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

G. Aad et al.**

(ATLAS Collaboration)

(Received 17 February 2022; accepted 15 March 2022; published 11 May 2022)

A search for resonant Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state is presented. The analysis uses 126 fb$^{-1}$–139 fb$^{-1}$ of $pp$ collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector at the Large Hadron Collider. The analysis is divided into two channels, targeting Higgs boson decays which are reconstructed as pairs of small-radius jets or as individual large-radius jets. Spin-0 and spin-2 benchmark signal models are considered, both of which correspond to resonant $HH$ production via gluon-gluon fusion. The data are consistent with Standard Model predictions. Upper limits are set on the production cross section times branching ratio to Higgs boson pairs of a new resonance in the mass range from 251 GeV to 5 TeV.

DOI: 10.1103/PhysRevD.105.092002

I. INTRODUCTION

The discovery of the 125 GeV Higgs boson ($H$) [1–4] at the Large Hadron Collider (LHC) enables and motivates the search for resonances decaying into Higgs boson pairs ($HH$). Such searches have been carried out by the ATLAS and CMS collaborations in $pp$ collisions at center-of-mass energies ($\sqrt{s}$) of both 8 TeV and 13 TeV. All of these have found results that are consistent with the Standard Model (SM) prediction that no such resonances exist. The ATLAS Collaboration has set constraints on resonant $HH$ production via vector-boson fusion in the $b\bar{b}b\bar{b}$ final state [5] and via gluon-gluon fusion (ggF) in the boosted $b\bar{b}\tau^+\tau^-$ final state [6] using the full Run 2 dataset. It also searched for resonant $HH$ production in the $b\bar{b}b\bar{b}$ decay mode via gluon-gluon fusion [7], in the $ggb\bar{b}$ [8], $b\bar{b}\tau^+\tau^-$ [9], $WWb\bar{b}$ [10], $WWWW$ [11], and $WW\gamma\gamma$ [12] decay modes, and in a combination [13], using up to 36.1 fb$^{-1}$ of the Run 2 dataset at $\sqrt{s} = 13$ TeV. The CMS Collaboration has similarly set constraints in the $b\bar{b}b\bar{b}$ [14], $b\bar{b}\gamma\gamma$ [15], $b\bar{b}\tau^+\tau^-$ [16], $b\bar{b}\ell\ell\nu\nu$ [17], and $b\bar{b}ZZ$ [18] decay modes, and a combination of these [19], using 35.9 fb$^{-1}$ of $\sqrt{s} = 13$ TeV data, as well as the $b\bar{b}\tau^+\tau^-$ [20] decay mode and final states with $b$-quarks and leptons [21] using the full Run 2 dataset.

This paper presents a search for resonant pair production of Higgs bosons via gluon-gluon fusion in the $b\bar{b}b\bar{b}$ final state, using the full LHC Run 2 dataset collected by ATLAS. The results are interpreted in terms of two representative benchmark models: a generic spin-0 boson, $X$ (as, for example, predicted by two-Higgs-doublet models [22] such as the Minimal Supersymmetric Standard Model [23,24]), and a spin-2 Kaluza–Klein graviton, $G_{\text{KK}}$, in the context of the bulk Randall-Sundrum (RS) model [25–28]. In both cases, only the gluon-gluon fusion production mode is considered. No further assumptions are made on the signal models except for the spin hypothesis and generated resonance width. Example production diagrams for these signal models are shown in Fig. 1. Throughout this analysis, the nominal $H \rightarrow b\bar{b}$ branching ratio is taken to be 0.582, corresponding to the SM value at a Higgs boson mass of 125 GeV [29].

The analysis is divided into two complementary channels: resolved, in which each of the four $b$-quarks from the $H$ decays leads to an individually reconstructed jet, and boosted, which targets the topology where each $H$ is produced with large transverse momentum ($p_T$) and its decay products are reconstructed as a single large-radius jet. The resolved and boosted channels target low and high resonance masses, respectively. The resolved channel covers resonance masses from 251 GeV to 1.5 TeV, and the boosted channel covers resonance masses from 900 GeV to 5 TeV. The two channels are statistically combined in the mass range where they overlap.

In addition to utilizing the full Run 2 dataset and benefiting from progress in the ATLAS $b$-jet identification

---

**Full author list given at the end of the article.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP3.
algorithms, this analysis is improved over the previous ATLAS search [7] in several ways. In the resolved channel, a machine-learning algorithm is used to pair the jets into Higgs boson candidates and the fully data-driven background model is substantially improved with a neural-network-based reweighting procedure. In the boosted channel, variable-radius track jets are used for $b$-tagging to recover signal acceptance for the highest resonance masses, which fell in the previous analysis as the Higgs boson decay products became very collimated, and the search is extended to cover the previously unexplored resonance mass range between 3 TeV and 5 TeV. In both channels, the analysis uses a new neural-network-based $b$-tagging algorithm DL1r, which performs better than the older MV2 algorithm [30,31].

II. ATLAS DETECTOR

The ATLAS detector [32] at the LHC covers nearly the entire solid angle around the collision point. It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadron calorimeters, and a muon spectrometer incorporating three large superconducting air-core toroidal magnets.

The inner-detector (ID) system is immersed in a 2 T axial magnetic field and provides charged-particle tracking in the range $|\eta| < 2.5$. The high-granularity silicon pixel detector covers the vertex region and typically provides four spacepoint measurements per track, the first hit normally being in the insertable B-layer installed before Run 2 [33,34]. The next layer outward is the silicon microstrip tracker, which usually provides eight measurements per track. These silicon detectors are complemented by the transition radiation tracker, which enables radially extended track reconstruction up to $|\eta| = 2.0$.

The calorimeter system covers the pseudorapidity range $|\eta| < 4.9$. Within the region $|\eta| < 3.2$, electromagnetic calorimetry is provided by barrel and endcap high-granularity lead/liquid-argon (LAr) calorimeters, with an additional thin LAr presampler covering $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters. Hadron calorimetry is provided by the steel/scintillator-tile calorimeter, segmented into three barrel structures within $|\eta| < 1.7$, and two copper/LAr hadron endcap calorimeters. The solid angle coverage is completed with forward copper/LAr and tungsten/LAr calorimeter modules optimized for electromagnetic and hadronic energy measurements respectively.

The muon spectrometer (MS) comprises separate trigger and high-precision tracking chambers measuring the deflection of muons in a magnetic field generated by the superconducting air-core toroidal magnets. The field integral of the toroids ranges between 2.0 and 6.0 T · m across most of the detector. A set of precision chambers covers the region $|\eta| < 2.7$ with three layers of monitored drift tubes, complemented by cathode-strip chambers in the forward region, where the background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate chambers in the barrel, and thin-gap chambers in the endcap regions.

Interesting events are selected by the first-level trigger system implemented in custom hardware, followed by selections made by algorithms implemented in software in the high-level trigger [35]. The first-level trigger accepts events from the 40 MHz bunch crossings at a rate below 100 kHz, which the high-level trigger further reduces in order to record events to disk at about 1 kHz.

An extensive software suite [36] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

III. DATA AND SIMULATED SAMPLES

A. Data sample

This analysis is performed using LHC $pp$ collision data at $\sqrt{s} = 13$ TeV. Only data collected during stable beam conditions with all relevant detector systems functional are used [37]. The resolved (boosted) channel uses 126 fb$^{-1}$ (139 fb$^{-1}$) of data collected in 2016–2018 (2015–2018).
The triggers for both channels are based on jets reconstructed using the anti-$k_t$ algorithm [38,39]. For the resolved (boosted) channel, the jets are clustered with a radius parameter of $R = 0.4$ ($R = 1.0$).

The resolved channel uses a combination of 12 triggers with various requirements on the transverse energy ($E_T$) and $b$-tagging status of the jets [40]. These include requirements on any one of four sets of objects:

(a) two $b$-jets plus two additional jets ($2b+2j$),
(b) two $b$-jets plus one additional jet ($2b+1j$),
(c) a single high-$E_T$b-jet, and
(d) two $b$-jets plus a high $H_T$, defined as the scalar sum of all jets’ $E_T$ ($2b+H_T$).

The minimum $E_T$ requirements on the jets are 35 GeV for the $b$-jets from the $2b+2j$ triggers, 55 GeV for the $b$-jets from the $2b+H_T$ triggers, 100 to 150 GeV (depending on the year) for the additional jet used for the $2b+1j$ triggers, and finally 225 GeV or 300 GeV for the single high-$E_T$b-jet trigger. The minimum $H_T$ requirement is 300 GeV for the triggers which include it. In the trigger, $H_T$ is computed using all jets with $E_T \geq 30$ GeV. The efficiency of individual triggers varies from a few percent to up to 80% depending on the kinematic and $b$-tagging requirements of the trigger, and the signal hypothesis. The choice of triggers is optimized to maximize the signal efficiency over the full range of hypothesized resonance mass values. The set of triggers used depends on the data-taking year, and the triggers from each year have different effects on kinematic distributions. This results in a different signal-to-background ratio in each year’s data. Therefore, the datasets from each year are treated independently until they are combined in the statistical analysis. During 2016 data taking, a fraction of the data (8.3 fb$^{-1}$) was affected by an inefficiency in the online vertex reconstruction, which reduced the efficiency of the algorithms used to identify $b$-jets; those events were not retained for further analysis. This results in an integrated luminosity of 24.6 fb$^{-1}$ for the 2016 data in the resolved channel. The integrated luminosities of the 2017 and 2018 datasets are 43.7 fb$^{-1}$ and 57.7 fb$^{-1}$, respectively.

In the boosted channel, events were selected from the 2015 dataset using a trigger that required a single jet, $J$, with $E_T > 360$ GeV. In 2016, 2017 and 2018 a similar trigger was used but requiring $E_T > 420$ GeV. The 2017 and 2018 triggers have additional requirements on the mass of the jet of $m_J > 40$ GeV and $m_J > 35$ GeV, respectively. The efficiency of these triggers is 98% for data passing the jet requirements described in Sec. VI, so the triggers do not significantly impact any relevant kinematic distributions and the datasets corresponding to each year are combined into one dataset.

**B. Simulated samples**

Monte Carlo (MC) simulation is used for the modeling of signal events and, in the case of the boosted channel, for the modeling of the $t\bar{t}$ background. The ATLAS detector response is simulated with GEANT4 [41] for background samples, spin-0 signal samples with a resonance mass of 1 TeV or higher, and all spin-2 signal samples. AtlFastII [42], which utilizes a fast calorimeter simulation, is used for spin-0 signals with resonance masses below 1 TeV.

The signal processes for both benchmark models were simulated at leading order (LO) in $a_s$, using MadGraph5_aMC@NLO2.6.1 [43] for the spin-0 samples and MadGraph5_aMC@NLO2.2.2 for the spin-2 samples. For both cases, the NNPDF2.3LO [44] parton distribution function (PDF) set was used. Both resonances were produced via gluon–gluon fusion and were forced to decay into a pair of SM Higgs bosons, as indicated in Fig. 1. Signal samples were generated for resonance masses in a range from 251 GeV to 5 TeV, with increased spacing between the higher mass-points. The spacing becomes larger than the detector resolution at high masses. For the spin-0 model, interpolation methods [45] are used to estimate the mass distribution between the points above 2 TeV for which samples were simulated. The spin-2 resonance is wide compared to the mass spacing, so no interpolation is required.

For the spin-0 case, a two-Higgs-doublet model was used in the event generation, taking the heavy $CP$-even neutral scalar as the resonance of interest. Its width was set to be much smaller than the detector resolution, and the other non-SM particles in this model do not enter the generation (At LO in QCD, there are no Feynman diagrams for this process containing them). Hadronization and parton showering were modeled with HERWIG7.1.3 [46], using the MMHT2014LO [47] PDF set for the parton shower and EvtGen 1.6.0 [48] to model heavy-flavor decays. The HERWIG7.1 default set of tuned underlying-event parameters was used. No theoretical cross sections are required for this process; the simulated events are used purely to model the kinematics.

For the spin-2 case, a Kaluza–Klein graviton with $k/M_{pl} = 1$ is taken as the benchmark, where $k$ is the curvature of the warped extra dimension and $M_{pl} = 2.4 \times 10^{18}$ GeV is the effective four-dimensional Planck scale. The generated width, based on the model prediction, ranges from 3% to 20% of the resonance mass, which is not negligible compared to the detector resolution. The hadronization and showering were modeled using PYTHIA8.186 [49] with EvtGen 1.2.0 for heavy-flavor decays. The A14 set of tuned underlying-event parameters [50] and the NNPDF2.3LO PDF set were used. The cross sections for this process are taken from Ref. [28]. They are used solely for setting limits on the graviton mass.

Top quark pair production ($t\bar{t}$) was simulated at next-to-leading order (NLO) in $a_s$ using Powheg Box v2 [51–54]. Parton showering, hadronization, and the underlying event were modeled using PYTHIA8.230 with EvtGen 1.6.0 for heavy-flavor decays. The matrix element calculation uses
NNPDF3.0NLO [55] as the PDF set, while the parton shower and underlying-event modeling uses NNPDF2.3LO [44] and the A14 set of tuned parameters. The damping parameter \( h_{\text{damp}} \), which effectively regulates radiation at high \( p_T \), was set to 1.5 times the top quark mass. The \( t\bar{t} \) simulation is normalized using the value of the inclusive cross section calculated with Top++ 2.0 [56,57]. This accounts for next-to-next-to-leading-order (NNLO) corrections in \( \alpha_s \), including next-to-next-to-leading logarithmic (NLLN) resummation of soft gluon terms.

Multijet background processes were modeled using PYTHIA8.235. This simulates pure quantum chromodynamics (QCD) 2-to-2 interactions at LO in \( \alpha_s \). Events were showered using the parton shower native to PYTHIA, which includes radiation and splitting that can result in additional jets, along with EvGen 1.6.0 for heavy-flavor decays. The NNPDF2.3LO PDF set were used.

Other background processes, such as SM \( H, HH \), and electroweak diboson production, have been estimated to give negligible contributions to the selected event yields and are therefore not included.

The effect of multiple interactions in the same and neighboring bunch crossings (pileup) was modeled by overlaying each simulated hard-scattering event with inelastic proton–proton (\( pp \)) events generated with PYTHIA8.186 using the NNPDF2.3LO PDF and the A3 set of tuned parameters [58].

### IV. OBJECT RECONSTRUCTION

Primary vertices from proton-proton interactions are reconstructed using at least two charged-particle tracks with \( p_T > 500 \) MeV measured with the ID [59]. The vertex which has the largest sum of squared track momenta (\( \sum p_T^2 \)) is selected as the hard-scatter primary vertex.

Hadronic jets are reconstructed using the anti-\( k_t \) algorithm [38,39]. Depending on the use case, different input objects and radius parameters \( R \) are used.

For the resolved channel, small-\( R \) jets are clustered using \( R = 0.4 \) with particle-flow objects as inputs [60]. Particle-flow objects are charged-particle tracks matched to the hard-scatter vertex and calorimeter energy clusters following an energy subtraction algorithm that removes the calorimeter deposits associated with good-quality tracks from any vertex. The tracking information is used to improve the clusters’ energy resolutions. The momenta of these jets are calibrated in a multistep procedure [61]. Jets with \( p_T < 60 \) GeV and \( |\eta| < 2.4 \) must also satisfy a requirement based on the output of the multivariate “jet vertex tagger” (JVT) algorithm, which is used to identify and reject jets in which much of the energy originates from pileup interactions [62].

The “Tight” working point, corresponding to an average signal efficiency of 96%, is used and jets failing this requirement are discarded. All small-\( R \) jets are required to have \( p_T > 40 \) GeV and \( |\eta| < 2.5 \). Any jets failing these requirements are discarded and not used further, except where stated explicitly.

Additional small-\( R \) jets are reconstructed from topological clusters of energy deposits in the calorimeter [63] instead of particle-flow objects. These jets are used exclusively for the purpose of applying quality criteria to identify events which are consistent with noise in the calorimeter or noncollision background [64]. They are calibrated in the same way as the small-\( R \) jets reconstructed from particle-flow objects. If an event contains at least one jet which has \( p_T > 20 \) GeV, passes the JVT, and fails to meet these quality criteria, the event is rejected.

For the boosted channel, large-\( R \) jets are clustered using \( R = 1.0 \) with topological clusters of energy deposits in calorimeter cells as the input. The clusters are locally calibrated [65] before being combined into jets. After these large-\( R \) jets are created, a trimming procedure [66] is applied to mitigate the effects of pileup: the constituents are reclusted into “subjets” using the \( k_t \) algorithm [67] with \( R = 0.2 \), and any of these subjets with less than 5% of the large-\( R \) jet’s \( p_T \) are removed. The large-\( R \) jets are calibrated following a procedure similar to that for the small-\( R \) jets; however, there is no area-based pileup subtraction step or global sequential calibration [68]. Additionally, the mass of each large-\( R \) jet is calibrated using both calorimeter and track information [69].

The boosted channel also makes use of track-jets to identify individual \( b \)-hadron decays within the large-\( R \) jet. ID tracks are clustered using the anti-\( k_t \) algorithm with a variable radius. The effective radius \( R \) is inversely proportional to the \( p_T \) of the constituent(s) in question: \( R = \rho/p_T \). Here, a value of \( \rho = 30 \) GeV is used. Minimum and maximum values of this effective radius are set at \( R_{\text{min}} = 0.02 \) and \( R_{\text{max}} = 0.4 \). Track jets do not have a dedicated calibration; their momenta are taken to be the vector sum of the momenta of their constituent tracks. After being reconstructed, these track jets are exclusively matched to large-\( R \) jets using the ghost association method [70].

A \( b \)-tagging algorithm [30,31] is applied to both the small-\( R \) jets and the track jets to identify those which are likely to have originated from a \( b \)-quark. The DL1r algorithm is used, at a working point chosen to have 77% efficiency on average for jets associated with true \( b \)-hadrons in simulated \( t\bar{t} \) events. This is a multivariate algorithm which uses a selection of inputs including information about the impact parameters of ID tracks, the presence of displaced secondary vertices, and the reconstructed flight paths of \( b \)- and \( c \)-hadrons inside the jet [71]. At the chosen working point, the light-jet (charm-jet) rejection measured in \( t\bar{t} \) MC simulation is about a factor of 130 (4.9) on average for small-\( R \) jets. The training and calibration of this algorithm is performed separately for each jet type [72,73]. Correction factors are applied to the simulated samples to compensate for differences between the \( b \)-tagging efficiencies in data and simulation.
Muons are reconstructed by matching ID tracks with either MS tracks or aligned individual hits in the MS and performing a combined track fit. They are required to have $p_T > 4$ GeV and $|\eta| < 2.5$, and to satisfy "Medium" identification criteria based on track-quality variables [74]. Muons are used only to apply corrections to jet momenta.

A momentum correction is applied to $b$-tagged small-$R$ jets to account for energy lost to soft out-of-cone radiation and to muons and neutrinos in semileptonic $b$-hadron decays. This correction follows the procedure used in Ref. [75] and consists of two parts. For the first, if any muon is within $\Delta R = 0.4$ of a $b$-tagged small-$R$ jet, the four-momentum of the muon is added to that of the jet. Any energy deposited in the calorimeter by the muon is then subtracted from the jet to prevent double counting; this is computed according to the description in Ref. [76]. In the second step a global scale factor is applied to each $b$-tagged small-$R$ jet based on its $p_T$ and whether or not it has a muon within $\Delta R = 0.4$ of the jet axis. These scale factors are derived from simulation.

To account for energy lost in semileptonic $b$-hadron decays, a similar muon-in-jet correction is applied to large-$R$ jets. If a muon is matched within a distance of $\Delta R = \min (0.4, 0.04 + 10 \text{ GeV}/p_T^{\text{muon}})$ to one of the two leading track jets associated with the large-$R$ jet, and if the track jet is $b$-tagged, the muon is considered part of the large-$R$ jet. Again, any energy deposited in the calorimeter by the muon is subtracted from the jet to prevent double counting. The muon four-momentum is then added to the calorimeter-based component of the large-$R$ jet four-momentum, and the jet mass is recalculated [77].

V. RESOLVED CHANNEL

A. Event selection

To be considered for analysis, events must pass the trigger requirements specified in Sec. III A. To simplify the modeling of trigger efficiencies, events are sorted into exactly one of four classes based on offline kinematic quantities, each of which requires one specific type of trigger to be passed. In decreasing order of priority, these are as follows:

(i) If the leading jet has $p_T > 325$ GeV and is $b$-tagged, the trigger requiring one high-$E_T$ $b$-jet is used.

(ii) If the leading jet has $p_T > 170$ GeV and is not $b$-tagged, the trigger requiring two $b$-jets and one additional jet is used.

(iii) If the $H_T$ in the event (computed using all jets with $p_T > 25$ GeV and $|\eta| < 2.5$) is greater than 900 GeV, the trigger requiring a high $H_T$ is used.

(iv) For all remaining events, the trigger requiring two $b$-jets and two additional jets is used.

The definitions of these classes are the result of a dedicated sensitivity optimization intended to minimize the expected limits on the signal cross section. Following the trigger selection, events are required to have at least four small-$R$ jets. Events are then divided into two categories, “2b” (where exactly two jets are $b$-tagged) and “4b” (where at least four jets are $b$-tagged). Exactly four jets are selected to construct the two $H$ candidates. For 4b events, the four $b$-tagged jets with the highest $p_T$ are selected. For 2b events, the two $b$-tagged jets and the two untagged jets with the highest $p_T$ are selected. The 2b events are needed to construct the background model for the 4b category. This selection of untagged jets can introduce a kinematic bias with respect to the 4b category, but this is accounted for by the reweighting function described in Sec. V B.

After the four jets are chosen, there are three possible combinations for pairing them into $H$ candidates. For a given pairing, the four-momentum of the $H$ candidate is defined as the sum of the four-momenta of the jets used to construct it. The pairing is chosen using a boosted decision tree (BDT). This is trained using LightGBM [78] to classify each of the three possible pairings in a signal event as either correct or incorrect. The correct pairing is defined by using the generator’s decay record to match jets to the parton-level $b$-quarks which result immediately from the $H$ decay. The classifier assigns each of the three candidate pairings a score, and the pairing with the highest score is chosen. The input variables to the BDT are the separations in pseudorapidity, azimuthal angle, and their quadratic sum ($\Delta \eta$, $\Delta \phi$, and $\Delta R$, respectively) between the two jets in each pair. Although the same information is contained in $\Delta \eta$ and $\Delta \phi$ alone, additional use of a precalculated $\Delta R$ leads to improved performance. The BDT is also parametrized in the invariant mass of the four identified $b$-jets [$m(4b)$], which is included as an additional input feature. However, $m(4b)$ cannot itself be used to discriminate between correct and incorrect pairings as it is independent of the pairing. Its inclusion as a parameter serves the purpose of ensuring optimal performance for all resonance masses. The BDT is trained on a sample consisting of one quarter of the simulated spin-0 signal events, across the full range of resonance masses considered in the resolved channel. A further selection on training events is applied, requiring them to contain at least four jets with $p_T > 35$ GeV, each of which is within $\Delta R = 0.4$ of a true $b$-quark, defined at parton level and originating from a Higgs boson decay. This $p_T$ requirement is loosened relative to the nominal selection in order to increase the number of selected events with low $m(4b)$. The events entering this training are not used anywhere else in the analysis. This technique results in less background in the signal region (defined below) compared to the strategy used in the previous ATLAS search [7], which was based on minimizing the difference between the two dijet invariant masses. The BDT algorithm finds the correct pairing in at least 65% of signal events. This efficiency is lowest for low resonance masses, but reaches almost 100% for resonance masses of 600 GeV.
and higher. This is a significant improvement with respect to the efficiency of the previous method, especially at low resonance masses.

After the $H$ candidates are formed, they are ordered by the scalar sum of the $p_T$ of their constituent jets; $H_1$ and $H_2$ denote the leading and subleading $H$ candidates, respectively. A pseudorapidity separation between the two $H$ candidates of $|\Delta \eta_{HH}| < 1.5$ is required in order to reduce the multijet background. Additionally, a “top veto” is applied, to reduce the background from hadronic top quark decays. This is defined by combining every possible pair of jets with $p_T > 40$ GeV and $|\eta| < 2.5$, including those that were not selected for the $H$ candidates, to form “$W$ candidates”. “Top quark candidates” are built by pairing $W$ candidates with each remaining jet that was selected for $H$ candidates. Events are rejected if any top quark candidate satisfies $X_{Wt} < 1.5$, where the discriminant $X_{Wt}$ is defined as

$$X_{Wt} = \sqrt{ \left( \frac{m(W) - m_{SM}(W)}{0.1 \times m(W)} \right)^2 + \left( \frac{m(t) - m_{SM}(t)}{0.1 \times m(t)} \right)^2 }.$$ 

Here, $m(W)$ and $m(t)$ denote the masses of the candidates under consideration, while $m_{SM}(W)$ and $m_{SM}(t)$ denote the measured masses of these particles (80.4 GeV and 172.5 GeV, respectively [79]). The denominators in the expression for $X_{Wt}$ represent the approximate mass resolution of the detector.

Finally, events are sorted into three kinematic regions based on the invariant masses of the $H$ candidates. The first of these is the signal region (SR), defined by requiring $X_{HH} < 1.6$, where

$$X_{HH} = \sqrt{ \left( \frac{m(H_1) - 120 \text{ GeV}}{0.1 \times m(H_1)} \right)^2 + \left( \frac{m(H_2) - 110 \text{ GeV}}{0.1 \times m(H_2)} \right)^2 }.$$ 

The shape of the SR in the $m(H_1) - m(H_2)$ plane is chosen to optimize the signal significance. The mass values of 120 GeV and 110 GeV correspond to the position of the peaks of the simulated signal $m(H_1)$ and $m(H_2)$ distributions. The deviations from the measured Higgs boson mass of 125 GeV [79] are due to detector effects, as well as energy lost to neutrinos from the $b$-hadron decays and to out-of-cone radiation. Jets which lose energy give rise to lower mass in their $H$ candidate. Since these jets are more likely to compose $H_2$ by definition, this results in a slightly lower mass for $H_2$ than $H_1$ on average. The validation region (VR) contains the events not in the SR which satisfy the condition

$$R_{HH}^{VR} \equiv \sqrt{(m(H_1) - 1.03 \times 120 \text{ GeV})^2 + (m(H_2) - 1.03 \times 110 \text{ GeV})^2} < 30 \text{ GeV}. $$

Finally, the control region (CR) contains the events not in the SR or VR which satisfy the condition

$$R_{HH}^{CR} \equiv \sqrt{(m(H_1) - 1.05 \times 120 \text{ GeV})^2 + (m(H_2) - 1.05 \times 110 \text{ GeV})^2} < 45 \text{ GeV}. $$

The centers of the VR and CR are shifted relative to that of the SR to ensure that the mean $H$ candidate masses are equal in the three regions. The shapes of these regions in the $m(H_1) - m(H_2)$ plane are shown with the $2b$ data for the full 2016–2018 dataset. $H_1$ and $H_2$ are the reconstructed Higgs boson candidates, sorted by their $p_T$.

The invariant mass of the sum of the two resulting four-momenta. This procedure improves the scale and resolution of the reconstructed signal mass distribution by correcting for detector effects and physical processes such as radiative emission outside the jet cones. This correction improves the signal mass resolution by up to 25% and shifts the mean of the mass distribution closer to the true value. It also modifies the background shape, but
The efficiency at low resonance masses is mainly limited by the trigger. At high resonance masses the jets start to merge together and the reconstruction and $b$-tagging efficiencies decrease. The efficiency is substantially larger for the spin-2 model than for the spin-0 model because the corrected $m(HH)$ distribution of the spin-2 model is much broader. This has an especially large effect at the lowest resonance masses, where the shape of the corrected $m(HH)$ distribution is distorted toward higher values. This is a result of both the $m(HH)$ correction and the natural mass shape of the graviton resonance near the production threshold.

B. Background estimation

After the selection described above, the background is dominated by pure QCD multijet processes (excluding top-quark production), with the approximately 5% remainder almost entirely composed of $t\bar{t}$. This background composition was determined by comparing $t\bar{t}$ MC simulation to the total background estimate in the SR; it is purely meant to be indicative and is not used in the statistical treatment. The background is modeled using a purely data-driven technique; the only MC simulation used is for the signal.

The background shape in the $4b$ SR is estimated from data in the $2b$ regions and the $4b$ CR. The signal contamination in the $2b$ dataset is evaluated and found to be negligible compared to the background uncertainties. The signal-to-background ratio in the $4b$ CR depends on the signal hypothesis and ranges from 10% to 25% of that in the SR. However, the impact on the analysis results was studied and found negligible. This was determined by injecting various signals into the CR (and VR) data using cross sections corresponding to existing experimental upper limits.

The event kinematics in the $2b$ and $4b$ regions are not expected to be identical, so a reweighting function which maps the $2b$ kinematic distributions onto the $4b$ distributions is derived,

$$w(\vec{x}) = \frac{p_{4b}(\vec{x})}{p_{2b}(\vec{x})},$$

where $p_{2b}(\vec{x})$ and $p_{4b}(\vec{x})$ are the probability density functions for $2b$ and $4b$ data, respectively, over a set of kinematic variables $\vec{x}$. This function is derived in the CR and then applied to the $2b$ SR in order to produce a model of the background in the $4b$ SR. It can also be applied more generally to any $2b$ region to produce a background model for the corresponding $4b$ region.

The computation of $w(\vec{x})$ is a density ratio estimation problem, for which a variety of approaches exist. The method employed in this analysis is modified from Refs. [80,81] and makes use of an artificial neural network (NN). This NN is trained on $2b$ and $4b$ CR data to minimize the loss function,

$$\mathcal{L}(w(\vec{x})) = \int d\vec{x}[\sqrt{w(\vec{x})}p_{2b}(\vec{x}) + \frac{1}{\sqrt{w(\vec{x})}}p_{4b}(\vec{x})].$$

The function in Eq. (1) optimizes this loss by equalizing the contributions from the two terms.

The kinematic variables used to make up $\vec{x}$ are chosen to be sensitive to the differences between the $2b$ and $4b$ events. These are:

1. $\log(p_T^2)$ of the selected jet with the second-highest $p_T$. 

FIG. 3. Cumulative acceptance times efficiency as a function of resonance mass for each event selection step in the resolved channel for (a) the spin-0 and (b) the spin-2 signal models. The local maximum at 251 GeV is a consequence of the near-threshold kinematics.
Here, “selected” jets refer to the four jets which are used to construct the $H$ candidates. The variables to which the reweighting is most sensitive are the jet multiplicity, $\Delta R$ between the two $H$ candidates, and $\log(p_T)$ of the $HH$ system. The NN has three densely connected hidden layers of 50 nodes, each with a rectified linear unit activation function [82], and a single-node linear output.

The training of the NN is subject to variation due to initial conditions and the limited size of the training samples. To account for these effects, the bootstrap resampling technique is used [83]. This entails constructing a set of training samples by sampling with replacement from the original. The NN is trained independently on each element of this set, using different initial conditions each time. This results in an ensemble of background estimates. Since the original training sample is large, the resulting background estimate in each bin can be approximated as being Gaussian distributed. Additionally, this sampling-with-replacement procedure can be approximated by applying a randomly distributed integer weight to each event, drawn from a Poisson distribution with a mean of 1. Both of these approximations are used in order to reduce the computational complexity of the problem. These “bootstrap weights” are independent of $w(x)$, which reweights the nominal 2b kinematic distributions to the nominal 4b kinematic distributions. To increase the stability of the background estimate, the median value of $w(x)$ for each event is calculated across the ensemble and used as the nominal background estimate. This ensemble of weights is also used to evaluate the uncertainty due to the finite training sample, as detailed in Sec. VC.

The effect of applying this reweighting to the CR, where it is derived, is shown in Fig. 4. The output of this procedure is an estimate of the corrected $m(HH)$ distribution in the 4b SR, which is then used as input to the statistical procedure detailed in Sec. VII. The optimization of the bin width of the corrected $m(HH)$ distribution is based on the detector resolution at low masses. The performance of this reweighting outside of the region where it is derived is checked using the VR. The scaling of the 2b distribution to the 4b sample size is always derived from the two respective CRs, as part of the reweighting procedure. The data are found to be compatible with the respective CRs as the reweighting procedure. The data are found to be compatible with the background model in the VR, as shown in Fig. 5. Residual differences between the CR and VR are shown (a) before and (b) after the kinematic reweighting procedure. In both cases the 2b distributions are normalized to the 4b event yields for a pure shape comparison. The final bin of each distribution includes overflow. The bottom panel shows the difference between the 4b and 2b distributions, relative to the 2b distribution.
distributions are then scaled to have the same normalization as the nominal distribution, multiplied by the ratio of the upper boundary value of the IQR of $\alpha^{(j)}$ to its nominal value. These rescaled varied distributions form an envelope around the nominal one, which specifies the size of this uncertainty in each corrected $m(HH)$ bin. As this uncertainty is statistical in origin, it is uncorrelated across $m(HH)$ bins.

Uncertainties in the background estimate also arise from kinematic differences between the CR and the SR. To evaluate these effects, an alternative background model is derived in the VR instead of the CR. The difference between the corrected $m(HH)$ distributions from the nominal and alternative background models is used to estimate the uncertainty in the shape of the $m(HH)$ distribution in the SR. To allow sufficient flexibility in the model, this uncertainty is parametrized in terms of two components: low-$H_T$ and high-$H_T$, where $H_T$ now denotes the scalar sum of the $p_T$ of the four jets constituting the $H$ candidates. This variable is chosen because it is correlated with $m(HH)$, but does not introduce a discontinuity in the $m(HH)$ spectrum when the two components are varied separately. The boundary between low-$H_T$ and high-$H_T$ events is chosen to be 300 GeV. Each of these two components is symmetrized around the nominal shape to construct a two-sided uncertainty. These uncertainties are taken to be uncorrelated across the different years to accommodate differences due to the varying triggers and run conditions.

Several detector modeling uncertainties are evaluated and included. These affect only the signal description, as the background is estimated entirely from data. Uncertainties in the jet energy scale and resolution are treated according to the prescription in Ref. [61]. Uncertainties in the $b$-tagging efficiency are treated according to the prescription in Ref. [30]. Uncertainties in the trigger efficiencies are evaluated from measurements of per-jet online efficiencies for both jet reconstruction and $b$-tagging, which are used to compute event-level uncertainties. These are then applied to the simulated events as overall weight variations. The uncertainty in the total integrated luminosity used in this analysis is 1.7% [84], obtained using the LUCID-2 detector for the primary luminosity measurements [85].

Several sources of theoretical uncertainty affecting the signal models were considered and are described as follows. Uncertainties due to modeling of the parton shower and underlying event are evaluated by comparing results between two generators for these parts of the calculation: the nominal HERWIG7.1.3 and the alternative PYTHIA8.235. This is found to have a 5% effect on the signal acceptance and a negligible impact on the $m(HH)$ distribution, independently of the resonance mass. Uncertainties in the matrix element calculation are evaluated by raising and lowering the factorization and renormalization scales used in the generator by a factor of two.
both independently and simultaneously. This results in an effect smaller than 1% for all variations and all masses; the impact of such uncertainties is therefore neglected. PDF uncertainties are evaluated using the PDF4LHC_NLO_MC set [86] by calculating the signal acceptance for each replica and taking the standard deviation. In all cases, these result in a less than 1% uncertainty in the signal acceptance, and therefore these are also neglected. Theoretical uncertainties in the $H \rightarrow b\bar{b}$ branching ratio [29] are included; they amount to a 2.5% overall uncertainty on the signal normalization.

D. Results

The corrected $m(HH)$ distributions for data and the estimated background after the fit to data described in Sec. VII are shown in Fig. 6. The data agree well with the background prediction and no significant excess is observed. The event yields for data, background, and several signal hypotheses are presented in Tables I and II. These are integrated over windows in the corrected $m(HH)$ spectrum containing approximately 90% of the signal in each case. These windows are defined such that their first and last bins contain the 5th and 95th percentile of the distribution, respectively. This range is larger for the spin-2 signal because the benchmark model predicts a wider resonance. The statistical interpretation of the data is discussed in Sec. VII.

VI. BOOSTED CHANNEL

A. Event selection

After passing the trigger requirement, each event is required to contain at least two large-$R$ jets with

![Graph showing corrected $m(HH)$ distribution for data and the estimated background.](image)

**FIG. 6.** Corrected $m(HH)$ distribution in the resolved $4b$ signal region (dots), after the fit under the background-only hypothesis. The error bars on the $4b$ points represent the Poisson uncertainties corresponding to their event yields. The background model (teal histogram) is shown with its total post-fit uncertainty (gray band). The final bin includes overflow. Representative spin-0 signal hypotheses (dashed, dotted, and dashed-dotted lines) are overlaid, normalized to the overall expected limits on their cross sections. The bottom panel shows the difference between the $4b$ distribution and the background model, relative to the background model. No significant excess of data relative to the SM background is observed.

**TABLE I.** Resolved $4b$ signal region data, estimated background, and signal event yields in corrected $m(HH)$ windows containing roughly 90% of each signal, for representative spin-0 mass hypotheses. The signal is normalized to the overall expected limit on its cross section; its uncertainties are evaluated by adding all individual components in quadrature. The background yields and uncertainties are evaluated after a background-only fit to the data.

<table>
<thead>
<tr>
<th>$m(X)$ [GeV]</th>
<th>Corrected $m(HH)$ range [GeV]</th>
<th>Data</th>
<th>Background model</th>
<th>Spin-0 signal model</th>
</tr>
</thead>
<tbody>
<tr>
<td>260</td>
<td>[250, 321]</td>
<td>18554</td>
<td>18300 ± 110</td>
<td>503 ± 43</td>
</tr>
<tr>
<td>500</td>
<td>[464, 536]</td>
<td>2827</td>
<td>2866 ± 22</td>
<td>105.4 ± 5.7</td>
</tr>
<tr>
<td>800</td>
<td>[750, 850]</td>
<td>358</td>
<td>366.2 ± 7.3</td>
<td>37.7 ± 1.7</td>
</tr>
<tr>
<td>1200</td>
<td>[1079, 1250]</td>
<td>68</td>
<td>52.6 ± 1.7</td>
<td>11.71 ± 0.62</td>
</tr>
</tbody>
</table>

**TABLE II.** Resolved $4b$ signal region data, estimated background, and signal event yields in corrected $m(HH)$ windows containing roughly 90% of each signal, for representative spin-2 mass hypotheses. The signal is normalized to the overall expected limit on its cross section; its uncertainties are evaluated by adding all individual components in quadrature. The background yields and uncertainties are evaluated after a background-only fit to the data.

<table>
<thead>
<tr>
<th>$m(G_{KK})$ [GeV]</th>
<th>Corrected $m(HH)$ range [GeV]</th>
<th>Data</th>
<th>Background model</th>
<th>Spin-2 signal model</th>
</tr>
</thead>
<tbody>
<tr>
<td>260</td>
<td>[250, 393]</td>
<td>26775</td>
<td>26650 ± 130</td>
<td>368 ± 25</td>
</tr>
<tr>
<td>500</td>
<td>[464, 636]</td>
<td>4655</td>
<td>4719 ± 37</td>
<td>138.6 ± 5.7</td>
</tr>
<tr>
<td>800</td>
<td>[707, 950]</td>
<td>795</td>
<td>811 ± 13</td>
<td>52.1 ± 1.9</td>
</tr>
<tr>
<td>1200</td>
<td>[993, 1279]</td>
<td>146</td>
<td>120.6 ± 2.8</td>
<td>14.45 ± 0.67</td>
</tr>
</tbody>
</table>
\( p_T \) > 250 GeV. The two highest-\( p_T \) jets are selected as the \( H \) candidates. The leading and subleading \( H \) candidates ordered by \( p_T \) are denoted by \( H_1 \) and \( H_2 \), respectively. Each \( H \) candidate is required to have \(|\eta| < 2.0\), \( m(H) > 50 \text{ GeV} \), and at least one associated track jet. Additionally, at least one \( H \) candidate must have \( p_T > 450 \text{ GeV} \), which is driven by the trigger threshold. In order to further reduce the background, an additional requirement of \( |\Delta R_{HH}| < 1.3 \) is placed on the \( H \) candidates.

Events are then categorized according to the multiplicity and \( b \)-tagging status of the track jets associated with each of the two \( H \) candidates. For \( H \) candidates with more than two associated track-jets, only the two with the highest \( p_T \) are considered.

Since track-jets have a variable radius, it is possible for a high-\( p_T \) jet to be contained completely within the catchment area of another low-\( p_T \) jet. This can lead to a pathological case in which the low-\( p_T \) jet’s axis is also contained within the high-\( p_T \) jet. This can result in misassignment of tracks to jets by the \( b \)-tagging algorithm (which is based purely on proximity to the jet axis). To avoid any degradation in performance resulting from this, events containing such collinear track-jets are vetoed from the boosted channel.

Three signal-enriched categories with four or fewer \( b \)-tagged track-jets are defined:

(a) Events in the “4b” category have two \( b \)-tagged track-jets associated with each \( H \) candidate.

(b) Events in the “3b” category have two \( b \)-tagged track-jets associated with one \( H \) candidate and exactly one \( b \)-tagged track-jet associated with the other \( H \) candidate.

(c) Events in the “2b” category have exactly one \( b \)-tagged track-jet associated with each \( H \) candidate.

These are collectively labeled as the high-tag categories. Including events with less than four \( b \)-tagged track-jets increases the sensitivity of the search especially for high resonance masses, where, due to the large boost, track-jets can become so close that they are often not reconstructed individually.

Additional low-tag categories with track-jets that fail the \( b \)-tagging requirement are also defined in order to estimate the background.

(a) Events in the “2b–2f” category (for modeling 4b) have one \( H \) candidate with two or more associated \( b \)-tagged track-jets and the other \( H \) candidate with no \( b \)-tagged track-jets but two or more untagged track-jets.

(b) Events in the “2b–1f” category (for modeling 3b) have one \( H \) candidate with two or more associated \( b \)-tagged track-jets and the other \( H \) candidate with no \( b \)-tagged track jets but one or more untagged track jets. Events in the “1b–1f” category (for modeling 2b) have one \( H \) candidate with exactly one associated \( b \)-tagged track jet and the other \( H \) candidate with no \( b \)-tagged track jets but one or more untagged track jets. In these low-tag categories, the \( H \) candidate that has no \( b \)-tagged track jets is also referred to as untagged, while the other one is labeled as tagged. The untagged \( H \) candidate in the 2b–1f region is allowed to have more than one track-jet because requiring exactly one would result in a very small number of events in this category. A diagram of events in these high-tag and low-tag categories is shown in Fig. 7.

Events satisfying the 2b–2f criteria also necessarily satisfy the 2b–1f criteria. To avoid overlap between the two categories, these events are distributed randomly between them, with 80% allocated as 2b–1f events and the remaining 20% allocated as 2b–2f events. This corresponds roughly to the ratio of background events present in the two categories.

Similarly to the resolved channel, events are sorted into signal, validation, and control regions based on the invariant masses of the \( H \) candidates. The SR is defined by requiring \( X_{HH} < 1.6 \), where

\[
X_{HH} = \sqrt{\left(\frac{m(H_1) - 124 \text{ GeV}}{0.1 \times m(H_1)}\right)^2 + \left(\frac{m(H_2) - 115 \text{ GeV}}{0.1 \times m(H_2)}\right)^2}. 
\]

This definition, as well as those of the validation and control regions, slightly differs from that in the resolved
channel. This is due to the different energy scale of the boosted jet reconstruction and the different background distribution. The VR contains the events not in the SR which satisfy the condition

$$R^{VR}_{HH} = \sqrt{(m(H_1) - 124 \text{ GeV})^2 + (m(H_2) - 115 \text{ GeV})^2}$$

$$< 33 \text{ GeV}.$$ 

Finally, the CR contains the events not in the SR or VR which satisfy the condition

$$R^{CR}_{HH} = \sqrt{(m(H_1) - 134 \text{ GeV})^2 + (m(H_2) - 125 \text{ GeV})^2}$$

$$< 58 \text{ GeV}.$$ 

The CR is shifted to higher masses relative to the signal and validation regions in order to maximize the number of selected events while avoiding the low-mass peak of the multijet background distribution. The definition of these regions in the $m(H_1)-m(H_2)$ plane are shown with the $2b-1f$ data in Fig. 8.

In order to ensure orthogonality between the resolved and boosted channels, any events passing the resolved signal region selection are vetoed from the boosted channel. This priority choice results in the best signal sensitivity.

The signal acceptance times efficiency for various steps of the selection is shown in Fig. 9.

### B. Background estimation

As in the resolved channel, the background in the boosted channel is dominated by QCD-induced jet production, which is separated into multijet (light quark) and $t\bar{t}$ production. The fractions of $t\bar{t}$ relative to the total background are 10%, 15%, and 30% for the $4b$, $3b$, and $2b$ regions, respectively. Other background sources, such as single Higgs boson production, SM $HH$ production, $(Z \rightarrow b\bar{b})$ + jets, and $ZZ \rightarrow b\bar{b}b\bar{b}$ account for $\lesssim 1\%$ of the total and are neglected.

A data-driven method is used to estimate the multijet background in each of the $4b$, $3b$, and $2b$ signal regions. The $t\bar{t}$ background is estimated from MC simulation, with corrections derived from data applied in the $3b$ and $2b$ regions.

The overall normalization of the multijet and $t\bar{t}$ estimates are obtained from a fit to the CR data in each category. Two normalization parameters $\mu_{MJ}$ and $\alpha_{t\bar{t}}$ per $b$-tagging category (here denoted $n_b$) are introduced as follows:

$$N^{hi}_{i,\text{data}} = \mu^{n_{b}}_{MJ}(N^{lo}_{i,\text{data}} - N^{lo}_{i,\bar{t},i}) + \alpha^{n_{b}}_{t\bar{t}} N^{hi}_{i,\bar{t},i}.$$ 

Here, $N^{hi}_{i,\text{data}}$ and $N^{lo}_{i,\text{data}}$ denote the number of events in bin $i$ of the $H_1$ mass distribution in the high-tag and corresponding low-tag regions, respectively. The parameter $\mu_{MJ}$ scales the multijet background from the low-tag CR to the high-tag CR, and $\alpha_{t\bar{t}}$ corrects the MC estimate in the high-tag region. The values of $\mu_{MJ}$ and $\alpha_{t\bar{t}}$ are determined using a maximum-likelihood fit to the mass of the leading $H$ candidate after the kinematic reweighting is applied. The mass of the leading $H$ candidate discriminates between the two background processes, as shown in Fig. 10. The data in the control regions of the $4b$, $3b$, and $2b$ categories are fitted separately, and therefore a total of six floating normalization factors are obtained. However, the value of $\alpha_{t\bar{t}}$ in the $4b$ region is fixed to 1, since the fit is insensitive to the $t\bar{t}$ MC normalization with the available dataset and therefore cannot constrain it. The results of these normalization fits are summarized in Table III. It is assumed that these values of $\mu_{MJ}$ and $\alpha_{t\bar{t}}$ are also applicable in the VR and SR; potential deviations from this assumption are accounted for by systematic uncertainties. The fact that $\mu_{MJ} \ll 1$ implies that the background is much larger in the low-tag categories than in the high-tag categories. As a result, any potential bias in the high-tag background estimate due to signal contamination in the low-tag categories is much smaller than the signal contribution itself in the high-tag regions.

For $3b$ and $2b$, a kinematic reweighting procedure is applied to each corresponding low-tag category, analogous to the resolved channel. For the $4b$ category, no kinematic reweighting is applied. This is because the effect of mismodelings due to $b$-tagging is small compared to the size of statistical uncertainties in this category. Instead of an NN for constructing the reweighting function, an iterative spline method based on the one used in Ref. [7] is implemented here.

The difference between these low-tag and high-tag regions is that the low-tag events have an untagged $H$ candidate (no $b$-tagged track jets), while high-tag events instead have a tagged $H$ candidate (exactly one $b$-tagged track jet, since only the $3b$ and $2b$ categories are considered here). Therefore, the reweighting applied to low-tag events...
FIG. 9. Cumulative signal acceptance times efficiency as a function of the resonance mass for various selection steps in the boosted channel. The steps up to the $b$-tag categorization are shown for (a) the spin-0 and (b) the spin-2 signal models. The efficiencies of the three $b$-tag categories are shown for (c) the spin-0 and (d) the spin-2 scenarios; this efficiency is obtained after the other selection steps including the SR definition. The signal efficiency in the 4$b$ region has a maximum around 1.5 TeV. Above that value the track jets start to merge together, and for the highest resonance masses the 2$b$ category becomes the most efficient.

FIG. 10. Reconstructed mass distributions of the leading $H$ candidate for the data (dots) and the background model (stacked histograms) in the (a) 2$b$, (b) 3$b$, and (c) 4$b$ control regions. The error bars on the data points represent the Poisson uncertainties corresponding to their event yields. The statistical uncertainty of the background model is represented by the gray band. This distribution is used to normalize the multijet and $t\bar{t}$ background components. The enhanced event rates at low and high masses are due to the geometry of the CR. The bottom panel shows the difference between the data and the background model, normalized to the background model.
is based on their untagged $H$ candidates, with the aim of matching the kinematics of the tagged $H$ candidates in high-tag events. This reweighting is derived purely in the low-tag regions; the tagged $H$ candidates in the $1b-1f$ category are used to define the target.

The following kinematic distributions are used to construct the reweighting, for which leading and subleading candidates with two track-jets, the leading and subleading track jets have roughly equal probabilities to be the $b$-tagged one, so this random selection does not introduce a significant bias. Separate distributions are constructed for leading and subleading $H$ candidates, as well as for leading and subleading track jets.

At each iteration $i$, cubic splines are fitted to the ratios of tagged to untagged distributions, and the weights are updated according to

$$w_i(\vec{x}) = w_{i-1}(\vec{x}) \times \left[ \prod_{j} f_{ij}(x_j) - 1 \right] \times r_i + 1,$$

where $j$ indexes the different reweighting variables $x_j$, $f_{ij}$ denotes the spline functions, and the “learning rate” $r_i$ controls how much the weight can change with each iteration. This is set to $r_i = 1 - 0.5^i$, and ten iterations are used, after which convergence is observed. Suppressing the learning rate for early iterations is intended to avoid instabilities. This reweighting function is applied to the low-tag data sample which contains multijet and $t\bar{t}$ events. In order to obtain the multijet model, the $t\bar{t}$ contribution is subtracted from data, and for that purpose the $t\bar{t}$ events in the low-tag regions are therefore also reweighted. The $t\bar{t}$ distributions in the high-tag regions are not reweighted.

In order to reduce the effect of statistical fluctuations in the background at high $m(HH)$, the following function is fitted to the reweighted multijet and $t\bar{t}$ distributions for $m(HH) \geq 1200 \text{ GeV}$,

$$f(x) = e^{-p_0} \frac{1}{x^2} \left(1 - x\right)^{p_1-p_2 \ln x}, \quad (2)$$

where the $p_i$ are dimensionless free parameters and $x \equiv m(HH)/\sqrt{s}$. This function and similar ones have been used to fit falling dijet and multijet spectra in similar analyses (e.g., Refs. [7,87]). The results of these smoothing fits for the multijet background model are shown in Fig. 11. Due to the small number of events in the $4b$ category, the shape of the $t\bar{t}$ distribution in this region is taken from the

---

**TABLE III.** Best-fit values for $\mu_{MJ}$ and $\alpha_{\bar{t}t}$, with their statistical uncertainties. The linear correlation coefficient between the two parameters is also given. The value of $\alpha_{\bar{t}t}$ in the $4b$ region is fixed to 1, since the data are unable to constrain it significantly.

<table>
<thead>
<tr>
<th>Region</th>
<th>2b</th>
<th>3b</th>
<th>4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{MJ}$</td>
<td>0.05435 ± 0.00056</td>
<td>0.1204 ± 0.0023</td>
<td>0.0272 ± 0.0015</td>
</tr>
<tr>
<td>$\alpha_{\bar{t}t}$</td>
<td>0.863 ± 0.011</td>
<td>0.786 ± 0.042</td>
<td>1</td>
</tr>
<tr>
<td>Correlation</td>
<td>−0.74</td>
<td>−0.74</td>
<td>0</td>
</tr>
</tbody>
</table>

---

**FIG. 11.** Smoothing fit (solid line) applied to the $HH$ invariant mass spectrum of the multijet background estimate (dots with error bars) in the (a) $2b$, (b) $3b$, and (c) $4b$ signal regions. The effects of the fit function parameter eigenvariations (dashed and dashed-dotted lines) are also shown, indicating the effective statistical uncertainties. The bottom panel shows the difference between the unsmoothed and smoothed multijet background estimates, normalized to the smoothed estimate.
3b category and scaled to the yield in the 4b category. This is found to be consistent with the shape in the 4b category within statistical uncertainties.

The background model outside the region where it is derived is checked using the VR, as shown in Fig. 12. Good agreement between the background model and the data in the VR is observed; residual differences between the CR and VR are used to estimate a systematic uncertainty as described in Sec. V IC.

C. Systematic uncertainties

The boosted channel is generally limited by statistical uncertainties, especially for higher resonance masses. Systematic uncertainties in the background and hypothesized signals are nonetheless fully accounted for and described here.

The uncertainty in the data-driven background estimate is split into three normalization components and five shape components. Each of these components is evaluated separately for the t\bar{t} background (from simulation) and the multijet background.

The first normalization component is associated with the assumption that the normalization factor \( \mu_{\text{MJ}} \) derived in the CR is also applicable in the SR. To evaluate this uncertainty, a Gaussian process technique [88] is used to interpolate the multijet contribution in the SR. First, a fit is performed on the two-dimensional \( m(H_1)-m(H_2) \) distribution, with the SR data removed, to determine the parameters for a Gaussian two-point correlation function in each of the low-tag and high-tag regions. The correlation scales are optimized as part of the fit and are found to be >100 GeV in all cases, so the interpolation is able to smoothly fit across the SR. Second, the resulting correlation function is used to construct an estimate for all points in the \( m(H_1)-m(H_2) \) plane. Event yields in each region, in particular the SR, are determined from this result. The uncertainty is finally defined as the difference of high-tag to low-tag yield ratio between the SR and the CR. This is applied only to the multijet component of the background; the t\bar{t} estimate from MC is subtracted from the data when deriving it.

The second normalization component is associated with the choice of control region. The suitability of this choice is validated by comparing the background model with data in signal-depleted regions; however, a small dependence of the background estimate on this choice remains. This is accounted for by defining slightly different CRs and rederviving the background model from each of them. The largest difference between any of these and the nominal background model is taken as a normalization uncertainty. The alternative CRs are defined by raising or lowering the values of \( m(H_1) \) and \( m(H_2) \) by 3 GeV (independently, for four variations). Additional variations are defined by increasing or decreasing the value of the \( R_{\text{SR}}^{HH} \) requirement by 3 GeV while applying the opposite change to the \( R_{\text{VR}}^{HH} \) requirement, effectively making the CR thicker or thinner in the \( m(H_1)-m(H_2) \) plane. In all cases, a veto on the SR is maintained. This procedure results in uncertainties of 0.9%, 1.6%, and 6.0% for the 2b, 3b, and 4b regions, respectively.

The third component of the background normalization uncertainty is associated with the statistical uncertainty of the normalization fit. The best-fit values of \( \mu_{\text{MJ}} \) and \( \alpha_{\text{t\bar{t}}} \) are varied along the eigenvectors of their covariance matrix. The varied values are propagated through the background estimation procedure to evaluate the resulting effect. Since \( \alpha_{\text{t\bar{t}}} \) is fixed to a value of 1 for the 4b category, only one variation is performed in that case, corresponding to \( \mu_{\text{MJ}} \). This variation changes the normalizations of the individual background components, resulting in different total background shapes.
The first component of the background shape uncertainty comes from the limited sample size used to fit the smoothing function in Eq. (2). This is taken into account by an eigenvariation method using the covariance matrix of the function parameters. Function variations are defined by varying the best-fit parameters according to the eigenvectors, scaled to the square root of the corresponding eigenvalues. These varied functions are then treated as components of the uncertainty in the background shape. Three variations are used for each of the multijet and $t\bar{t}$ background components.

Two further components of the shape uncertainty in the background model are associated with choosing the fit function that is used in Eq. (2). As the function and fit range are arbitrary and chosen based only on empirical results from the CR and VR, both are varied. Seven alternative functional forms are used, based on the scheme from Ref. [87]. The nominal fit is compared with the other seven, and those which differ the most from the nominal one in each direction are used to define an envelope for the shape uncertainty. The result of this procedure is shown in Fig. 11. The fit range is also varied by changing its upper and lower bounds by 100 GeV in each direction. Again, the results which differ the most from the nominal one are used to construct an envelope, defining another shape uncertainty. In cases where all variations fall on one side of the nominal result, an envelope is constructed by symmetrization.

The fourth component of the shape uncertainty is the “residual” one, associated with shape differences between the CR and SR. This is associated only with the multijet background, as direct simulation of the SR is used for the $t\bar{t}$ background shape. This uncertainty is evaluated by comparing the multijet background model with the data in the VR, after subtracting the $t\bar{t}$ component determined from simulation. Since the events in the VR are kinematically closer to the SR than events in the CR are, this comparison of the background models covers residual mismodeling connected to the extrapolation over the regions. The multijet models agree within statistical uncertainties for the 3$b$ and 4$b$ regions, so this uncertainty is only derived for 2$b$ events. The shape of the uncertainty is defined as the ratio of the multijet background model to the $t\bar{t}$-subtracted VR data. Empty bins and bins with relative statistical uncertainty over 50% are discarded to suppress unphysical effects from statistical fluctuations, and the remainder are smoothed and symmetrized to obtain the final shape.

The fifth component of the shape uncertainty is associated with the nonclosure of the background estimation method itself. This is evaluated by running the full background estimation procedure on multijet MC simulation. The reweighting is derived in the CR and applied to the low-tag events in both the VR and SR, which are summed into a single region for the purpose of evaluating this uncertainty. The ratio of the resulting estimate to the actual MC prediction in that summed region is determined as a function of $m(HH)$. A linear function is fitted to this set of ratios in each high-tag category, in order to mitigate the effects of statistical fluctuations. The result is then symmetrized to form an envelope, which defines this component of the shape uncertainty. It is applied only to the multijet component of the background. For 2$b$ and 3$b$ categories, the relative size of this uncertainty is 5% or less for $m(HH) < 2$ TeV, ranging up to about 15% for the largest values of $m(HH)$. For the 4$b$ category, the relative size of this uncertainty is less than 15% over the whole $m(HH)$ range.

Several experimental uncertainties are considered. These affect only the signal and $t\bar{t}$ background models directly, as the multijet background is estimated from data alone. However, they have indirect effects on the multijet background model due to the $t\bar{t}$ subtraction applied in the data-driven procedure. These are accounted for by propagating the corresponding variations through the background estimation procedure. Uncertainties in the jet energy and mass scales and resolutions are treated according to the prescriptions in Refs. [89,90]. Uncertainties in the $b$-tagging efficiency are treated according to the prescription in Ref. [30]; this follows exactly the same procedure as in the resolved channel. Unlike in the resolved channel, no trigger efficiency uncertainties are required; no $b$-tagging is done at trigger level and the jet $p_T$ requirement is high enough that the triggers are fully efficient. As in the resolved channel, the uncertainty in the total integrated luminosity is 1.7% [84], obtained using the LUCID-2 detector [85].

Theory uncertainties in the $t\bar{t}$ background model are evaluated for the matrix element and parton shower parts of the calculation. Matrix element uncertainties are computed by raising and lowering the factorization and renormalization scales by a factor of two. The Powheg damping parameter $h_{\text{damp}}$ is also varied upwards by a factor of two. Additionally, a comparison with an alternative matrix element generator, MadGraph5_aMC@NLO2.6.0, is performed and the difference is taken as an uncertainty component. Uncertainties in the initial-state and final-state radiation modeling are evaluated using both scale variations and eigenvariations of the A14 tune [50]. Parton shower uncertainties are obtained by generating alternative samples which are showered using HERWIG7.1.3 instead of PYTHIA8.230 and taking the difference between the resulting $m(HH)$ distributions. The effects of PDF uncertainties are evaluated by comparing the nominal $m(HH)$ distribution with those obtained from a set of 100 weight variations, but are found to be much smaller than the statistical uncertainty and are therefore neglected.

Theory uncertainties in the signal models are evaluated using exactly the same method as for the resolved channel. Here, the parton shower and underlying event are found to have a 10% effect on the signal acceptance, independently
of the resonance mass. Uncertainties in the matrix element calculation are again found to be negligible, and are not included in the statistical interpretation.

D. Results

The $m(HH)$ distributions for data and the estimated background are shown in Fig. 13 for the three categories. The data agree well with the background and no significant excess is observed. The numbers of events for data, background, and several signal hypotheses are presented in Tables IV and V. These event yields are integrated over a set of $m(HH)$ bins containing approximately 90% of the signal in each case. These windows are defined such that their first and last bins contain the 5th and 95th percentile of the distribution, respectively. This range is larger for the spin-2 signal because the benchmark model predicts a wider
TABLE IV. Boosted signal region data, estimated background, and signal event yields in \(m(HH)\) windows containing roughly 90\% of each signal, for representative spin-0 mass hypotheses. The signal is normalized to the overall expected limit on its cross section; its uncertainties are evaluated by adding all individual components in quadrature. The background yields and uncertainties are evaluated after a background-only fit to the data. The 4\(b\) category is not used for \(m(X) > 3\) TeV and the 2\(b\) category is not used for \(m(X) < 2\) TeV.

<table>
<thead>
<tr>
<th>(m(X)) [GeV]</th>
<th>Category</th>
<th>(m(HH)) range [GeV]</th>
<th>Data</th>
<th>Background model</th>
<th>Spin-0 signal model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>3(b)</td>
<td>[900, 1200]</td>
<td>1076</td>
<td>1109 ± 28</td>
<td>11.2 ± 2.0</td>
</tr>
<tr>
<td></td>
<td>4(b)</td>
<td>[900, 1200]</td>
<td>62</td>
<td>58.4 ± 5.8</td>
<td>3.85 ± 0.67</td>
</tr>
<tr>
<td>2000</td>
<td>2(b)</td>
<td>[1800, 2200]</td>
<td>77</td>
<td>77.8 ± 3.5</td>
<td>2.69 ± 0.92</td>
</tr>
<tr>
<td></td>
<td>3(b)</td>
<td>[1800, 2200]</td>
<td>14</td>
<td>11.47 ± 0.76</td>
<td>5.2 ± 1.6</td>
</tr>
<tr>
<td></td>
<td>4(b)</td>
<td>[1800, 2200]</td>
<td>2</td>
<td>0.76 ± 0.11</td>
<td>2.5 ± 1.1</td>
</tr>
<tr>
<td>3000</td>
<td>2(b)</td>
<td>[2600, 3200]</td>
<td>7</td>
<td>5.78 ± 0.45</td>
<td>2.01 ± 0.84</td>
</tr>
<tr>
<td></td>
<td>3(b)</td>
<td>[2600, 3200]</td>
<td>0</td>
<td>0.61 ± 0.16</td>
<td>2.37 ± 0.76</td>
</tr>
<tr>
<td></td>
<td>4(b)</td>
<td>[2600, 3200]</td>
<td>0</td>
<td>0.042 ± 0.055</td>
<td>0.72 ± 0.35</td>
</tr>
<tr>
<td>5000</td>
<td>2(b)</td>
<td>[4100, 5200]</td>
<td>0</td>
<td>0.65 ± 0.23</td>
<td>2.5 ± 1.1</td>
</tr>
<tr>
<td></td>
<td>3(b)</td>
<td>[4200, 5200]</td>
<td>0</td>
<td>0.012 ± 0.034</td>
<td>1.09 ± 0.29</td>
</tr>
</tbody>
</table>

TABLE V. Boosted signal region data, estimated background, and signal event yields in \(m(HH)\) windows containing roughly 90\% of each signal, for representative spin-2 mass hypotheses. The signal is normalized to the overall expected limit on its cross section; its uncertainties are evaluated by adding all individual components in quadrature. The background yields and uncertainties are evaluated after a background-only fit to the data. The 4\(b\) category is not used for \(m(G'_{KK}) > 3\) TeV and the 2\(b\) category is not used for \(m(G'_{KK}) < 2\) TeV.

<table>
<thead>
<tr>
<th>(m(G'_{KK})) [GeV]</th>
<th>Category</th>
<th>(m(HH)) range [GeV]</th>
<th>Data</th>
<th>Background model</th>
<th>Spin-2 signal model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>3(b)</td>
<td>[900, 1400]</td>
<td>1282</td>
<td>1298 ± 29</td>
<td>12.0 ± 2.1</td>
</tr>
<tr>
<td></td>
<td>4(b)</td>
<td>[900, 1500]</td>
<td>72</td>
<td>71.0 ± 6.9</td>
<td>4.14 ± 0.80</td>
</tr>
<tr>
<td>2000</td>
<td>2(b)</td>
<td>[1600, 2400]</td>
<td>223</td>
<td>221.0 ± 8.4</td>
<td>3.7 ± 1.2</td>
</tr>
<tr>
<td></td>
<td>3(b)</td>
<td>[1600, 2300]</td>
<td>38</td>
<td>33.5 ± 1.7</td>
<td>7.1 ± 2.1</td>
</tr>
<tr>
<td></td>
<td>4(b)</td>
<td>[1600, 2300]</td>
<td>3</td>
<td>2.15 ± 0.26</td>
<td>3.2 ± 1.3</td>
</tr>
<tr>
<td>3000</td>
<td>2(b)</td>
<td>[2300, 3400]</td>
<td>19</td>
<td>16.5 ± 1.2</td>
<td>2.44 ± 0.93</td>
</tr>
<tr>
<td></td>
<td>3(b)</td>
<td>[2200, 3300]</td>
<td>3</td>
<td>2.81 ± 0.42</td>
<td>2.79 ± 0.78</td>
</tr>
<tr>
<td></td>
<td>4(b)</td>
<td>[1900, 3300]</td>
<td>1</td>
<td>0.63 ± 0.17</td>
<td>0.82 ± 0.39</td>
</tr>
<tr>
<td>5000</td>
<td>2(b)</td>
<td>[2400, 5300]</td>
<td>18</td>
<td>13.6 ± 1.1</td>
<td>3.5 ± 1.4</td>
</tr>
<tr>
<td></td>
<td>3(b)</td>
<td>[1600, 5200]</td>
<td>41</td>
<td>35.5 ± 1.9</td>
<td>2.06 ± 0.54</td>
</tr>
</tbody>
</table>

resonance. The statistical interpretation of the data is discussed in Sec. VII.

VII. COMBINED RESULTS AND INTERPRETATIONS

For each signal model, the hypothesis of the presence of a signal is tested using a profile likelihood ratio [91]. The likelihood fit is carried out in bins of corrected \(m(HH)\) for the resolved channel and in bins of \(m(HH)\) for the boosted channel. In the resolved channel, data from 2016, 2017, and 2018 are included separately in a simultaneous fit. In the boosted channel, data from the 2\(b\), 3\(b\), and 4\(b\) signal regions are included separately in a simultaneous fit. The relative contribution of each \(b\)-tagging category to the sensitivity varies significantly with the resonance mass. At very high masses, the track-jets associated with one Higgs boson candidate tend to be highly boosted and are often not reconstructed as individual jets. Therefore, the 4\(b\) category is used only for signal hypotheses with \(m(X/G'_{KK}) \leq 3\) TeV, and the 2\(b\) category is used only for signal hypotheses with \(m(X/G'_{KK}) \geq 2\) TeV. For resonance masses in the range 900 GeV–1.5 TeV, the resolved and boosted channels are fitted simultaneously.

The likelihood function used to construct the test statistic has a standard form, consisting of a product of Poisson distributions for the yields in each bin and constraint functions for nuisance parameters describing systematic uncertainties. For uncertainties due to the limited sample size in data or MC simulation, the constraint is a Poisson distribution. For all other systematic uncertainties, the constraint is a Gaussian distribution. Any systematic uncertainty which is treated as uncorrelated between different regions or bins has a separate independent nuisance parameter for each of them. Uncertainties in the luminosity and signal modeling are treated as fully correlated between the resolved and boosted channels. All other uncertainties in the background model are treated as uncorrelated between the resolved and boosted channels. The statistical model is implemented using HistFactory [92].
The global significance is evaluated according to the procedure detailed in Ref. [93]. Pseudoexperiments are generated from the background-only model that was fitted to data, and used to construct a local $p$-value distribution as a function of the resonance mass. The number of level crossings below a reference level of $p = 0.5$ is used together with the local $p$-value to compute a global $p$-value. The most significant excess is found for a signal mass of 1100 GeV. The local significance of this excess is 2.3σ for the spin-0 signal model and 2.5σ for the spin-2 signal model. Its global significance is 0.4σ for the spin-0 signal model and 0.8σ for the spin-2 signal model.

Upper limits on the cross section of resonant Higgs boson pair production via gluon–gluon fusion ($\sigma_{ggF}$) are set in each of the benchmark models. These are based on the CL(s) method [94], where a cross-section value is considered excluded at the 95% confidence level (C.L.) when $CL(s)$ is less than 0.05. For signal masses up to 3 TeV, the limits are computed using asymptotic formulae [91]. At higher masses, the asymptotic approximation is inaccurate, so the limits are instead computed by sampling pseudoexperiments. The results are shown in Fig. 14. The theoretical prediction for the bulk RS model with $k/M_{Pl} = 1$ [28] (solid red line) is shown; the decrease below 350 GeV is due to a sharp reduction in the $G_{KK} \rightarrow HH$ branching ratio. The nominal $H \rightarrow b\bar{b}$ branching ratio is taken as 0.582.

The impacts of the main systematic uncertainties on the expected 95% C.L. upper limits on the signal cross section for four illustrative values of $m(X)$. These are defined as the relative decrease in the expected limit when each relevant nuisance parameter is held fixed to its best-fit value instead of being assigned an uncertainty. The spin-0 signal model is used here.

<table>
<thead>
<tr>
<th>Uncertainty category</th>
<th>280 GeV</th>
<th>600 GeV</th>
<th>1600 GeV</th>
<th>4000 GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background ($m(HH)$ shape)</td>
<td>12.5</td>
<td>8.7</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Jet momentum/ mass scale</td>
<td>0.6</td>
<td>0.1</td>
<td>1.2</td>
<td>1.7</td>
</tr>
<tr>
<td>Jet momentum/ mass resolution</td>
<td>2.1</td>
<td>1.5</td>
<td>7.1</td>
<td>7.8</td>
</tr>
<tr>
<td>$b$-tagging calibration</td>
<td>0.7</td>
<td>0.4</td>
<td>2.1</td>
<td>7.0</td>
</tr>
<tr>
<td>Theory (signal)</td>
<td>0.6</td>
<td>0.6</td>
<td>1.4</td>
<td>1.2</td>
</tr>
<tr>
<td>Theory ($t\bar{t}$ background)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>All systematic uncertainties</td>
<td>15.9</td>
<td>10.9</td>
<td>13.4</td>
<td>15.6</td>
</tr>
</tbody>
</table>

FIG. 14. Expected (dashed black lines) and observed (solid black lines) 95% C.L. upper limits on the cross section of resonant $HH$ production in the (a) spin-0 and (b) spin-2 signal models. The $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty ranges for the expected limits (colored bands) are shown. Expected limits using each of the resolved and boosted channels individually (dashed colored lines) are shown. The theoretical prediction for the bulk RS model with $k/M_{Pl} = 1$ [28] (solid red line) is shown; the decrease below 350 GeV is due to a sharp reduction in the $G_{KK} \rightarrow HH$ branching ratio. The nominal $H \rightarrow b\bar{b}$ branching ratio is taken as 0.582.
“systematic” for the purposes of this evaluation, and no corresponding fit parameters are held fixed. For all signal mass hypotheses, statistical uncertainties are dominant. At low masses, uncertainties in the shape of the background $m(HH)$ distribution from the data-driven estimate also contribute significantly. These uncertainties get considerably constrained (typically by a factor of 2–3) by the fit. Detector and theoretical uncertainties have only a very small impact on the sensitivity of the search.

VIII. CONCLUSION

A search for resonant pair production of Higgs bosons in the $bb\bar{b}$ final state was carried out using up to 139 fb$^{-1}$ of LHC $pp$ collision data collected by the ATLAS detector at $\sqrt{s} = 13$ TeV. Results are reported for the resolved channel, where each $\bar{b}$ pair is reconstructed as two separate small-$R$ jets, and the boosted channel, where each $\bar{b}$ pair is reconstructed as a single large-$R$ jet. The sensitivity of this analysis is improved relative to previous searches by using more sophisticated background modeling techniques, machine-learning methods, and variable-radius track-jets with optimized $b$-tagging in addition to the full ATLAS Run 2 dataset. The expected upper limits on the cross section are reduced relative to the previous ATLAS search in this final state by approximately 20% at low resonance masses and more than 80% at high masses. This search also covers resonance masses in the range from 3 TeV to 5 TeV for the first time.

No significant evidence of a signal is observed. Upper limits are set on the cross section of resonant Higgs boson pair production for two benchmark models: a generic narrow spin-0 resonance, and a spin-2 graviton in the context of a bulk Randall–Sundrum model with $k/M_{Pl} = 1$, both of which are assumed to be produced via gluon-gluon fusion. The bulk Randall–Sundrum model is excluded for graviton masses between 298 GeV and 1460 GeV.

ACKNOWLEDGMENTS

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; YerPhl, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; STFC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MEYS CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRI, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MEiN, Poland; FCT, Portugal; MNE/IFA, Romania; JINR; MES of Russia and NRC KI, Russian Federation; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZS, Slovenia; DSI/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; COST, ERC, ERDF, Horizon 2020 and Marie Sklodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; Norwegian Financial Mechanism 2014-2021, Norway; NCN and NAWA, Poland; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [95].


[9] ATLAS Collaboration, Search for Resonant and Nonresonant Higgs Boson Pair Production in the $b\bar{b}r\tau\tau^* \rightarrow b\bar{b}r\tau\tau^*$ Decay Channel in $pp$ Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector, Phys. Rev. Lett. 121, 191801 (2018); 122, 089901 (E) (2019).


SEARCH FOR RESONANT PAIR PRODUCTION OF HIGGS …

PHYS. REV. D 105, 092002 (2022)

53a Dipartimento di Fisica, Università di Genova, Genova, Italy
53b INFN Sezione di Genova, Italy
54 II. Physikalisches Institut, Justus-Liebig-Universität Gießen, Gießen, Germany
55 SUPA—School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
56 LPSC, Université Grenoble Alpes, CNRS/IN2P3, Grenoble, France
57 Laboratory for Particle Physics and Cosmology, Harvard University, Cambridge Massachusetts, USA
58a Department of Modern Physics and State Key Laboratory of Particle Detection and Electronics, University of Science and Technology of China, Hefei, China
58b Institute of Frontier and Interdisciplinary Science and Key Laboratory of Particle Physics and Particle Irradiation (MOE), Shandong University, Qingdao, China
58c School of Physics and Astronomy, Shanghai Jiao Tong University, Key Laboratory for Particle Astrophysics and Cosmology (MOE), SKLPPC, Shanghai, China
58d Tsung-Dao Lee Institute, Shanghai, China
59 Kirchhoff-Institut für Physik, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany
59a Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany
60a Department of Physics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
60b Department of Physics, University of Hong Kong, Hong Kong, China
60c Department of Physics and Institute for Advanced Study, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
60d ICTP, Trieste, Italy
61a Dipartimento Politecnico di Ingegneria e Architettura, Università di Udine, Sezione di Trieste, Udine, Italy
61b INFN Sezione di Trieste, Udine, Italy
62 JICLab, Université Paris-Saclay, CNRS/IN2P3, 91405, Orsay, France
63 Department of Physics, Indiana University, Bloomington Indiana, USA
64a INFN Gruppo Collegato di Udine, Sezione di Trieste, Udine, Italy
64b ICTP, Trieste, Italy
64c INFN Sezione di Lecce, Italy
65 Dipartimento di Matematica e Fisica, Università del Salento, Lecce, Italy
65a INFN Sezione di Lecce, Italy
66a Dipartimento di Fisica, Università di Milano, Milano, Italy
66b INFN Sezione di Milano, Italy
67a Dipartimento di Fisica, Università di Napoli, Napoli, Italy
67b INFN Sezione di Napoli, Italy
68a Dipartimento di Fisica, Università di Pavia, Pavia, Italy
68b INFN Sezione di Pavia, Italy
69a Dipartimento di Fisica E. Fermi, Università di Pisa, Pisa, Italy
69b INFN Sezione di Roma, Italy
70a Dipartimento di Fisica, Sapienza Università di Roma, Roma, Italy
70b INFN Sezione di Roma Tre, Roma, Italy
71a INFN Sezione di Roma Tor Vergata, Roma, Italy
71b Dipartimento di Fisica, Università di Roma Tor Vergata, Roma, Italy
72a INFN Sezione di Roma Tre, Rome, Italy
72b Dipartimento di Matematica e Fisica, Università Roma Tre, Roma, Italy
73a INFN-TIFPA, Italy
73b Università degli Studi di Trento, Trento, Italy
74 Institut für Astro- und Teilchenphysik, Leopold-Franzens-Universität, Innsbruck, Austria
75 University of Iowa, Iowa City Iowa, USA
76 Department of Physics and Astronomy, Iowa State University, Ames Iowa, USA
77 Joint Institute for Nuclear Research, Dubna, Russia
78a Departamento de Engenharia Elétrica, Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, Brazil
78b Universidade Federal do Rio De Janeiro COPPE/EE/IF, Rio de Janeiro, Brazil
79a Instituto de Física, Universidade de São Paulo, São Paulo, Brazil
79b KEK, High Energy Accelerator Research Organization, Tsukuba, Japan
80 Graduate School of Science, Kobe University, Kobe, Japan
81a AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Krakow, Poland
81b Marian Smoluchowski Institute of Physics, Jagiellonian University, Krakow, Poland
82 Institute of Nuclear Physics Polish Academy of Sciences, Krakow, Poland
83 Faculty of Science, Kyoto University, Kyoto, Japan
84 Kyoto University of Education, Kyoto, Japan

092002-33
SEARCH FOR RESONANT PAIR PRODUCTION OF HIGGS ... PHYS. REV. D 105, 092002 (2022)
Deceased.

Also at Department of Physics, King’s College London, London, United Kingdom.

Also at Istanbul University, Dept. of Physics, Istanbul, Turkey.

Also at Instituto de Fisica Teorica, IFT-UAM/CSIC, Madrid, Spain.

Also at TRIUMF, Vancouver BC, Canada.

Also at Physics Department, An-Najah National University, Nablus, Palestinian Authority.

Also at Department of Physics, University of Fribourg, Fribourg, Switzerland.

Also at Department of Physics and Astronomy, University of Louisville, Louisville, Kentucky, USA.

Also at Department of Physics, California State University, Fresno, USA.

Also at Department of Financial and Management Engineering, University of the Aegean, Chios, Greece.

Also at Centro Studi e Ricerche Enrico Fermi, Italy.

Also at Institute of Particle Physics (IPP), Canada.

Also at Bruno Kessler Foundation, Trento, Italy.

Also at Department of Physics, St. Petersburg State Polytechnical University, St. Petersburg, Russia.

Also at Borough of Manhattan Community College, City University of New York, New York New York, USA.

Also at Department of Physics, California State University, East Bay, USA.

Also at Institucio Catalana de Recerca i Estudis Avancats, ICREA, Barcelona, Spain.

Also at Graduate School of Science, Osaka University, Osaka, Japan.

Also at Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany.

Also at University of Chinese Academy of Sciences (UCAS), Beijing, China.

Also at Institute of Physics, Azerbaijan Academy of Sciences, Baku, Azerbaijan.

Also at Yeditepe University, Physics Department, Istanbul, Turkey.

Also at Institute of Theoretical Physics, Ilia State University, Tbilisi, Georgia.

Also at CERN, Geneva, Switzerland.

Also at Joint Institute for Nuclear Research, Dubna, Russia.

Also at Hellenic Open University, Patras, Greece.

Also at Center for High Energy Physics, Peking University, China.

Also at The City College of New York, New York New York, USA.

Also at Department of Physics, California State University, Sacramento, USA.

Also at Department de Physique Nucléaire et Corpusculaire, Université de Genève, Genève, Switzerland.

Also at Faculty of Physics, M.V. Lomonosov Moscow State University, Moscow, Russia.

Also at Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany.

Also at CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France.

Also at National Research Nuclear University MEPhI, Moscow, Russia.

Also at Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary.

Also at Giresun University, Faculty of Engineering, Giresun, Turkey.

Also at Department of Physics and Astronomy, Michigan State University, East Lansing Michigan, USA.