Properties of $g \to b\bar{b}$ at small opening angles in $pp$ collisions
with the ATLAS detector at $\sqrt{s} = 13$ TeV

M. Aaboud et al.¹
(ATLAS Collaboration)

(Received 24 December 2018; published 14 March 2019)

The fragmentation of high-energy gluons at small opening angles is largely unconstrained by present measurements. Gluon splitting to $b$-quark pairs is a unique probe into the properties of gluon fragmentation because identified $b$-tagged jets provide a proxy for the quark daughters of the initial gluon. In this study, key differential distributions related to the $g \to b\bar{b}$ process are measured using 33 $fb^{-1}$ of $\sqrt{s} = 13$ TeV $pp$ collision data recorded by the ATLAS experiment at the LHC in 2016. Jets constructed from charged-particle tracks, clustered with the anti-$k_t$ jet algorithm with radius parameter $R = 0.2$, are used to probe angular scales below the $R = 0.4$ jet radius. The observables are unfolded to particle level in order to facilitate direct comparisons with predictions from present and future simulations. Multiple significant differences are observed between the data and parton shower Monte Carlo predictions, providing input to improve these predictions of the main source of background events in analyses involving boosted Higgs bosons decaying into $b$-quarks.

DOI: 10.1103/PhysRevD.99.052004

I. INTRODUCTION

The search for highly Lorentz-boosted Higgs bosons produced by Standard Model (SM) processes [1] or by beyond-the-SM (BSM) processes [2–7] is of crucial importance at the Large Hadron Collider (LHC). As the branching ratio for the Higgs boson to decay into bottom quark pairs dominates the total decay rate, the boosted $H \to b\bar{b}$ channel can be the most sensitive to BSM effects entering at high Higgs boost. Algorithms for identifying jets resulting from bottom quark fragmentation are very powerful, so the main background for searches with boosted Higgs bosons contains $b$-quarks. The main contribution to this background is gluon splitting to $b\bar{b}$ pairs at small opening angles since the angle between the $b$-quarks in $H \to b\bar{b}$ scales with the Higgs boson mass ($m_H$) and momentum ($p_H$) as $m_H / p_H$. The $g \to b\bar{b}$ process also contributes to many other important SM measurements and searches by providing a source of additional real $b$-quark jets that can fake a signal for $b$-quarks originating from other processes (see e.g., Refs. [8–12]).

The modeling of $g \to b\bar{b}$ fragmentation is complex and provides a useful probe of quantum chromodynamics (QCD). The large mass of the $b$-quark introduces a significant modification to the massless QCD splitting functions by screening the soft-emission singularity. Trijet measurements from LEP [13–15] and SLD [16] provide valuable information about the rate of $g \to b\bar{b}$, but have not explored the differential properties of the fragmentation in the small opening-angle regime. Previous measurements that include the $b\bar{b}$ final state at the $Sp$S, Tevatron, and LHC using inclusive [17–36], multijet [37–39], and associated production [40–49] topologies have focused on well-separated quark pairs (dominated by fixed-order instead of parton-splitting effects) and were limited in their kinematic reach due in part to small datasets and low momentum transfers.

The high transverse momentum and low angular separation regime for $g \to b\bar{b}$ can be probed at the LHC using $b$-tagged small-radius jets within large-radius jets. This topology is used to calibrate $b$-tagging in dense environments [50–52] and is studied phenomenologically [53,54]. The measurement shown in this paper builds on these studies by using data collected by the ATLAS detector from $\sqrt{s} = 13$ TeV $pp$ collisions in order to perform a differential cross-section measurement of $g \to b\bar{b}$ inside jets at high transverse momentum—see Fig. 1 for a representative Feynman diagram. Small-radius jets built from charged-particle tracks are used as proxies for $b$-quarks and can be used as precision probes of the small opening-angle regime.

This paper is organized as follows. After a brief introduction to the ATLAS detector in Sec. II, the data and simulations used for the measurement are documented in Sec. III. Section IV describes the event selection and

¹Full 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.
pseudorapidity is defined in terms of the polar angle as \( \eta \). The first-level trigger is implemented in hardware and fast trigger chambers covering a system of precision tracking chambers covering magnets with eight coils each. The muon spectrometer has a spectrometer including three large superconducting toroidal meters with \( 8 \) coils each. Surrounding the calorimeters, the muon detector and the detector with a forward/backward-symmetric cylindrical geometry. The detector has a nearly \( 4\pi \) coverage in solid angle\(^1\) and consists of an inner tracking detector, electromagnetic and hadronic calorimeters, and a muon spectrometer. The inner detector (ID) is surrounded by a superconducting solenoid providing a \( 2 \text{T} \) magnetic field and covers a pseudorapidity range of \( \eta < 2.5 \). The ID is composed of silicon pixel and microstrip detectors as well as a transition radiation tracker. For the LHC run, the silicon pixel detector has been upgraded to include an additional layer close to the beam interaction point \( [56] \). The lead/liquid-argon electromagnetic sampling calorimeters measure electromagnetic energies with high granularity for the pseudorapidity region of \( |\eta| < 3.2 \). Hadron energies are measured by a hadronic (steel/scintillator tile) calorimeter with \( |\eta| < 1.7 \). The forward and end cap regions between \( 1.5 < |\eta| < 4.9 \) are instrumented with liquid-argon calorimeters for both the electromagnetic and hadronic measurements. Surrounding the calorimeters, the muon spectrometer includes three large superconducting toroidal magnets with eight coils each. The muon spectrometer has a system of precision tracking chambers covering \( |\eta| < 2.7 \) and fast trigger chambers covering \( |\eta| < 2.4 \). A two-level trigger system is used to select events for further analysis \([57]\). The first-level trigger is implemented in hardware and utilizes partial detector information to reduce the accepted event rate to \( 100 \text{kHz} \). The high-level trigger is based on software and accepts events at a rate of \( 1 \text{kHz} \).

## III. DATASETS

This measurement uses the dataset of \( pp \) collisions recorded by the ATLAS detector in 2016, corresponding to an integrated luminosity of \( 33 \text{fb}^{-1} \) at a center-of-mass energy of \( \sqrt{s} = 13 \text{TeV} \). Events are considered only if they were collected during stable beam conditions and satisfy data quality requirements. Due to the high instantaneous luminosity and the large total inelastic proton-proton scattering cross section, on average there are about \( 25 \) simultaneous (pileup) collisions in each bunch crossing.

The measurement presented in this paper uses a variety of Monte Carlo (MC) simulated event samples to correct for detector effects as well as for direct comparisons with the unfolded data. Inclusive jet events were generated at leading order in the strong coupling constant, \( \alpha_s \), with \textsc{Pythia} 8.1 \([58]\) using a \( 2 \to 2 \) matrix element, the NNPDF2.3LO parton distribution function (PDF) set \([59]\), and a set of generator parameter values called the A14 tune \([60]\). The \textsc{EvtGen} \([61]\) program was used to model the heavy-flavor decays to agree with experimental data. Additional inclusive jet events were simulated using a different generator in order to study the impact of modeling uncertainties related to both the perturbative and nonperturbative aspects of fragmentation. \textsc{Sherpa} 2.1 \([62]\) generates events using multileg \( 2 \to 2 \) and \( 2 \to 3 \) matrix elements, which are matched to parton showers following the Catani-Krauss-Kuhn-Webber (CKKW) prescription \([63]\). These \textsc{Sherpa} events were simulated using the CT10 PDF set \([64]\) and the default parameter tune in \textsc{Sherpa}.

Energy depositions from particles in MC event samples interacting with the detector and the subsequent detector readout were modeled using a full simulation of the ATLAS detector \([65]\) implemented in \textsc{Geant4} \([66]\). The effects of pileup were simulated with unbiased \( pp \) collisions using \textsc{Pythia} 8.1 and overlaid on the nominal inclusive jet events.

## IV. OBJECT AND EVENT SELECTION

Section IV A describes the definition of collision vertices, charged-particle tracks, jets, and \( b \)-jets. Following the discussion of objects, Section IV B describes the particle-level definition of the measurement phase space and the detector-level selections used to create an enriched sample of jets resulting from \( g \to bb \).

### A. Object reconstruction, identification and association

Collision vertices are reconstructed from ID tracks. Each vertex is required to be associated with at least two tracks with \( p_T > 0.4 \text{GeV} \). The primary hard-scattering vertex of the event is chosen to be the vertex having the highest \( \sum p_T^2 \) calculated using all associated tracks. Particle-level events

\(^1\)ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the center of the LHC ring, and the y-axis points upward. Cylindrical coordinates \( (r, \phi) \) are used in the transverse plane, \( \phi \) being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle as \( \eta = -\ln \tan(\text{polar angle}/2) \).
in simulation are considered before the addition of pileup and therefore there is no ambiguity in selecting the collision vertex.

Calorimeter jets are built from calibrated topological calorimeter-cell clusters [67] using the anti- \( k_t \) [68] algorithm with radius parameter \( R = 1.0 \) as implemented in FastJet [69]. Jets are groomed using a trimming procedure [70]. This procedure reclusters the constituents of a jet into subjets with a smaller radius of size \( R_{\text{sub}} = 0.2 \) and removes those subjets with a low fraction of the full jet momentum \( f_{\text{cut}} = 0.05 \). Following jet grooming, the mass and momentum of the resulting jets are corrected so that the detector-level values match the particle-level values on average [71]. These large-radius jets are proxies for the gluons.

Smaller-radius jets are used as proxies for the \( b \)-quarks originating from the gluons. For this purpose, jets are clustered using tracks as inputs (track-jets). Track-jets clustered from tracks with \( p_T > 500 \) MeV that have been well matched to the primary vertex\(^2\) and are composed of at least one pixel detector hit and at least six hits in the silicon tracker [72] are constructed using the anti- \( k_t \) algorithm with \( R = 0.2 \). Track-jets are required to have at least two tracks. Four-vectors are calculated for each track, assuming the mass of the charged pion. Small-radius jets are matched to large-radius jets via ghost association [73]. This matching procedure creates ghost versions of the small-radius jets with the same direction but infinitesimal \( p_T \). Jet clustering is repeated and small-radius jets are assigned to the large-radius jet that contains their ghosted version. Since the jet finding algorithm is infrared safe, the four-momenta of the jets are unaffected by the addition of ghosts.

Particle-level jets are clustered using the same algorithms as for detector-level jets, except the inputs to jet finding are all stable particles (\( ct > 10 \) mm) excluding all muons and neutrinos. The same trimming algorithm applied to calorimeter jets is also applied to the large-radius particle-level jets. Particle-level track-jets are formed from all stable charged particles that have \( p_T > 500 \) MeV and \( |\eta| < 2.5 \), excluding muons.

Track-jets that are likely to have originated from the fragmentation of a \( b \)-quark (\( b \)-jet) are identified using the MV2c10 algorithm [74,75], which is a combination of three baseline algorithms IP3D, SV, and JetFitter. The IP3D algorithm uses log-likelihood ratios of the three-dimensional signed impact parameter significance of tracks associated with jets. The SV tagger reconstructs the secondary decay vertices of \( b \)-hadrons. The JetFitter algorithm reconstructs the topology of detached vertices along the \( b \)-hadron decay axis. Finally, the MV2c10 algorithm combines the outputs of the baseline algorithms with a boosted decision tree and assigns a probability of a jet being a \( b \)-jet, \( c \)-jet, or light-flavor jet. The selected working point corresponds to a \( b \)-jet efficiency, a \( c \)-jet efficiency, and a light-flavor-jet rejection of \( \epsilon_b = 60\% \), \( \epsilon_c = 15\% \), and \( 1/\epsilon_{\text{light}} = 480 \), respectively, as measured in \( \bar{t}t \) events for jets with \( p_T > 10 \) GeV and \( |\eta| < 2.5 \). A \( p_T \) and \( \eta \)-dependent scale factor is applied to MC events to account for the measured efficiency difference between data and MC events at the chosen working point [76]. This scale factor is consistent with unity for \( b \)-jets with uncertainties ranging from a few to 10\% and the scale factor for light-flavor jets is between 1.5 and 2 with an uncertainty that ranges from about 30\%–50\%.

Particle-level track-jets are tagged as \( b \)-jets if there is a \( b \)-hadron from the simulated event record with \( p_T > 5 \) GeV that is ghost-associated with the jet. If instead a hadron containing a \( c \)-quark from the event record with \( p_T > 5 \) GeV can be matched to the jet by ghost-association, it is declared a \( c \)-jet. All other jets are declared to be light-flavor jets.

### B. Event selection

At detector level events are selected using single-jet triggers. In the first-level trigger, a sliding-window algorithm based on low-granularity calorimeter towers records events with transverse energy greater than 100 GeV. In the high-level trigger, \( R = 1.0 \) anti- \( k_t \) jets are formed from calibrated calorimeter-cell clusters and the \( p_T \) threshold is 420 GeV. This trigger scheme is fully efficient for calibrated offline jets with \( p_T > 450 \) GeV and \( |\eta| < 2 \) and therefore these kinematic requirements are used to select jets for the measurement. The offline analysis requires the highest- \( p_T \) calorimeter jet to have at least two associated track-jets with \( p_T > 10 \) GeV and \( |\eta| < 2.5 \). If there are more than two track-jets, only the leading two are used for subsequent analysis. In order to enhance the \( g \to b\bar{b} \) purity, the leading track-jet associated with the selected calorimeter jet must be \( b \)-tagged by the MV2c10 algorithm at the 60\% efficiency working point. Requiring both track-jets to be \( b \)-tagged increases the purity but degrades the precision of the background fit described in Sec. VI and so only one is required.

---

\(^2\) The transverse impact parameter must be less than 2 mm and the longitudinal impact parameter multiplied by the sine of the polar angle must be less than 3 mm.
At particle level events are required to have at least one large-radius jet with $p_T > 450$ GeV. The leading jet needs to have at least two associated particle-level track-jets with $p_T > 10$ GeV. Both of the associated small-radius jets must be tagged as $b$-jets.

This inclusive event selection produces a sample where QCD scattering processes dominate.

**V. OBSERVABLES**

The kinematic properties of the $g \to b\bar{b}$ process are characterized by three quantities: the opening angle between the $b$-quarks, the momentum sharing between the $b$-quarks, and the orientation of the gluon splitting relative to the gluon production plane. The first of these quantities is probed by measuring the $\Delta R(b, \bar{b}) = \sqrt{\Delta \phi(b, \bar{b})^2 + \Delta \eta(b, \bar{b})^2}$ between track-jets. Momentum sharing is explored using the quantity $z(p_T) = p_{T,1}/(p_{T,1} + p_{T,2})$, where $p_{T,1}$ and $p_{T,2}$ are the transverse momenta of the leading and subleading track-jets, respectively. A quantity sensitive to the relative orientation of the gluon splitting is $\Delta \theta_{ppg, gb\bar{b}}$, which is the angle between the plane spanned by the beam line and the vector sum of the two track-jets and the plane spanned by the two track-jets.

The angular quantities $\Delta R(b, \bar{b})$ and $\Delta \theta_{ppg, gb\bar{b}}$ are depicted in Fig. 2.

**FIG. 3.** The detector response is represented as the conditional probability of the detector-level quantity given the particle-level quantity, written as $Pr(\text{detector-level} | \text{particle-level})$, in simulation for $\Delta R(b, \bar{b})$ (top left), $\Delta \theta_{ppg, gb\bar{b}}$ (top right), $z(p_T)$ (bottom left), and $\log(m_{b\bar{b}}/p_T)$ (bottom right). The small antidiagonal component for $\Delta \theta_{ppg, gb\bar{b}}$ is due to cases where the leading and subleading track-jets are swapped between detector level and particle level so $\Delta \theta_{ppg, gb\bar{b}} \mapsto \pi - \Delta \theta_{ppg, gb\bar{b}}$. 

At particle level events are required to have at least one large-radius jet with $p_T > 450$ GeV. The leading jet needs to have at least two associated particle-level track-jets with $p_T > 10$ GeV. Both of the associated small-radius jets must be tagged as $b$-jets.

This inclusive event selection produces a sample where QCD scattering processes dominate.
In addition to these quantities, the dimensionless mass log($m_{bb}/p_T$) is also measured, where the mass and $p_T$ in the logarithm are computed from the four-vector sum of the two track-jets. The $b\bar{b}$ mass is an important observable for measurements and searches with Higgs and $Z$ bosons. Track-jets, in contrast to the calorimeter-based subjets from trimming, are used due to their excellent angular resolution; in simulation, there is little difference between using the directions from the track-jets, the full jets (including neutrals), or the $b$-hadrons from the $b$-quark fragmentation. There is some discrepancy between these different objects for the energy-dependent observables, but the track-jets are still useful due to their excellent momentum resolution in the gluon $p_T$ range probed in this measurement. The excellent angular and momentum resolutions are presented in Fig. 3, which shows the detector response for all four observables targeted with this measurement.

VI. BACKGROUND ESTIMATION

After the event selection, the contribution from large-radius jets that do not have two associated track-jets containing $b$-hadron is subtracted from data, as described below, before correcting for detector effects. The fraction of background events may not be well modeled by the simulation, so correction factors are determined from data template fits to the impact parameter distribution and applied for each bin of the four target observables of the analysis prior to subtraction.\textsuperscript{3} In each bin of the target observable distributions, the distribution of the signed impact parameter significance $s_{d_0}$ is fitted to data using templates from simulation while letting the fraction of each flavor component float in the fit [50,51]. For a given track, $s_{d_0} = s_j d_0 / \sigma(d_0)$, where $d_0$ is the transverse impact parameter relative to the beam line and $\sigma(d_0)$ is the uncertainty in $d_0$ from the track fit and the variable $s_j$ is the sign of $d_0$ with respect to the jet axis: $s_j = +1$ if $\sin(\phi_{jet} - \phi_{track}) \cdot d_0 > 0$ and $s_j = -1$ otherwise. The transverse impact parameter itself is signed, with sign($d_0$) = sign($\hat{p}_T,track \times (\vec{r}_{IP,xy} - \vec{r}_{PV,xy})$), where $\vec{r}_{IP,xy}$ and $\vec{r}_{PV,xy}$ are the locations of the track impact parameter and primary vertex, respectively, in the transverse plane.

Due to the long lifetime of $b$-hadrons, the values of $s_{d_0}$ for tracks in $b$-jets tend to be larger than those for tracks in $c$-jets and light-flavor jets. Therefore, the distribution of $s_{d_0}$ can be used to extract the fractions of $b$-jets, $c$-jets, and light-flavor jets using templates from simulation. For each track-jet $j$, the $s_{d_0}$ from the track with the second largest $|s_{d_0}|$, called $s_{d_0}^{ab}(j_1)$, is used for the extraction. The leading and third-leading $s_{d_0}$ values (ordered by $|s_{d_0}|$) are used as a validation of the flavor-fraction fitting procedure and produce consistent results. The leading $s_{d_0}$ is not as well modeled as $s_{d_0}^{ab}$ and therefore the $\chi^2$ resulting from the fit procedure described below is slightly worse. The value of $s_{d_0}^{ab}$ does not have a strong dependence on jet $p_T$, so the fit is performed inclusively. This choice was validated by using $p_T$-binned fits, which produce results consistent with the inclusive approach.

A binned maximum-likelihood fit to $s_{d_0}^{ab}(j_1)$ and $s_{d_0}^{ab}(j_2)$ is performed to extract the flavor fractions. Given the flavors of track-jets $j_1$ and $j_2$ ($p_{T,1} \geq p_{T,2}$), $s_{d_0}^{ab}(j_1)$ and $s_{d_0}^{ab}(j_2)$ are well approximated as being statistically independent (linear correlation is less than 5%). Therefore, the probability distribution $p(s_{d_0}^{ab}(j_1), s_{d_0}^{ab}(j_2))$ can be approximated by the product of marginals $p(s_{d_0}^{ab}(j_1)) \times p(s_{d_0}^{ab}(j_2))$. This approximation reduces a two-dimensional fit to a simultaneous fit of two one-dimensional distributions. In order to increase the robustness of the fit, flavor combinations with similar templates are merged. The three templates used for each bin of the target observable are BB (signal), B, and L + C. The BB template only includes events where both jets are labeled as $b$-jets using MC particle-level flavor labeling. The B template is an aggregation of $bl$ and $bc$ events in which the leading track-jets are labeled as $b$ and the subleading track-jets are labeled as light-flavor or $c$ in particle-level flavor labeling. This template also includes cases in which a $g \rightarrow b\bar{b}$ splitting was fully contained inside one small-radius track-jet and the second track-jet is due to a light quark or gluon. The rest of the events are merged into the L + C template. The nominal results use fits from PYTHIA templates. Uncertainties in the templates, in particular resulting from the template merging scheme and from the choice of generator, are described in Sec. VIII A.

Examples of the flavor-fraction determination fits are shown in Fig. 4 for one bin of $\Delta R(b, b')$. The template binning was chosen to have enough sensitivity in the tails while also having sufficient simulation statistics to populate the templates; for reference, the MC statistical uncertainty is shown in the error bands. Since the leading track-jet is required to be $b$-tagged, the distribution of $s_{d_0}^{ab}(j_1)$ is broader and shifted toward more positive values than the distribution of $s_{d_0}^{ab}(j_2)$. In contrast, the subleading jet is most often a light-flavor jet for both the B and L + C categories and therefore $s_{d_0}^{ab}(j_2)$ is nearly symmetric about zero. The BB template is similar in shape for the leading and subleading track-jets. For this particular fit, the $\chi^2$ per degree of freedom improves from 72/22 (prefit) to 13.5/22 (postfit). As a result of the fit, the background fraction changes from 79.6% to 82.8%. A comparison of the prefit and postfit flavor fractions for all bins of $\Delta R$ as well as the other observables is presented in Fig. 5. Except for $\Delta \theta_{ppg,bb}$ and the highest bin of $z(p_T)$ the PYTHIA simulation prediction for the shape of the flavor-fraction

\textsuperscript{3}As these fits are performed at detector level, there is no statistical correlation between bins. Therefore, the fits are performed independently for each observable bin.
FIG. 4. The distribution of $s_{d_{\text{sub}}}$ in data and in simulation, postfit, for the higher-$p_T$ track-jet (left) and for the lower-$p_T$ track-jet (right) in the bin $0.25 < \Delta R(b, b) < 0.3$. The three components are the signal double-$b$ (BB), the background single $b$ (B), and the background non-$b$ components (L + C). Percentages reported in the legend indicate the prefit and postfit fraction of each component. Only data and MC statistical uncertainties are shown. The lower panel shows the ratio between data and the postfit simulation.

FIG. 5. The prefit (MC) and postfit (data) flavor fractions for $\Delta R(b, b)$ (top left), $\Delta \theta_{ppg, gb}$ (top right), $z(p_T)$ (bottom left), and $\log(m_{pp}/p_T)$ (bottom right) are indicated with open and solid markers respectively. The error bars include only statistical uncertainties from the flavor-fraction fit. The fit’s systematic uncertainties are comparable in magnitude, but correlated across the bins. The impact of both the flavor-fraction fit’s statistical and systematic uncertainties on the final results is presented in Table 1.
distribution is accurate for the BB fraction. In all cases the BB yield is slightly overestimated. The flavor fractions from B and L + C are inverted between PYTHIA and data.

VII. UNFOLDING

After subtracting the background from the detector-level distributions, as described in Sec. VI, the data are corrected for resolution and acceptance effects. The fiducial volume of the measurement is described by the particle-level object and event selection in Sec. IV. First, the data are corrected for events that pass the detector-level selection but not the particle-level selection using the simulations introduced in Sec. III. Then, the iterative Bayes (IB) unfolding technique [77] is used to correct for the detector resolution in events that pass both the detector-level and particle-level selections. The IB method is applied with four iterations implemented in the RooUnfold framework [78]. After the application of the response matrix, a final correction is applied to account for events that pass the particle-level but not detector-level selection. Uncertainties in the unfolding procedure are described in Sec. VIII.

VIII. UNCERTAINTIES

Systematic and statistical uncertainties were assessed for each aspect of the analysis, including the background subtraction, acceptance and efficiency correction factors, response matrix, and unfolding method. For each uncertainty, a component of the analysis chain is varied and then the entire procedure including the background subtraction is repeated. Table I provides a summary of the systematic uncertainties for each observable. The jet energy scale, the unfolding, and the theoretical modeling uncertainties dominate. Further details about each category are provided in Sec. VIII A for the uncertainties associated with each analysis object, in Sec. VIII B for the background fit procedure uncertainty, and in Sec. VIII C for the unfolding method and theoretical modeling uncertainties.

<table>
<thead>
<tr>
<th>TABLE I. Summary of systematic uncertainty sizes for each observable for the normalized differential cross sections.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ΔR</strong>(b, b)</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Calorimeter jet energy</td>
</tr>
<tr>
<td>Flavor tagging</td>
</tr>
<tr>
<td>Tracking</td>
</tr>
<tr>
<td>Background fit</td>
</tr>
<tr>
<td>Unfolding method</td>
</tr>
<tr>
<td>Theoretical modeling</td>
</tr>
<tr>
<td>Statistical</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

A. Object reconstruction

Each object used in the analysis has an associated uncertainty. These uncertainties affect the acceptance factors and the response matrix, as well as the background fit templates.

1. Calorimeter jets

The energies of large-radius jets are shifted and smeared to account for uncertainties in both the bias and variance of the reconstructed energy. Jet energy scale uncertainties are determined by comparing calorimeter-based and tracker-based jet energy measurements in inclusive dijet events, and range from 2% to 6% [79].

2. Flavor tagging

Data/MC corrections (“scale factors”) and uncertainties in the b-tagging efficiencies and c-jet misidentification rates are determined from t\overline{t} events [80]. Light-flavor misidentification rates are studied using dijet events. The b-jet uncertainties are 5%–10%, while the c-jet uncertainties are 20% and light-flavor jet uncertainties are about 50%. Uncertainties for track-jets with p_T > 300 GeV are extrapolated from low-p_T jets as there are too few events with high-p_T jets for a proper calibration. Extrapolation uncertainties are evaluated by varying quantities such as impact parameter resolutions and descriptions of the detector material. The extrapolated uncertainties are added in quadrature on top of the data-based p_T and η-dependent uncertainties and range from 20% to 100% depending on the p_T and the flavor of the jet. As the flavor fractions are constrained in situ, there is a significant reduction from the prior flavor-tagging uncertainty values described above. In particular, there is little sensitivity to the background scale factors and there is no sensitivity to inclusive flavor-tagging scale factors for b-jets. Residual ΔR(b, b)-dependent scale factor uncertainties that account for differences in the modeling of isolated versus nonisolated b-jets are derived from dedicated performance studies [50,51].

3. Tracking

Systematic uncertainties are estimated for the track reconstruction efficiency, fake rate, and track parameter scales and resolutions. The main source of inclusive tracking inefficiency is multiple scattering in the ID, so the uncertainty is set by the accuracy with which the ID material is simulated [81]. This leads to a 0.5% uncertainty for |η| < 0.1, which grows to 2.7% at the end of the ID acceptance. An additional source of inefficiency arises inside the high-multiplicity environment in the cores of jets due to silicon pixel and microstrip cluster merging. The uncertainty in the modeling of this density-induced inefficiency is about 0.8% [82,83]. Fake tracks are due to combinations from multiple charged particles. The track selection described in Sec. IV reduces the contribution of
fake tracks to much less than 1% with a relative uncertainty that is about 30% [84]. The track parameters that are most relevant for this analysis are $p_T$ and $d_0$. Weak modes in the ID alignment cause a bias in the track sagitta that is corrected for using a dedicated calibration and the calibration uncertainty is propagated through to the measurement [72]. The modeling of the $d_0$ scale and resolution was studied in $Z \rightarrow \mu^+\mu^-$ events at low $p_T$ and in dijet events at high $p_T$ [72] and is used to assign an uncertainty to the modeling of these important track properties.

**B. Background fit**

In addition to the fit validations described in Sec. VI, several aspects of the fit are varied in order to assess the uncertainty in the extracted flavor fractions.

**1. Fit range**

The nominal flavor fraction fit is performed for $s_{d_0}^{ub} \in [-40, 70]$. In order to assess the impact of this choice and the sensitivity to the $s_{d_0}^{ub}$ tails, the fit is repeated while excluding the left and right tails of the distributions, corresponding to $s_{d_0}^{ub} \in [-30, 70]$ and $s_{d_0}^{ub} \in [-40, 60]$.

**2. Template merging scheme**

Merging background components to form three aggregated templates fixes the relative fractions of the template subcomponents. The sensitivity of the fitted flavor fractions to this choice is estimated by varying each merged background component up or down by a factor of 2. The fit range has a bigger impact on the uncertainty than the merging variations of the flavor fractions.

**C. Unfolding method and theoretical modeling**

An uncertainty resulting from the unfolding method described in Sec. VII is determined by unfolding the prediction from a different simulation with the nominal procedure. The alternative simulation is constructed by reweighting the nominal particle-level spectrum so that the simulated detector-level spectrum, obtained by propagating the reweighted particle-level spectrum through the response matrix, agrees well with the data. The modified detector-level distribution is unfolded with the nominal response matrix and the difference between this and the reweighted particle-level spectrum is an indication of the bias due to the unfolding method (in particular, the choice of prior) [85].

The unfolded result can depend on the modeling of jet fragmentation through the background fit, the prior, the response matrix, and the correction factors. The $s_{d_0}$ distribution does not strongly depend on the properties of the jet radiation pattern, but an uncertainty is determined by taking the fitted background using templates from SHERPA instead of PYTHIA. Variations in the prior are already accounted for in the data-driven nonclosure uncertainty described above. The rest of the contributions are evaluated by comparing the result using PYTHIA with the result using the alternative SHERPA sample described in Sec. VI. This comparison is decomposed into components corresponding to varying only the response matrix or only the initial/final correction factors. Varying only one component at a time is possible by reweighting a component of the PYTHIA simulation to match the SHERPA simulation and then evaluating the relative difference in the unfolded result. All of the components are added in quadrature to determine the total uncertainty due to fragmentation modeling. Each component is treated as uncorrelated because the uncertainty is based on only two fragmentation models and therefore a potential reduced uncertainty from exploiting potentially unphysical correlations between kinematic properties impacting acceptance and substructure attributes impacting the response is avoided.

**IX. RESULTS**

The unfolded results along with multiple parton shower MC predictions are presented in Fig. 6. By construction, $0.2 \leq \Delta R \leq 1.0$ and the peak around 0.3 is due to the radius of the track-jets. Furthermore, $0 \leq z(p_T) \leq 0.5$, with some distortions to the natural distribution at low values due to the $p_T$ threshold applied to the small-radius track-jets.

The SHERPA predictions are generally more accurate than those from PYTHIA, although there are significant differences between both generators and the data at low mass, low $z(p_T)$ and for all $\Delta \theta_{ppg,gbb}$. The $\Delta \theta_{ppg,gbb}$ distribution in data appears to be inverted with respect to the one from PYTHIA (with a minimum instead of maximum at $\pi/2$) while SHERPA predicts a relatively uniform distribution. For comparison, the figure contains a version of PYTHIA$^4$ with the azimuthal asymmetries induced by gluon polarization turned off. This sample appears to be closer to SHERPA and also to the data, consistent with the studies in Ref. [86], which also suggest a good agreement with simulations that include higher-order effects. In general, the properties of gluon polarization inside unpolarized hadrons are largely unconstrained by experimental data (see e.g., Ref. [87] and references therein). This and future measurements of $\Delta \theta_{ppg,gbb}$ may provide a new way to extract $p_T$-dependent parton distributions in order to better understand proton structure and further improve the precision of various cross-section calculations [88].

In addition to studying gluon production properties, $g \rightarrow bb$ provides a handle on gluon fragmentation. Due to the large $b$-quark mass and in general the large $m_{bb}$ mass that is

---

$^4$Due to the available settings in PYTHIA, different versions were used for these comparisons. It was checked that there is no significant difference in $\Delta \theta_{ppg,gbb}$ between PYTHIA versions 8.186 and 8.230.

possible after splitting, there are many formally equivalent model choices in describing gluon fragmentation. For example, the scale at which the strong coupling constant acts (renormalization scale) may be better described as scaling with \( m_{b\bar{b}}^2 \) instead of the PYTHIA default \( p_T^{b\bar{b}} \). To illustrate the sensitivity of \( \Delta R(b, \bar{b}) \), \( z(p_T) \) (bottom left), and \( \log(m_{b\bar{b}}/p_T) \) (bottom right) to fragmentation settings in PYTHIA, the plots of Fig. 6 show the final-state radiation variations of the A14 tune (indicated as an uncertainty band in the plot) as well as a different way to treat the \( b\bar{b} \) pair mass in the QCD splitting kernels indicated by the \( m_{b\bar{b}}^2/4 \) variation (\( m_{b\bar{b}}^2/4 \) instead of \( p_T^{b\bar{b}} \) for the renormalization scale). No variation describes all of the data, but some variations are worse than others.

For example, the \( \text{Var2} + \text{A14} \) variation, which increases the final-state shower \( \alpha_S(M_Z) \) value to 0.139, moves the prediction further from the data in nearly all measurement bins. Related variations (not shown) such as using \( m_{b\bar{b}}^2 \) instead of \( p_T^{b\bar{b}} \) as the renormalization scale, adding additional phase-space factors, or suppressing high-mass \( b\bar{b} \) pairs, are not significantly different from the nominal PYTHIA setup.\(^7\)

\(^6\)TimeShower:weightGluonToQuark=5 and TimeShower:scaleGluonToQuark=0.25.

\(^7\)Variations of TimeShower:weightGluonToQuark and TimeShower:scaleGluonToQuark: when TimeShower:weightGluonToQuark=1, then the \( g \to b\bar{b} \) kernel is weighted by an extra \( \beta \) phase-space factor; when TimeShower:weightGluonToQuark=5, this kernel is re-weighted to \( \alpha_S(m_{b\bar{b}}^2) \) instead of \( \alpha_S(p_T^2) \); when TimeShower:weightGluonToQuark=8, there is an additional factor that suppresses the rate of high-mass \( b\bar{b} \) pairs.
X. CONCLUSION

This paper presents a measurement of various properties of $g\rightarrow b\bar{b}$ at high $p_T$ and low $\Delta R(b,\bar{b})$ from 33 fb$^{-1}$ of $\sqrt{s}=13$ TeV $pp$ collisions recorded by the ATLAS detector at the LHC. A flavor-fraction fit is used to remove contributions from processes other than $g\rightarrow b\bar{b}$. The fitted fractions significantly disagree with the prefit PYTHIA predictions and suggest that further studies could improve the modeling of analyses sensitive to these fractions. The measured properties are unfolded to correct for the detector acceptance and resolution for direct comparison with particle-level models. Comparisons are made at the particle level between the distributions and various models of jet formation. Simulations from the SHERPA event generator generally provide a better model than PYTHIA, especially for the $\Delta\phi_{p,q,b\bar{b}}$ observable, which is sensitive to the modeling of the gluon polarization. The particle-level spectra are publicly available \[89\] for further interpretation and can be used to validate QCD MC predictions and tune their models’ free parameters.

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; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; CONICYT, Chile; CAS, MOST and NSFC, China; COLCIENCIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS, CEA-DRF/IFRF, France; SRNSFG, Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russian Federation; JINR; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; and DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, CRC and Compute Canada, Canada; COST, ERC, ERDF, Horizon 2020, and Marie Skłodowska-Curie Actions, European Union; Investissements d’ Avenir Labex and Idex, ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programs cofinanced by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; CERCA Programme Generalitat de Catalunya, Spain; and 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); BNL (USA); the Tier-2 facilities worldwide; and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. \[90\].

[3] ATLAS Collaboration, Search for heavy resonances decaying to a $W$ or $Z$ boson and a Higgs boson in the $q\bar{q}/gg\rightarrow b\bar{b}$ final state in $pp$ collisions at $\sqrt{s}=13$ TeV with the ATLAS detector, Phys. Lett. B 774, 494 (2017).


[23] D0 Collaboration, The $b\bar{b}$ production cross section and angular correlations in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV, Phys. Lett. B 487, 264 (2000).


[26] D0 Collaboration, Inclusive $\mu$ and $b$-Quark Production Cross Sections in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV, Phys. Rev. Lett. 74, 3548 (1995).


[28] LHCb Collaboration, Measurement of $\sigma(pp \rightarrow b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region, Phys. Lett. B 694, 209 (2010).


[42] CMS Collaboration, Measurement of the production cross sections for a $Z$ boson and one or more $b$ jets in $pp$ collisions at $\sqrt{s} = 7$ TeV, J. High Energy Phys. 06 (2014) 120.


[44] D0 Collaboration, Measurement of the ratio of differential cross sections $\sigma(pp \rightarrow Z + b$ jet)/$\sigma(pp \rightarrow Z + jet)$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. D 87, 092010 (2013).

[45] D0 Collaboration, Measurement of the ratio of inclusive cross sections $\sigma(pp \rightarrow Z + 2b$ jets)/$\sigma(pp \rightarrow Z + 2$ jets) in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. D 91, 052010 (2015).

PROPERTIES OF $g \to b\bar{b}$ AT ...

PHYS. REV. D 99, 052004 (2019)
106 Research Institute for Nuclear Problems of Byelorussian State University, Minsk, Belarus
107 Group of Particle Physics, University of Montreal, Montreal Québec, Canada
108 P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow, Russia
109 Institute for Theoretical and Experimental Physics of the National Research Centre Kurchatov Institute, Moscow, Russia
110 National Research Nuclear University MEPhI, Moscow, Russia
111 D.V. Skobeltsyn Institute of Nuclear Physics, M.V. Lomonosov Moscow State University, Moscow, Russia
112 Fakultät für Physik, Ludwig-Maximilians-Universität München, München, Germany
113 Max-Planck-Institut für Physik (Werner-Heisenberg-Institut), München, Germany
114 Nagasaki Institute of Applied Science, Nagasaki, Japan
115 Graduate School of Science and Kobayashi-Maskawa Institute, Nagoya University, Nagoya, Japan
116 Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico, USA
117 Institute for Mathematics, Astrophysics and Particle Physics, Radboud University Nijmegen/Nikhef, Nijmegen, Netherlands
118 Nikhef National Institute for Subatomic Physics and University of Amsterdam, Amstterdam, Netherlands
119 Department of Physics, Northern Illinois University, DeKalb, Illinois, USA
120 Budker Institute of Nuclear Physics and NSU, SB RAS, Novosibirsk, Russia
121 Institute for High Energy Physics of the National Research Centre Kurchatov Institute, Protvino, Russia
122 Department of Physics, New York University, New York, New York, USA
123 The Ohio State University, Columbus, Ohio, USA
124 Faculty of Science, Okayama University, Okayama, Japan
125 Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman, Oklahoma, USA
126 Department of Physics, Oklahoma State University, Stillwater, Oklahoma, USA
127 Palacký University, RCPTM, Joint Laboratory of Optics, Olomouc, Czech Republic
128 Center for High Energy Physics, University of Oregon, Eugene, Oregon, USA
129 LAL, Université Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
130 Graduate School of Science, Osaka University, Osaka, Japan
131 Department of Physics, University of Oslo, Oslo, Norway
132 Department of Physics, Oxford University, Oxford, United Kingdom
133 LPNHE, Sorbonne Université, Paris Diderot Sorbonne Paris Cité, CNRS/IN2P3, Paris, France
134 Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania, USA
135 Konstantinov Nuclear Physics Institute of National Research Centre “Kurchatov Institute”, PNPI, St. Petersburg, Russia
136 Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
137 Departamento de Física y el Cosmos, Universidad de Granada, Granada, Spain
138 Dep Física and CEFITEC of Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
139 Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic
140 Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic
141 Particle Physics Department, Rutherford Appleton Laboratory, Didcot, United Kingdom
142 IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
143 Santa Cruz Institute for Particle Physics, University of California Santa Cruz, Santa Cruz, California, USA
144 Instituto de Física, Pontificia Universidad Católica de Chile, Santiago, Chile
145 Departamento de Física, Universidad Técnica Federico Santa María, Valparaiso, Chile
146 Department of Physics, University of Washington, Seattle, Washington, USA
147 Department of Physics and Astronomy, University of Sheffield, Sheffield, United Kingdom
148 K. A. Tolmachev Institute for Theoretical and Experimental Physics, Frunze 5, Moscow, Russia
149 Department of Physics, Simon Fraser University, Burnaby British Columbia, Canada
150 SLAC National Accelerator Laboratory, Stanford, California, USA