Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment

Aaboud, M.; The ATLAS Collaboration

DOI
10.1103/PhysRevLett.120.211802

Publication date
2018

Document Version
Final published version

Published in
Physical Review Letters

License
CC BY

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment

M. Aaboud et al.*
(ATLAS Collaboration)

(Received 14 February 2018; published 22 May 2018)

A direct search for the standard model Higgs boson decaying to a pair of charm quarks is presented. Associated production of the Higgs and Z bosons, in the decay mode $ZH \rightarrow \ell^+\ell^-\bar{c}c$ is studied. A data set with an integrated luminosity of 36.1 fb$^{-1}$ of $pp$ collisions at $\sqrt{s} = 13$ TeV recorded by the ATLAS experiment at the LHC is used. The $H \rightarrow \bar{c}c$ signature is identified using charm-tagging algorithms. The observed (expected) upper limit on $\sigma(pp \rightarrow ZH) \times B(H \rightarrow \bar{c}c)$ is 2.7 (3.9 $^{+1.1}_{-1.0}$) pb at the 95% confidence level for a Higgs boson mass of 125 GeV, while the standard model value is 26 fb.

DO: 10.1103/PhysRevLett.120.211802

In July 2012, the ATLAS and CMS collaborations announced the discovery of a new particle with a mass of approximately 125 GeV [1,2] in searches for the standard model (SM) Higgs boson at the Large Hadron Collider (LHC) [3]. Subsequent measurements indicate that this particle is consistent with the SM Higgs boson [4–10]. Direct evidence for the Yukawa coupling of the Higgs boson to the top [11] and bottom [12,13] quarks was recently obtained. Measurements of the Yukawa coupling of the Higgs boson to quarks in generations other than the third are difficult at hadron colliders, due to small branching fractions, large backgrounds, and challenges in jet flavor identification [14,15]. This Letter presents a direct search by the ATLAS experiment for the decay of the Higgs boson to a pair of charm ($c$) quarks. This search targets the production of the Higgs boson in association with a $Z$ boson decaying to charged leptons: $Z(\ell^+\ell^-)H(c\bar{c})$, where $\ell = e, \mu$.

The SM branching fraction for a Higgs boson with a mass of 125 GeV to decay to a pair of charm quarks is predicted to be 2.9% [16]. The inclusive cross section for $\sigma(pp \rightarrow ZH) \times B(H \rightarrow \bar{c}c)$ is 26 fb at $\sqrt{s} = 13$ TeV [17]. Rare exclusive decays of the Higgs boson to a light vector meson or quarkonium state and a photon can also probe the couplings of the second-generation quarks to the Higgs boson [18–21]. Previously, the ATLAS Collaboration presented an indirect search for the decay of the Higgs boson to $c$ quarks via the decay to $J/\psi\gamma$, obtaining a branching fraction limit of $1.5 \times 10^{-3}$ at the 95% confidence level (C.L.), which approximately corresponds to a limit of 540 times the SM branching fraction prediction [14,20]. Bounds on the Higgs boson branching fractions to unobserved final states and fits to global rates constrain $B(H \rightarrow \bar{c}c) < 20\%$ at the 95% C.L., assuming SM production cross sections [22]. These limits can still accommodate large modifications to the Higgs boson coupling to charm quarks from new physics [22]. In this Letter, a new approach is introduced to investigate the coupling of the Higgs boson to charm quarks.

The search is performed using $pp$ collision data recorded in 2015 and 2016 with the ATLAS detector [23] at $\sqrt{s} = 13$ TeV. The ATLAS detector at the LHC covers nearly the entire solid angle around the collision point [24]. It consists of an inner tracking detector surrounded by a thin superconducting solenoid, electromagnetic and hadronic calorimeters, and a muon spectrometer incorporating three large superconducting toroidal magnets. An additional pixel layer was installed for the $\sqrt{s} = 13$ TeV running period [25]. After the application of beam, detector, and data-quality requirements, the integrated luminosity corresponds to 36.1 $\pm$ 0.8 fb$^{-1}$, measured following Ref. [26]. Events are required to contain exactly two same-flavor leptons with an invariant mass consistent with that of the $Z$ boson, and at least two jets of which one or two are identified as charm jets ($c$ jets). In this Letter, lepton refers to only electrons or muons. The analysis procedure is validated by measuring the yield of $ZW$ and $ZZ$ production, where the sample is enriched in $W \rightarrow cs, cd$ and $Z \rightarrow \bar{c}c$ decays. Further details can be found in Ref. [12].

Monte Carlo (MC) simulated samples were produced for signal and background processes using the full ATLAS detector simulation [27] using GEANT4 [28]. Table I provides details of the event generators used for each signal and background sample. Signal events were produced at next-to-leading order (NLO) for the $q\bar{q} \rightarrow ZH$ process and at leading order (LO) for the $gg \rightarrow ZH$ process with POWHEG-BOX v2 [32]. The dominant $Z +$ jets background and the resonant diboson $ZW$ and $ZZ$ processes were generated using SHERPA 2.2.1 [54]. The $t\bar{t}$ background was
TABLE I. The configurations used for event generation of the signal and background processes. If two parton distribution functions (PDFs) are shown, the first is for the matrix element calculation and the second for the parton shower, otherwise the same is used for both. Alternative event generators and configurations, used to estimate systematic uncertainties, are in parentheses. Tune refers to the underlying-event tuned parameters of the parton shower event generator. MG5_AMC refers to MadGRAPH5_AMC@NLO 2.2.2 [29]; PYTHIA 8 refers to version 8.212 [30]. Heavy-flavor hadron decays modeled by EvtGen 1.2.0 [31] are used for all samples except those generated using SHERPA. The order of the calculation of the cross sections used to normalize the predictions is indicated. The $q\bar{q} \rightarrow ZH$ cross section is estimated by subtracting the $gg \rightarrow ZH$ cross section from the $pp \rightarrow ZH$ cross section. The asterisk (*) in the last column denotes that the indicated order is for the $pp \rightarrow ZH$ cross section. NNLO denotes next-to-next-to-leading order; NLL denotes next-to-leading log and NNLL denotes next-to-next-to-leading log.

<table>
<thead>
<tr>
<th>Process</th>
<th>Event Generator (alternative)</th>
<th>Parton Shower (alternative)</th>
<th>PDF (alternative)</th>
<th>Tune</th>
<th>Cross section</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q\bar{q} \rightarrow ZH$</td>
<td>POWHEG-BOX v2 +MiNLO [45,46]</td>
<td>(HERWIG 7 [47])</td>
<td>PDF4LHC15NLO [33]</td>
<td>AZNLO [34]</td>
<td>NNLO (QCD)*</td>
</tr>
<tr>
<td>$gg \rightarrow ZH$</td>
<td>POWHEG-BOX v2</td>
<td>(HERWIG 7)</td>
<td>PDF4LHC15NLO</td>
<td>A14 [48]</td>
<td>NLO+NLL (QCD)</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>POWHEG-BOX v2</td>
<td>(HERWIG 7)</td>
<td>NNPDF3.0NNLO [52]</td>
<td>A14</td>
<td>NLO+NNL [53]</td>
</tr>
<tr>
<td>$ZW, ZZ$</td>
<td>SHERPA 2.2.1 [54] (POWHEG-BOX)</td>
<td>(PYTHIA 8)</td>
<td>NNPDF2.3LO</td>
<td>NLO</td>
<td></td>
</tr>
<tr>
<td>$Z + \text{jets}$</td>
<td>SHERPA 2.2.1 (MG5_AMC)</td>
<td>(PYTHIA 8)</td>
<td>NNPDF2.3LO</td>
<td>NLO</td>
<td></td>
</tr>
</tbody>
</table>

Events are required to contain at least two jets. Jets are reconstructed from topological energy clusters in the calorimeters consistent with the passage of minimum-ionizing particles [58]. For data recorded in 2015, the single-electron trigger required a candidate with $p_T > 24(20) \text{ GeV}$; in 2016 the lepton $p_T$ threshold was raised to 26 GeV. Events are required to contain a pair of same-flavor leptons, both satisfying $p_T > 7 \text{ GeV}$ and $|\eta| < 2.5$. At least one lepton must have $p_T > 27 \text{ GeV}$ and correspond to a lepton that passed the trigger. The two leptons are required to satisfy loose track-isolation criteria with an efficiency greater than 99%. They are required to have opposite charge in dimuon events, but not in dielectron events due to the non-negligible charge misidentification rate of electrons. The invariant mass of the dilepton system is required to be consistent with the mass of the Z boson: $81 \text{ GeV} < m_{\ell\ell} < 101 \text{ GeV}$.

Jets are reconstructed from topological energy clusters in the calorimeters [59,60] using the anti-$k_T$ algorithm [61] with a radius parameter of 0.4 implemented in the FASTJET package [62]. The jet energy is corrected using a jet-area-based technique [63,64] and calibrated [65,66] using $p_T$- and $\eta$-dependent correction factors determined from simulation, with residual corrections from internal jet properties. Further corrections from in situ measurements are applied to data. Selected jets must have $p_T > 20 \text{ GeV}$ and $|\eta| < 2.5$. Events are required to contain at least two jets. If a muon is found within a jet, its momentum is added to the selected jet. An overlap removal procedure resolves cases in which the same physical object is reconstructed multiple times, e.g. an electron also reconstructed as a jet.

![FIG. 1. The $c$-jet-tagging efficiency (colored scale) as a function of the $b$ jet and $l$ jet rejection as obtained from simulated $t\bar{t}$ events. The cross, labeled as working point, WP, denotes the selection criterion used in this analysis. The solid and dotted black lines indicate the contours in rejection space for the fixed $c$-tagging efficiency used in the analysis and two alternatives.](211802-2)
TABLE II. Breakdown of the relative contributions to the total uncertainty in $\mu$. The statistical uncertainty includes the contribution from the floating $Z +$ jets normalization parameters. The sum in quadrature of the individual components differs from the total uncertainty due to correlations between the components.

<table>
<thead>
<tr>
<th>Source</th>
<th>$\sigma/\sigma_{\text{tot}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td>49%</td>
</tr>
<tr>
<td>Floating $Z +$ jets</td>
<td>31%</td>
</tr>
<tr>
<td>Systematic</td>
<td>87%</td>
</tr>
<tr>
<td>Flavor tagging</td>
<td>73%</td>
</tr>
<tr>
<td>Background modeling</td>
<td>47%</td>
</tr>
<tr>
<td>Lepton, jet luminosity</td>
<td>28%</td>
</tr>
<tr>
<td>Signal modeling</td>
<td>28%</td>
</tr>
<tr>
<td>MC statistical</td>
<td>6%</td>
</tr>
</tbody>
</table>

Jets in simulated events are labeled according to the presence of a heavy-flavor hadron with $p_T > 5$ GeV within $\Delta R = 0.3$ from the jet axis. If a $b$ hadron is found the jet is labeled as a $b$ jet. If no $b$ hadron is found, but a $c$ hadron is present, then the jet is labeled as a $c$ jet. Otherwise the jet is labeled as a light-flavor jet ($l$ jet).

Flavor-tagging algorithms exploit the different lifetimes of $b$, $c$, and light-flavor hadrons. A $c$-tagging algorithm is used to identify $c$ jets. Charm jets are particularly challenging to tag because $c$ hadrons have shorter lifetimes and decay to fewer charged particles than $b$ hadrons. Boosted decision trees are trained to obtain two multivariate discriminants: to separate $c$ jets from $l$ jets and $c$ jets from $b$ jets. The same variables used for $b$ tagging [67,68] are used. Figure 1 shows the selection criteria applied in the two-dimensional multivariate discriminant space, to obtain an efficiency of 41% for $c$ jets and rejection factors of 4.0 and 20 for $b$ jets and $l$ jets. The efficiencies are calibrated to data using $b$ quarks from $t \rightarrow Wb$ and $c$ quarks from $W \rightarrow cs$, $cd$ with methods identical to the $b$-tagging algorithms [67]. Statistical uncertainties in the simulation are reduced, by weighting events according to the tagging efficiencies of their jets, parametrized as a function of jet flavor, $p_T$, $\eta$ and the angular separation between jets, rather than imposing a direct requirement on the $c$-tagging discriminants.

Data are analyzed in four categories with different expected signal purities. The dijet invariant mass, $m_{c\bar{c}}$, constructed using the two highest-$p_T$ jets, is the discriminating variable in each category. Categories are defined using the transverse momentum of the reconstructed Z boson, $p_T^Z$ (75 GeV $\leq p_T^Z < 150$ GeV and $p_T^Z \geq 150$ GeV) and the number of $c$ tags amongst the leading jets (either one or two). The $p_T^Z$ requirements exploit the harder $p_T^Z$ distribution in $ZH$ compared to $Z +$ jets production. Background events are rejected by requiring the angular separation between the two jets constituting the dijet system, $\Delta R_{c\bar{c}}$, to be less than 2.2, 1.5, or 1.3 for events satisfying $75 \leq p_T^Z < 150$ GeV, $150 \leq p_T^Z < 200$ GeV, or $p_T^Z \geq 200$ GeV. The signal acceptance ranges from 0.5% to 3.4% depending on the category. A joint binned maximum-profile-likelihood fit to $m_{c\bar{c}}$ in the categories is used to extract the signal yield and the $Z +$ jets background normalization. The fit uses 15 bins in each category within the range of 50 GeV $< m_{c\bar{c}} < 200$ GeV, with a bin width of 10 GeV. The parameter of interest, $\mu$, common to all categories, is the signal strength, defined as the ratio of the measured signal yield to the SM prediction.

Systematic uncertainties affecting the signal and background predictions include theoretical uncertainties in the signal and background modeling and experimental uncertainties. Table II shows their relative impact on the fitted value of $\mu$. Uncertainties in the $m_{c\bar{c}}$ shape of the backgrounds are assessed by comparisons between nominal and alternative event generators as indicated in Table I.

Systematic uncertainties are incorporated within the statistical model through nuisance parameters that modify the shape and/or normalization of the distributions. Statistical uncertainties in the simulation samples are accounted for. The $Z +$ jets background is normalized from the data through the inclusion of an unconstrained normalization parameter for each category. The fitted

TABLE III. Postfit yields for the signal and background processes in each category from the profile likelihood fit. Uncertainties include statistical and systematic contributions. The prefit SM expected $ZH(c\bar{c})$ signal yields are indicated in parenthesis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$Y_{1 \text{c tag}}, 50$ GeV $&lt; m_{c\bar{c}} &lt;$ 200 GeV</th>
<th>$Y_{2 \text{c tags}}, 50$ GeV $&lt; m_{c\bar{c}} &lt;$ 200 GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$75 \leq p_T^Z &lt; 150$ GeV</td>
<td>$p_T^Z \geq 150$ GeV</td>
</tr>
<tr>
<td>$Z +$ jets</td>
<td>69400 $\pm$ 500</td>
<td>15650 $\pm$ 180</td>
</tr>
<tr>
<td>$ZW$</td>
<td>750 $\pm$ 130</td>
<td>290 $\pm$ 50</td>
</tr>
<tr>
<td>$ZZ$</td>
<td>490 $\pm$ 70</td>
<td>180 $\pm$ 28</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>2020 $\pm$ 280</td>
<td>130 $\pm$ 50</td>
</tr>
<tr>
<td>$ZH(b\bar{b})$</td>
<td>32 $\pm$ 2</td>
<td>19.5 $\pm$ 1.5</td>
</tr>
<tr>
<td>$ZH(c\bar{c})$ (SM)</td>
<td>$-143 \pm 170 (2.4)$</td>
<td>$-84 \pm 100 (1.4)$</td>
</tr>
<tr>
<td>Total</td>
<td>72500 $\pm$ 320</td>
<td>16180 $\pm$ 140</td>
</tr>
<tr>
<td>Data</td>
<td>72504</td>
<td>16181</td>
</tr>
</tbody>
</table>

$211802-3$
normalization parameters range between 1.13 and 1.30. All other background normalization factors are correlated between categories, with acceptance uncertainties of order 10% to account for relative variations between categories.

The dominant contributions to the uncertainty in $\mu$ are the efficiency of the tagging algorithms, the jet energy scale and resolution, and the background modeling. The largest uncertainty is due to the normalization of the dominant $Z+\text{jets}$ background. The typical uncertainty in the tagging efficiency is 25% for $c$ jets, 5% for $b$ jets, and 20% for $l$ jets.

Table III shows the fitted signal and background yields. The $m_{c\bar{c}}$ distributions in the 2 $c$ tag categories are shown in Fig. 2 with the background shapes and normalizations according to the result of the fit. Good agreement is observed between the postfit shapes of the distributions and the data.

The analysis procedure is validated by measuring the yield of $ZV$ production, where $V$ denotes a $W$ or $Z$ boson, with the same event selection. The fraction of the $ZZ$ yield from $Z \rightarrow c\bar{c}$ decays is $\sim 55%$ (20%) in the 2 $c$ tag (1 $c$ tag) category, while the fraction of the $ZW$ yield from $W \rightarrow c\bar{s}$, $cd$ is $\sim 65\%$ for both the 2 and 1 $c$ tag categories. Contributions of Higgs boson decays to $cc$ and $b\bar{b}$ are treated as background and constrained to the SM predictions within its theoretical uncertainties. The diboson signal strength is measured to be $\mu_{ZV} = 0.6_{-0.3}^{+0.3}$ with an observed (expected) significance of 1.4 (2.2) standard deviations.

The best-fit value for the $ZH(c\bar{c})$ signal strength is $\mu_{ZH} = -69 \pm 101$. By assuming a signal with the kinematics of the SM Higgs boson, model-dependent corrections are made to extrapolate to the inclusive phase space. Hence, an upper limit on $\sigma(pp\rightarrow ZH) \times B(H \rightarrow c\bar{c})$ is computed using a modified frequentist CL$_s$ method [69,70] with the profile likelihood ratio as the test statistic. The observed (expected) upper limit is found to be $2.7 (3.9_{-1.1}^{+2.1})$ pb at the 95% C.L. This corresponds to an observed (expected) upper limit on $\mu$ at the 95% C.L. of $110 (150_{-80}^{+40})$. The uncertainties in the expected limits correspond to the $\pm 1\sigma$ interval of background-only pseudoexperiments. With the current sensitivity, the result depends weakly on the assumption of the SM rate for $H \rightarrow b\bar{b}$. The observed limit remains within 5% of the nominal value when the assumed value for normalization of the $ZH(b\bar{b})$ background is varied from zero to twice the SM prediction.

A search for the decay of the Higgs boson to charm quarks has been performed using 36.1 fb$^{-1}$ of data collected with the ATLAS detector in $pp$ collisions at $\sqrt{s} = 13$ TeV at the LHC. No significant excess of $ZH(c\bar{c})$ production is observed over the SM background expectation. The observed upper limit on $\sigma(pp\rightarrow ZH) \times B(H \rightarrow c\bar{c})$ is 2.7 pb at the 95% C.L. The corresponding expected upper limit is $3.9_{-1.1}^{+2.1}$ pb. This is the most stringent limit to date in direct searches for the inclusive decay of the Higgs boson to charm quarks.

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/IRFU, France; SRNSFG,
Georgia; BMBF, HGF, and MPG, Germany; GSRT, Greece; RGC, Hong Kong SAR, China; ISF, I-CORE and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MINISW and NDN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC NI, Russian Federation; JINR, MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MINECO, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, the Canada Council, CANARIE, CRC, Compute Canada, FQRNT, and the Ontario Innovation Trust, Canada; EPLANET, ERC, ERDF, FP7, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex and Idex, ANR, Région Auvergne and Fondation Partager le Savoir, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; BRF, GIF and Minerva, Israel; BRF, Norway; CERCA Programme Generalitat de Catalunya, Generalitat Valenciana, Spain; the Royal Society and Leverhulme Trust, United Kingdom. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [71].

[24] 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 upwards. Cylindrical coordinates $(r, \phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the z axis. The pseudorapidity is defined in terms of
the polar angle \( \theta \) as \( \eta = -\ln \tan(\theta/2) \). Angular distance is measured in units of \( \Delta R = \sqrt{\Delta \eta^2 + (\Delta \phi)^2} \).


