CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch 2022/02/28

Direct search for the standard model Higgs boson decaying to a charm quark-antiquark pair

The CMS Collaboration

Abstract

A search for $H \to c\bar{c}$ in associated VH production with leptonically decaying V (W or Z) boson is performed with $138\,fb^{-1}$ of proton-proton collisions at $\sqrt{s} = 13\,TeV$ in the CMS experiment. Novel charm jet identification and analysis techniques are used. A search for $Z \to c\bar{c}$ in VZ events is used to validate the strategy and yields a first observation at a hadron collider with a significance of 5.7 standard deviations. The analysis yields $\sigma\,(VH)\,\mathcal{B}\,(H \to c\bar{c}) < 0.94\,pb$ at 95% CL. For κ_c, the Higgs-charm Yukawa coupling modifier, the observed 95% CL interval (expected upper limit) is $1.1 < |\kappa_c| < 5.5$ ($|\kappa_c| < 3.4$) — the most stringent to date.

The discovery of a Higgs boson (H) with the LHC Run–1 data by the ATLAS [1] and CMS [2] experiments in 2012, confirms the electroweak symmetry breaking mechanism. The measured mass is $m_{\mathrm{H}} = 125.38 \pm 0.14\,\mathrm{GeV}$ [3]. The observation of Higgs boson decays to gauge bosons, to third generation fermions [4–14] and all measured properties [3, 15–23] support the hypothesis that this new particle is very standard model (SM) like. However, there is more to be learned. A high priority of the LHC physics program is the measurement of the couplings of the Higgs boson to other SM fields. Recently, the CMS Collaboration reported the first evidence of Higgs boson decays to second generation leptons [24]. The next milestone is the coupling to second generation quarks. In this note we focus on the search for Higgs bosons decaying to a charm quark and antiquark, c and $\bar{\mathrm{c}}$, respectively. The corresponding Yukawa coupling, y_{c}, can be significantly modified in the presence of physics beyond the SM [25–28]. However, the smallness of the branching ratio, very large QCD multijet background and difficulty of identifying charm quark jets in a hadronic environment makes this a very challenging measurement at the LHC. The direct searches for $\mathrm{H} \to \mathrm{c\bar{c}}$ at the LHC, reported in [29–31] by the ATLAS and CMS collaborations, target the associated production of a Higgs boson with a $\mathrm{V} = \mathrm{W}$ or Z boson. Using $139\,\mathrm{fb}^{-1}$ of data at 13 TeV, the most recent search by the ATLAS collaboration obtains an observed (expected) upper limit on the product of the production cross section $\sigma\,(\mathrm{VH})$ and branching fraction $\mathcal{B}\,(\mathrm{H} \to \mathrm{c\bar{c}})$ of 26 (31) times the SM prediction at 95% confidence level (CL) [31].

This note presents the CMS Run–2 legacy analysis using $138\,\mathrm{fb}^{-1}$ of proton-proton (pp) collision data at 13 TeV, to significantly improve upon the results of the previous CMS analysis [30]. Building on the study reported in [30], a set of c jet reconstruction and identification algorithms, and sophisticated analysis techniques based upon machine learning (ML) techniques, have been developed.

The CMS apparatus [32] is a multipurpose, nearly hermetic detector, designed to trigger on [33, 34] and identify electrons, muons, photons, and (charged and neutral) hadrons [35–38]. A global reconstruction "particle-flow" (PF) algorithm [39] combines the information provided by the all-silicon inner tracker and by the crystal electromagnetic and brass-scintillator hadron calorimeters, operating inside a 3.8 T superconducting solenoid, with data from gas-ionization muon detectors interleaved with the solenoid return yoke, to build τ leptons, jets, missing transverse momentum, and other physics objects [40–42].

Two collections of jets are used in the search. In the first, PF candidates are clustered using the anti-k_{T} algorithm [43, 44] with a distance parameter $R = 0.4$, and will be referred to as "small-R jets". The second uses $R = 1.5$ and contains what are referred to as "large-R jets". Contamination due to additional pp interactions within the same or nearby bunch crossings (pileup) is mitigated via the charged hadron subtraction algorithm [39] for small-R jets, and via the PUPPI [45, 46] algorithm for large-R jets. A regression algorithm [47] is developed to improve the large-R jet mass reconstruction. The algorithm exploits properties of the PF candidates and secondary vertices associated to the jet using the ParticleNet graph neural network [48]. The mass resolution is improved by ≈50% over traditional jet grooming algorithms [49, 50]. The small-R (large-R) jets are required to have $p_{\mathrm{T}} > 25(200)\,\mathrm{GeV}$ and to be within the tracker acceptance, $|\eta| < 2.4$.

Signal and background processes are simulated using various Monte Carlo (MC) event generators, while the CMS detector response is modeled with GEANT4 [51]. The quark-induced ZH and WH signal processes are generated at next-to-leading order (NLO) accuracy in QCD using the POWHEG v2 [52–54] event generator extended with the multi-scale improved NLO (MiNLO) procedure [55, 56], while the gluon-induced ZH process is generated at leading order (LO) accuracy with POWHEG v2. The Higgs boson mass is set to 125 GeV for all simulated signal samples. The production cross sections of the signal processes [57] are corrected as a function of $p_{\mathrm{T}}(\mathrm{V})$ to next-to-next-to-leading order (NNLO) QCD + NLO electroweak (EW) accuracy combining the VHNNLO [58–61], VH@NNLO [62, 63], and HAWK v2.0 [64] generators as described in [57].

The V+jets background samples are generated with MADGRAPH5_aMC@NLO v2.6.0 [65] at NLO with up to two additional partons. The top quark pair ($\mathrm{t\bar{t}}$) [66] and single top quark production processes [67–69] are generated to NLO accuracy with POWHEG v2. The production cross sections for the $\mathrm{t\bar{t}}$ samples are scaled to the NNLO prediction with the next-to-next-to-leading-log result obtained from TOP++ v2.0 [70], and the differential cross sections as a function of top quark p_{T} are corrected to the NNLO QCD + NLO EW prediction [71]. Diboson backgrounds are generated at NLO with POWHEG v2 (MADGRAPH5_aMC@NLO v2.4.2) for the WW [72] (WZ and ZZ) process. Production cross sections of the diboson processes are reweighted as a function of the subleading vector boson p_{T} to NNLO QCD + NLO EW accuracy [73].

The NLO NNPDF3.0 [74] (NNLO NNPDF3.1 [75]) parton distribution function (PDF) set is used for the 2016 (2017–2018) simulations. For parton showering and hadronization, including the $\mathrm{H} \to \mathrm{c\bar{c}}$ decay, the matrix element generators are interfaced with PYTHIA v8.230 [76] with the CUETP8M1 [77] (CP5 [78]) underlying event tune for the 2016 (2017–2018) samples. The matching of jets from matrix element calculations and those from parton showers is done with the FxFx [79] (MLM [80]) prescription for NLO (LO) samples. For all samples, pileup interactions are simulated and added to the hard-scattering process. Events are then reweighted to match the pileup profile observed in data.

The analysis is carried out in mutually exclusive channels targeting leptonic decays of the vector bosons: $\mathrm{Z} \to \nu\nu$, $\mathrm{W} \to \ell\nu$, and $\mathrm{Z} \to \ell\ell$, where ℓ is an electron or a muon, and referred to as the 0L, 1L and 2L channels. Events are collected using triggers based on missing transverse momentum, one, or two leptons. The event selection criteria are detailed in [30].

As was true in [30], the analysis reconstructs the Higgs boson candidate ($\mathrm{H_{cand}}$) under the assumption of either a “merged-jet” topology, in which the hadronization products of the two charm quarks are reconstructed as a single large-R jet, or a “resolved-jet” topology, in which the $\mathrm{H_{cand}}$ is reconstructed from two well-separated and individually resolved c jets. In practice, the two topologies can have significant overlap and are made distinct by defining them in reference to whether a given $\mathrm{H_{cand}}$, identified through a large-R jet in the event, has p_{T} above or below a threshold of 300 GeV, chosen to maximize the sensitivity to the $\mathrm{VH}(\mathrm{H} \to \mathrm{c\bar{c}})$ process.

The merged-jet analysis targets moderately to highly Lorentz-boosted Higgs bosons, as flagged by $p_{\mathrm{T}}(\mathrm{H}) \gtrsim 300\,\mathrm{GeV}$, where the decay products are contained in a single large-R jet. On average, Higgs bosons in the signal process have larger p_{T} than those from the V+jets and $\mathrm{t\bar{t}}$ backgrounds. Thus, the high-p_{T} regime explored in this topology, although only about 5% of the available signal cross section, provides powerful sensitivity to the search. Moreover, the possibility for both c quarks to reside in a single large-R jet enhances the signal acceptance, improves the identification of the correct pair of c jets to use to reconstruct the Higgs boson, and facilitates the task of taking into account final-state radiation (FSR) emitted by the quarks. A more detailed discussion of potential advantages can be found in [30, 50, 81].

State-of-the-art performance in the reconstruction and identification of the pair of c quarks from the Higgs boson decay is achieved with ParticleNet [48], a novel jet identification algorithm. Using PF candidates and secondary vertices associated to large-R jets as inputs, Parti-

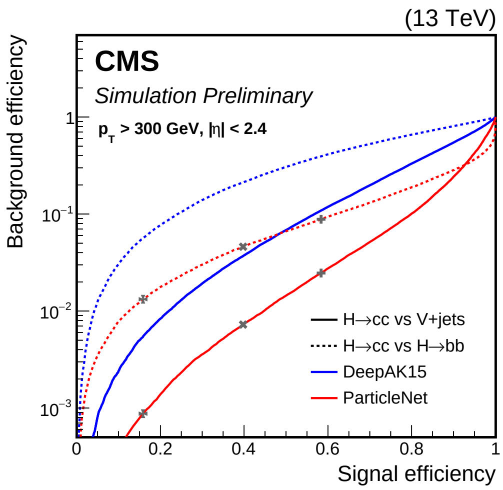

Figure 1: Performance of ParticleNet (red lines) for identifying a $c\bar{c}$ pair for large-R jets with $p_T > 300$ GeV. The solid (dashed) line shows the efficiency to correctly identify $H \to c\bar{c}$ vs. the efficiency of misidentifying quarks or gluons from the V+jets process (vs. $H \to b\bar{b}$). The gray crosses represent the three working points used in the merged-jet analysis. The performance of DeepAK15 (blue lines) used in [30] is shown for comparison.

cleNet exploits information related to jet substructure, flavor, and pileup simultaneously with an advanced graph neural network, yielding substantial gains over other approaches [82, 83]. Decorrelation of the algorithm's response with the jet mass is achieved by training it with a dedicated set of MC samples produced with the same jet mass distributions for the signal and background processes [82]. The performance of the $c\bar{c}$ discriminant for identifying a pair of c quarks from Higgs boson decay for large-R jets with $p_T > 300$ GeV is shown in Fig. 1. ParticleNet is compared to the previous state-of-the-art $c\bar{c}$ discriminant "DeepAK15" [30, 84] yielding improvement of a factor of 4 to 7 in the rejection of other jet flavors. Three working points (WPs) are defined on the $c\bar{c}$ discriminant distribution with approximately 58, 40, and 16% efficiencies for identifying a $c\bar{c}$ pair. The corresponding misidentification rates of light quark and gluon jets (bottom jets) are 2(9), 0.7(5), and 0.08(1)%. These WPs separate events into three mutually exclusive categories with different $c\bar{c}$ purity to improve the sensitivity of the analysis. The $c\bar{c}$ identification is calibrated using a data sample enriched in gluons splitting to $c\bar{c}$. The p_T-dependent data-to-simulation ratios (used as corrective scale factors below) are typically 0.9–1.3 with corresponding uncertainties of 20–30%.

The main backgrounds, $t\bar{t}$ and V+jets, are further suppressed by developing a separate boosted decision tree (BDT) for each channel, using event-level kinematical variables as inputs. The BDT design relies heavily on previous developments [30] with improvements in variable selection and training procedure, leading to ≈15% enhancement of the sensitivity of the analysis.

The signal regions (SRs) in the 0L and 2L channels are defined for BDT values greater than 0.55, whereas two SRs with BDT values 0.55–0.7 and >0.7, are defined in the 1L channel. Events in the SRs are further subdivided into the three $c\bar{c}$ discriminant categories mentioned above. As both the BDTs and the $c\bar{c}$ discriminant are designed to be largely independent of the H_{cand} mass ($m\,(H_{cand})$), the $m\,(H_{cand})$ distributions are used to separate signal and background contributions in each SR.

More than 95% of the VH events have a Higgs boson with $p_T(H) < 300$ GeV, corresponding to the phase space region where the Higgs boson decay products generally give rise to two distinctly reconstructed small-R jets. The resolved-jet topology exploits a large fraction of this phase space which, however, contains higher background contamination than that used in the merged-jet analysis. The H_{cand} is reconstructed as two distinct small-R jets. The identification of c jets relies on the ML-based DeepJet algorithm [85, 86]. The discrimination between c jets and light quark or gluon jets (b jets) is achieved via the probability ratio defined as *CvsL* (*CvsB*) [87]. Potential differences in the discriminant shapes between data and simulation are accounted for with flavor-dependent simulation-to-data scale factors [87], which are typically in the 0.9–1.0 range. The corresponding uncertainties range from 2% for light quark and gluon jets or b jets, to 5% for c jets.

The two jets with the highest *CvsL* discriminants in each event are selected to reconstruct the H_{cand} four-vector. To improve sample purity, criteria are imposed on *CvsL* and *CvsB* for the leading jet that correspond to ≈40% c jet efficiency, for ≈4% (≈16%) light quark or gluon jet (b jet) misidentification rate. A number of steps are then taken to improve the $m\left(H_{cand}\right)$ reconstruction. To account for an underestimation of the c jet energy, due to the potential presence of undetected neutrinos in c hadron decays, an ML-based jet energy regression algorithm [88], originally developed for b jets, is utilized. In addition, small-R jets reconstructed in the vicinity of H_{cand} jets, often stemming from FSR, are included in the H_{cand} reconstruction. Furthermore, the $m\left(H_{cand}\right)$ resolution in the 2L channel is improved via a kinematic fit by balancing the momenta of the two small-R jets and the lepton pair within experimental uncertainties [89]. These steps improve the $m\left(H_{cand}\right)$ resolution up to 20%. Finally, a BDT is developed to maximize the discrimination power between signal and background processes in each channel, using event-level kinematical variables, c jet identification discriminants and properties of H_{cand} as inputs [30].

The signal strength modifier μ, defined as $(\sigma\mathcal{B})_{obs}$ / $(\sigma\mathcal{B})_{SM}$ where σ indicates the signal production cross section and $\mathcal{B}$ is the branching fraction, is measured via a binned maximum likelihood fit to data. The fitted variable is $m\left(H_{cand}\right)$ in the merged-jet analysis, and the BDT discriminant in the resolved-jet analysis. The normalizations of the main backgrounds, namely V+jets and $t\bar{t}$, are estimated by including dedicated control regions in the fit following the strategy detailed in [30]. Contributions from single top, diboson and $VH(H \to b\bar{b})$ processes are estimated from simulation. Due to improvements in charm identification, the $VH(H \to b\bar{b})$ contribution in the high-purity signal regions is suppressed to about twice that expected for $VH(H \to c\bar{c})$, largely negligible compared to the leading backgrounds. Similarly, the $VZ(Z \to b\bar{b})$ yield is about 10% of that expected for $VZ(Z \to c\bar{c})$. Contributions from $H \to \tau\tau$ decay are negligible after the above mentioned selection criteria, and thus are not considered.

Systematic uncertainties affecting normalizations and shapes of fitted variables are taken into account via nuisance parameters. Table 1 shows the relative impact of each uncertainty on the fitted μ. The leading uncertainty is statistical because of the limited number of events in the SRs as well as the control regions used to extract background normalizations. The main experimental systematic uncertainties are associated with limited simulated sample sizes and the charm identification efficiencies, representing ≈37% and ≈23% of the total, respectively. Theoretical uncertainties in the cross sections, p_T spectra, PDFs, renormalization and factorization scales, represent ≈22% of the total uncertainty in μ.

The analysis is validated by a search for the analogous SM process $VZ(Z \to c\bar{c})$. The BDTs in the resolved-jet topology are modified to treat the $VZ(Z \to c\bar{c})$ process instead of the $VH(H \to c\bar{c})$ process as signal. No modification is needed for the BDTs in the merged-jet topology. The best fit μ of this process is $\mu_{VZ(Z\to c\bar{c})} = 1.01^{+0.23}_{-0.21}$, in agreement with the SM

Table 1: The relative contributions to the total uncertainty on the signal strength modifier μ for the $\mathrm{VH}(\mathrm{H} \to \mathrm{c}\bar{\mathrm{c}})$ process.

Uncertainty source	$\Delta\mu / (\Delta\mu)_{\mathrm{tot}}$
Statistical	**85%**
Background normalizations	37%
Experimental	**48%**
Sizes of the simulated samples	37%
Charm identification efficiencies	23%
Jet energy scale and resolution	15%
Simulation modeling	11%
Luminosity	6%
Lepton identification efficiencies	4%
Theory	**22%**
Backgrounds	17%
Signal	15%

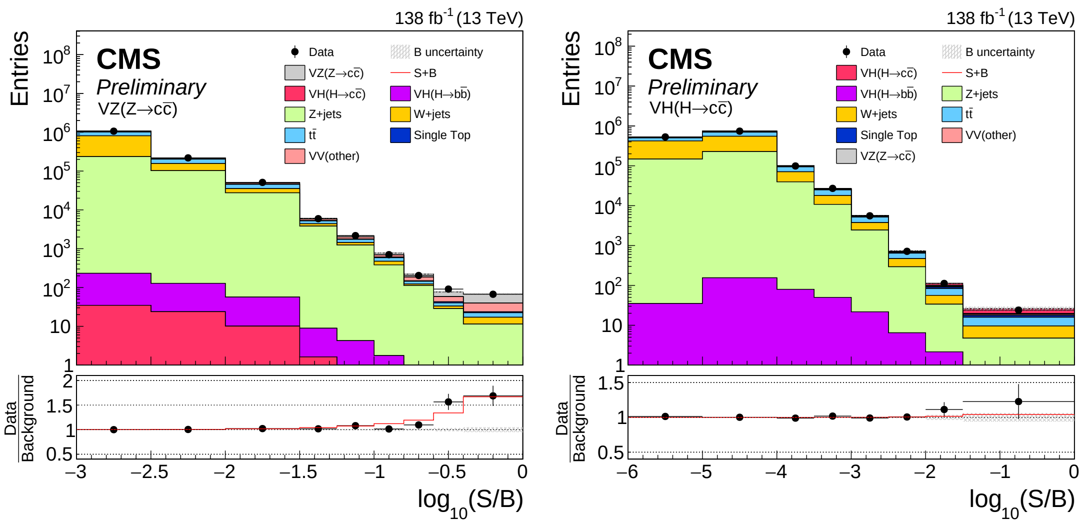

Figure 2: Distribution of events as a function of S/B in the $\mathrm{VZ}(\mathrm{Z} \to \mathrm{c}\bar{\mathrm{c}})$ (left) and $\mathrm{VH}(\mathrm{H} \to \mathrm{c}\bar{\mathrm{c}})$ (right) searches, where S and B are the postfit signal and background yields, respectively, in each bin of the fitted $m\,(\mathrm{H}_{\mathrm{cand}})$ or BDT discriminant distributions. The bottom panel shows the ratio of data to the total background, with the uncertainty in background indicated by gray hatching. The red line represents background plus SM signal divided by background.

expectation. Figure 2 (left) shows the distribution of events in all channels, sorted into bins of similar signal-to-background ratios. The observed data shows a visible excess over the expected non-$\mathrm{VZ}(\mathrm{Z} \to \mathrm{c}\bar{\mathrm{c}})$ backgrounds. The significance of the excess is computed using the asymptotic distribution of a test statistic based on the profile likelihood ratio [90, 91]. The observed (expected) significance is 4.4 (4.7) standard deviations for the merged-jet analysis, 3.1 (3.3) standard deviations for the resolved-jet analysis, and 5.7 (5.9) standard deviations for their combination. *This is the first observation of $Z \to c\bar{c}$ at a hadron collider.*

Figure 2 (right) compares the observed data to the SM prediction in the search of $\mathrm{VH}(\mathrm{H} \to \mathrm{c}\bar{\mathrm{c}})$, where the best fit μ is $\mu_{\mathrm{VH}(\mathrm{H}\to\mathrm{c}\bar{\mathrm{c}})} = 7.7^{+3.8}_{-3.5}$. The fitted $m\,(\mathrm{H}_{\mathrm{cand}})$ distribution in the merged-jet topology is displayed in Fig. 3. No significant excess over the background-only hypothesis is observed. An upper limit on $\mu_{\mathrm{VH}(\mathrm{H}\to\mathrm{c}\bar{\mathrm{c}})}$ is extracted via the asymptotic $\mathrm{CL_s}$ method [90–93]. The observed (expected) 95% CL upper limit on $\mu_{\mathrm{VH}(\mathrm{H}\to\mathrm{c}\bar{\mathrm{c}})}$ is 14 ($7.6^{+3.4}_{-2.3}$), which is equivalent to

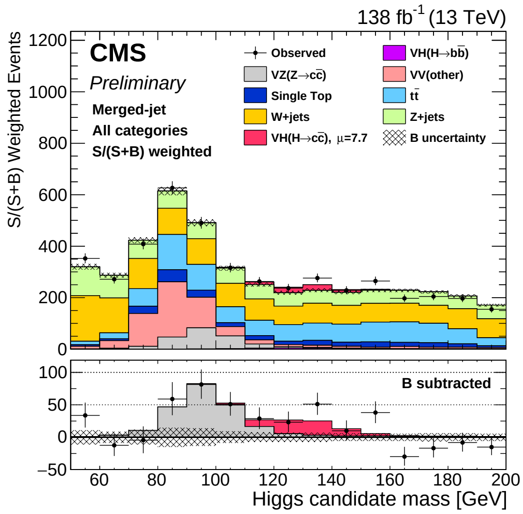

Figure 3: Combined $m\left(\mathrm{H_{cand}}\right)$ distribution in all channels of the merged-jet analysis. The fitted $m\left(\mathrm{H_{cand}}\right)$ distribution in each signal region is weighted by $S/(S+B)$, where S and B are the postfit $\mathrm{VH(H \to c\bar{c})}$ signal and total background yields. The lower panel shows data (points) and the fitted $\mathrm{VH(H \to c\bar{c})}$ (red) and $\mathrm{VZ(Z \to c\bar{c})}$ background (grey) distributions after subtracting all other processes. Error bars represent pre-subtraction statistical uncertainties on data, while the gray hatching indicates the total uncertainty on the signal and all background processes.

an observed (expected) upper limit on $\sigma\left(\mathrm{VH}\right)\mathcal{B}\left(\mathrm{H \to c\bar{c}}\right)$ of 0.94 ($0.50^{+0.22}_{-0.15}$) pb. Contributions of the individual channels are summarized in Fig. 4.

The result is interpreted in the κ-framework [57, 94] by reparameterizing $\mu_{\mathrm{VH(H \to c\bar{c})}}$ in terms of the Higgs-charm Yukawa coupling modifier κ_c, assuming only the Higgs boson decay widths are altered:

$$\mu_{\mathrm{VH(H \to c\bar{c})}} = \frac{\kappa_c^2}{1 + \mathcal{B}_{\mathrm{SM}}\left(\mathrm{H \to c\bar{c}}\right) \times \left(\kappa_c^2 - 1\right)}. \tag{1}$$

The observed 95% CL interval is $1.1 < |\kappa_c| < 5.5$, and the expected upper limit is $|\kappa_c| < 3.4$.

In summary, a direct search for the SM Higgs boson decaying to a pair of charm quarks in the CMS experiment is presented. Novel jet reconstruction and identification tools and analysis techniques are developed for this analysis, which is validated by measuring the $\mathrm{VZ(Z \to c\bar{c})}$ process. The observed Z boson signal relative to the SM prediction is $\mu_{\mathrm{VZ(Z \to c\bar{c})}} = 1.01^{+0.23}_{-0.21}$, with an observed (expected) significance of 5.7 (5.9) standard deviations. This is the first observation of $\mathrm{Z \to c\bar{c}}$ at a hadronic collider.

The observed (expected) upper limit on $\sigma\left(\mathrm{VH}\right)\mathcal{B}\left(\mathrm{H \to c\bar{c}}\right)$ is 0.94 ($0.50^{+0.22}_{-0.15}$) pb, corresponding to 14 ($7.6^{+3.4}_{-2.3}$) times the theoretical prediction for an SM Higgs boson mass of 125.38 GeV. The observed 95% CL interval on the modifier, κ_c, for the Yukawa coupling of the Higgs boson to the charm quark is $1.1 < |\kappa_c| < 5.5$, and the expected upper limit is $|\kappa_c| < 3.4$. This is the most stringent limit on κ_c to date.

References

[1] ATLAS Collaboration, "Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1,

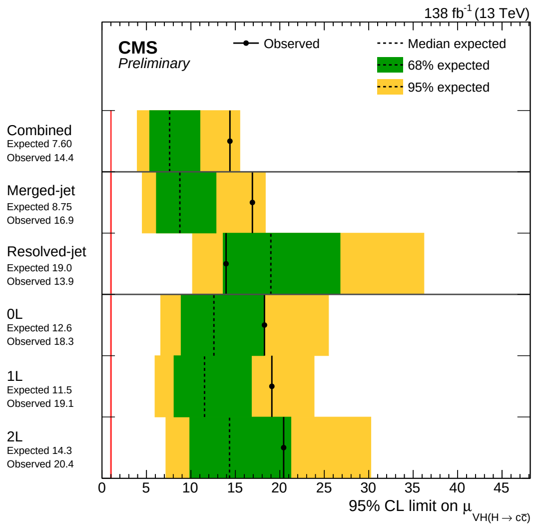

Figure 4: The 95% CL upper limits on $\mu_{\mathrm{VH(H\to c\bar{c})}}$. Green and yellow bands indicate the expected 68% and 95% CL regions, respectively, under the background-only hypothesis. The vertical red line indicates the SM value $\mu_{\mathrm{VH(H\to c\bar{c})}} = 1$.

`doi:10.1016/j.physletb.2012.08.020`, `arXiv:1207.7214`.

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30, `doi:10.1016/j.physletb.2012.08.021`, `arXiv:1207.7235`.

[3] CMS Collaboration, "A measurement of the Higgs boson mass in the diphoton decay channel", *Phys. Lett. B* **805** (2020) 135425, `doi:10.1016/j.physletb.2020.135425`, `arXiv:2002.06398`.

[4] ATLAS Collaboration, "Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector", *Phys. Rev. D* **90** (2014) 112015, `doi:10.1103/PhysRevD.90.112015`, `arXiv:1408.7084`.

[5] CMS Collaboration, "Observation of the diphoton decay of the Higgs boson and measurement of its properties", *Eur. Phys. J. C* **74** (2014) 3076, `doi:10.1140/epjc/s10052-014-3076-z`, `arXiv:1407.0558`.

[6] ATLAS Collaboration, "Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector", *Phys. Rev. D* **91** (2015) 012006, `doi:10.1103/PhysRevD.91.012006`, `arXiv:1408.5191`.

[7] CMS Collaboration, "Measurement of the properties of a Higgs boson in the four-lepton final state", *Phys. Rev. D* **89** (2014) 092007, `doi:10.1103/PhysRevD.89.092007`, `arXiv:1312.5353`.

[8] ATLAS Collaboration, "Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector", *Phys. Rev. D* **92** (2015) 012006, `doi:10.1103/PhysRevD.92.012006`, `arXiv:1412.2641`.

[9] ATLAS Collaboration, "Study of (W/Z)H production and Higgs boson couplings using $H \to WW^*$ decays with the ATLAS detector", *JHEP* **08** (2015) 137, doi:10.1007/JHEP08(2015)137, arXiv:1506.06641.

[10] CMS Collaboration, "Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states", *JHEP* **01** (2014) 096, doi:10.1007/JHEP01(2014)096, arXiv:1312.1129.

[11] ATLAS Collaboration, "Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector", *JHEP* **04** (2015) 117, doi:10.1007/JHEP04(2015)117, arXiv:1501.04943.

[12] CMS Collaboration, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons", *JHEP* **05** (2014) 104, doi:10.1007/JHEP05(2014)104, arXiv:1401.5041.

[13] CMS Collaboration, "Observation of the Higgs boson decay to a pair of tau leptons", *Phys. Lett. B* **779** (2017) 283, doi:10.1016/j.physletb.2018.02.004, arXiv:1708.00373.

[14] CMS Collaboration, "Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **791** (2019) 96, doi:10.1016/j.physletb.2018.12.073, arXiv:1806.05246.

[15] ATLAS Collaboration, "Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment", *Eur. Phys. J. C* **76** (2016) 6, doi:10.1140/epjc/s10052-015-3769-y, arXiv:1507.04548.

[16] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", *Eur. Phys. J. C* **75** (2015) 212, doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.

[17] CMS Collaboration, "Study of the mass and spin-parity of the Higgs boson candidate via its decays to Z boson pairs", *Phys. Rev. Lett.* **110** (2013) 081803, doi:10.1103/PhysRevLett.110.081803, arXiv:1212.6639.

[18] ATLAS Collaboration, "Evidence for the spin-0 nature of the Higgs boson using ATLAS data", *Phys. Lett. B* **726** (2013) 120, doi:10.1016/j.physletb.2013.08.026, arXiv:1307.1432.

[19] ATLAS and CMS Collaborations, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.

[20] ATLAS and CMS Collaborations, "Combined measurement of the Higgs boson mass in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments", *Phys. Rev. Lett.* **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803, arXiv:1503.07589.

[21] CMS Collaboration, "Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV", *JHEP* **11** (2017) 047, doi:10.1007/JHEP11(2017)047, arXiv:1706.09936.

[22] ATLAS Collaboration, "Combined measurement of differential and total cross sections in the $H \to \gamma\gamma$ and the $H \to ZZ^* \to 4\ell$ decay channels at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B* **786** (2018) 114, `doi:10.1016/j.physletb.2018.09.019`, `arXiv:1805.10197`.

[23] ATLAS Collaboration, "Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector", *Phys. Lett. B* **784** (2018) 345, `doi:10.1016/j.physletb.2018.07.050`, `arXiv:1806.00242`.

[24] CMS Collaboration, "Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.* **122** (2019) 021801, `doi:10.1103/PhysRevLett.122.021801`, `arXiv:1807.06325`.

[25] D. Ghosh, R. S. Gupta, and G. Perez, "Is the Higgs mechanism of fermion mass generation a fact? A Yukawa-less first-two-generation model", *Phys. Lett. B* **755** (2016) 504, `doi:10.1016/j.physletb.2016.02.059`, `arXiv:1508.01501`.

[26] F. J. Botella, G. C. Branco, M. N. Rebelo, and J. I. Silva-Marcos, "What if the masses of the first two quark families are not generated by the standard model Higgs boson?", *Phys. Rev. D* **94** (2016) 115031, `doi:10.1103/PhysRevD.94.115031`, `arXiv:1602.08011`.

[27] R. Harnik, J. Kopp, and J. Zupan, "Flavor violating Higgs decays", *JHEP* **03** (2013) 026, `doi:10.1007/JHEP03(2013)026`, `arXiv:1209.1397`.

[28] W. Altmannshofer et al., "Collider signatures of flavorful Higgs bosons", *Phys. Rev. D* **94** (2016) 115032, `doi:10.1103/PhysRevD.94.115032`, `arXiv:1610.02398`.

[29] ATLAS Collaboration, "Search for the decay of the Higgs boson to charm quarks with the ATLAS experiment", *Phys. Rev. Lett.* **120** (2018) 211802, `doi:10.1103/PhysRevLett.120.211802`, `arXiv:1802.04329`.

[30] CMS Collaboration, "A search for the standard model Higgs boson decaying to charm quarks", *JHEP* **03** (2020) 131, `doi:10.1007/JHEP03(2020)131`, `arXiv:1912.01662`.

[31] ATLAS Collaboration, "Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector", 1, 2022. `arXiv:2201.11428`. Submitted to *EPJC*.

[32] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, `doi:10.1088/1748-0221/3/08/S08004`.

[33] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **15** (2020) P10017, `doi:10.1088/1748-0221/15/10/P10017`, `arXiv:2006.10165`.

[34] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, `doi:10.1088/1748-0221/12/01/P01020`, `arXiv:1609.02366`.

[35] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, `doi:10.1088/1748-0221/10/06/P06005`, `arXiv:1502.02701`.

[36] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018) P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[37] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV", *JINST* **10** (2015) P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[38] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[39] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[40] CMS Collaboration, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018) P10005, doi:10.1088/1748-0221/13/10/P10005, arXiv:1809.02816.

[41] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[42] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *JINST* **14** (2019) P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

[43] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{T} jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[44] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[45] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *JINST* **15** (2020) P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503.

[46] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup per particle identification", *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[47] CMS Collaboration, "Mass regression of highly-boosted jets using graph neural networks", CMS Detector Performance Note CMS-DP-2021-017, 2021.

[48] H. Qu and L. Gouskos, "ParticleNet: Jet Tagging via Particle Clouds", *Phys. Rev. D* **101** (2020) 056019, doi:10.1103/PhysRevD.101.056019, arXiv:1902.08570.

[49] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, "Towards an understanding of jet substructure", *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.

[50] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new Higgs search channel at the LHC", *Phys. Rev. Lett.* **100** (2008) 242001, doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.

[51] GEANT4 Collaboration, "GEANT4 — A simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[52] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[53] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[54] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[55] K. Hamilton, P. Nason, and G. Zanderighi, "MINLO: Multi-scale improved NLO", *JHEP* **10** (2012) 155, doi:10.1007/JHEP10(2012)155, arXiv:1206.3572.

[56] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, "$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO", *JHEP* **10** (2013) 083, doi:10.1007/JHEP10(2013)083, arXiv:1306.2542.

[57] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector", *CERN* (2016) doi:10.23731/CYRM-2017-002, arXiv:1610.07922.

[58] G. Ferrera, M. Grazzini, and F. Tramontano, "Higher-order QCD effects for associated WH production and decay at the LHC", *JHEP* **04** (2014) 039, doi:10.1007/JHEP04(2014)039, arXiv:1312.1669.

[59] G. Ferrera, M. Grazzini, and F. Tramontano, "Associated ZH production at hadron colliders: the fully differential NNLO QCD calculation", *Phys. Lett. B* **740** (2015) 51, doi:10.1016/j.physletb.2014.11.040, arXiv:1407.4747.

[60] G. Ferrera, M. Grazzini, and F. Tramontano, "Associated WH production at hadron colliders: a fully exclusive QCD calculation at NNLO", *Phys. Rev. Lett.* **107** (2011) 152003, doi:10.1103/PhysRevLett.107.152003, arXiv:1107.1164.

[61] G. Ferrera, G. Somogyi, and F. Tramontano, "Associated production of a Higgs boson decaying into bottom quarks at the LHC in full NNLO QCD", *Phys. Lett. B* **780** (2018) 346, doi:10.1016/j.physletb.2018.03.021, arXiv:1705.10304.

[62] O. Brein, R. V. Harlander, and T. J. E. Zirke, "vh@nnlo — Higgs Strahlung at hadron colliders", *Comput. Phys. Commun.* **184** (2013) 998, doi:10.1016/j.cpc.2012.11.002, arXiv:1210.5347.

[63] R. V. Harlander, S. Liebler, and T. Zirke, "Higgs Strahlung at the Large Hadron Collider in the 2-Higgs-doublet model", *JHEP* **02** (2014) 023, doi:10.1007/JHEP02(2014)023, arXiv:1307.8122.

[64] A. Denner, S. Dittmaier, S. Kallweit, and A. Mück, "HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders", *Comput. Phys. Commun.* **195** (2015) 161, doi:10.1016/j.cpc.2015.04.021, arXiv:1412.5390.

[65] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[66] S. Frixione, P. Nason, and G. Ridolfi, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.

[67] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J. C* **71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[68] R. Frederix, E. Re, and P. Torrielli, "Single-top *t*-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO", *JHEP* **09** (2012) 130, doi:10.1007/JHEP09(2012)130, arXiv:1207.5391.

[69] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO single-top production matched with shower in POWHEG: *s*- and *t*-channel contributions", *JHEP* **09** (2009) 111, doi:10.1088/1126-6708/2009/09/111, arXiv:0907.4076. [Erratum: doi:10.1007/JHEP02(2010)011].

[70] M. Czakon and A. Mitov, "Top++: A program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[71] M. Czakon et al., "Top-pair production at the LHC through NNLO QCD and NLO EW", *JHEP* **10** (2017) 186, doi:10.1007/JHEP10(2017)186, arXiv:1705.04105.

[72] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, "W+W-, WZ and ZZ production in the POWHEG BOX", *JHEP* **11** (2011) 078, doi:10.1007/JHEP11(2011)078, arXiv:1107.5051.

[73] M. Grazzini et al., "NNLO QCD + NLO EW with Matrix+OpenLoops: precise predictions for vector-boson pair production", *JHEP* **02** (2020) 087, doi:10.1007/JHEP02(2020)087, arXiv:1912.00068.

[74] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[75] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

[76] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[77] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[78] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[79] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[80] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[81] T. Plehn, G. P. Salam, and M. Spannowsky, "Fat jets for a light Higgs", *Phys. Rev. Lett.* **104** (2010) 111801, doi:10.1103/PhysRevLett.104.111801, arXiv:0910.5472.

[82] CMS Collaboration, "Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques", CMS Detector Performance Note CMS-DP-2020-002, 2020.

[83] A. Butter et al., "The Machine Learning Landscape of Top Taggers", *SciPost Phys.* **7** (2019) 014, doi:10.21468/SciPostPhys.7.1.014, arXiv:1902.09914.

[84] CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", *JINST* **15** (2020) P06005, doi:10.1088/1748-0221/15/06/P06005, arXiv:2004.08262.

[85] E. Bols et al., "Jet Flavour Classification Using DeepJet", *JINST* **15** (2020) P12012, doi:10.1088/1748-0221/15/12/P12012, arXiv:2008.10519.

[86] CMS Collaboration, "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector", CMS Detector Performance Note CMS-DP-2018-058, 2018.

[87] CMS Collaboration, "A new calibration method for charm jet identification validated with proton-proton collision events at $\sqrt{s}$ =13 TeV", 11, 2021. arXiv:2111.03027. Submitted to *JINST*.

[88] CMS Collaboration, "A Deep Neural Network for Simultaneous Estimation of b Jet Energy and Resolution", *Comput. Softw. Big Sci.* **4** (2020) 10, doi:10.1007/s41781-020-00041-z, arXiv:1912.06046.

[89] CMS Collaboration, "Observation of Higgs boson decay to bottom quarks", *Phys. Rev. Lett.* **121** (2018) 121801, doi:10.1103/PhysRevLett.121.121801, arXiv:1808.08242.

[90] The ATLAS and CMS Collaborations, The LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, 2011.

[91] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: doi:10.1140/epjc/s10052-013-2501-z].

[92] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

[93] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[94] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 3. Higgs properties", *CERN* (2013) doi:10.5170/CERN-2013-004, arXiv:1307.1347.