exceeds that of the previous analysis of the $\sqrt{s} = 13$ TeV 2015 dataset [10] across the entire mass range. Analysis improvements have led to large sensitivity gains for non-resonant signal and for resonance signals with $m_{hh} < 500$ GeV and $m_{hh} > 2.5$ TeV.
8 Conclusion

A search for both resonant and non-resonant production of pairs of Standard Model Higgs bosons has been carried out in the dominant $b\bar{b}b\bar{b}$ channel, using 13.3 fb$^{-1}$ of LHC $pp$ collision data at $\sqrt{s} = 13$ TeV collected by ATLAS in 2015 and 2016.

Results are reported for the resolved analysis with each $h \rightarrow b\bar{b}$ decay reconstructed as two separate $b$-tagged jets and for the boosted analysis with each $h \rightarrow b\bar{b}$ decay reconstructed as a single large-radius jet associated with at least one small-radius $b$-tagged track-jet. No significant data excess is observed above the estimated background consisting mainly of multijet and $t\bar{t}$ events. Upper limits on the production cross section times branching ratio to the $b\bar{b}b\bar{b}$ final state are set for spin-2 resonances with values ranging between 1000 and 2 fb (at 95% CL) for resonance masses in the range between 300 and 3000 GeV. For non-resonant production, the upper limit is 330 fb (at 95% CL). The search sensitivity of this analysis exceeds that of the previous analysis [10] of the $\sqrt{s} = 13$ TeV 2015 dataset for non-resonant signals and for resonant signals across the mass range $300 \text{ GeV} < m_{hh} < 3000 \text{ GeV}$.
References


Figure 12: The expected and observed upper limit for $pp \rightarrow G^*_{KK} \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ in the bulk RS model with $k/M_{Pl} = 1$ at the 95% confidence level. The results of the resolved analysis are used up to a mass of 1000 GeV and those of the boosted analysis are used at higher mass where its expected sensitivity is higher. The red curves show the predicted cross sections as a function of resonance mass for the bulk RS model with $k/M_{Pl} = 1$. The drop in cross section for masses below $m_{G^*_{KK}} = 350$ GeV is due to a sharp drop in $\text{Br}(G^*_{KK} \rightarrow hh)$. The previous analysis is documented in arXiv:1606.04782.