IL NUOVO CIMENTO **47 C** (2024) 142
DOI 10.1393/ncc/i2024-24142-8

Colloquia: IFAE 2023

Search for the Higgs boson decay to charm quarks at the CMS experiment(*)

A. Zaza(1)(2) on behalf of the CMS Collaboration

(1) *INFN, Sezione di Bari - Bari, Italy*
(2) *Dipartimento di Fisica, Università di Bari - Bari, Italy*

received 13 February 2024

Summary. — Searches for the decay of the Higgs boson in a pair of charm quarks ($H \to c\bar{c}$) were carried out by the CMS and ATLAS experiments with data collected during the Run-2 of the LHC in proton-proton collisions. The most stringent upper limit on the signal strength was set by the CMS Collaboration in the VH production mode, with the Higgs boson produced in association with a leptonically decaying vector boson (W or Z). The observed upper limit on $\sigma(VH)B(H \to c\bar{c})$ is 14 times the standard model prediction. During Run-3, tighter constraints could be established by extending the search to other production mechanisms of the Higgs boson, like the vector boson fusion. The search for $H \to c\bar{c}$ performed by CMS with Run-2 data is described in this work.

1. – Introduction

Since its discovery in 2012 [1-3], characterizing the Higgs boson properties has become one of the highest priority goals of general-purpose high energy physics (HEP) collider experiments, like CMS [4] and ATLAS. So far, the Higgs boson couplings with gauge bosons and third generation fermions and muons have been established, while its couplings with second generation quarks are still out of reach [5]. This work focuses on the search for the Higgs boson decay in a charm quark-antiquark pair performed by the CMS Collaboration with data collected in proton-proton collisions in 2016-2018 (Run-2) at a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}. It is a particularly challenging search because of the small branching ratio of the $H \to c\bar{c}$ decay ($B_{SM} \sim 3\%$) predicted by the standard model (SM), the charm jet identification (c-tagging) complexity and the overwhelming QCD multijet background. By investigating the VH production mechanism, with the Higgs boson generated in association with a W or Z boson, the CMS Collaboration set the tightest constraints on the Higgs-charm coupling modifier k_c to date [6]. In addition, the gluon fusion production channel (ggF) was explored in a high transverse momentum (p_T) phase space. An overview of these studies is provided in the following sections.

(*) IFAE 2023 - "Poster" session

2. – Search for $H \to c\bar{c}$ in the VH production channel

The search for $H \to c\bar{c}$ in the VH production mode was performed by the CMS Collaboration with data collected during the Run-2 of the LHC. By targeting the leptonic decays of the vector boson (W or Z) produced in association with the Higgs boson, the QCD multijet background is highly suppressed. Three mutually exclusive analysis categories are defined according to the leptonic decay channels of the W/Z boson: $Z \to \nu\nu$ (0L), $W \to l\nu$ (1L) and $Z \to ll$ (2L), where l can be an electron or a muon. Two complementary approaches are adopted in this search, involving different topologies: the "resolved-jet" and the "merged-jet" topology, used respectively in the transverse momentum (p_T) region of the Higgs boson candidate (H_{cand}), reconstructed as a single large-radius jet, below and above 300 GeV.

2˙1. *Resolved-jet topology*. – The H_{cand} is reconstructed from two well-separated and individually resolved c-tagged jets, reconstructed with a distance parameter R=0.4 of the anti-kt algorithm. Heavy flavor tagging is performed, in this topology, by means of the DeepJet algorithm [7]. A boosted decision tree (BDT) algorithm is trained in each category to discriminate the signal against the dominant physics background processes. The signal strength modifier μ, defined as $\frac{(\sigma B)_{obs}}{(\sigma B)_{SM}}$, where σ is the signal production cross section and B is the branching fraction, is extracted from a maximum likelihood fit to data of the BDT output score. Dedicated control regions (CRs) are defined to determine, with fits to data, the normalizations of the V+jets and $t\bar{t}$ backgrounds. Figure 1 (left) shows the post-fit distribution of the BDT output score in the 2L(ee) category in 2017 data. Analogous figures for the other categories can be found at [6].

2˙2. *Merged-jet topology* . – The hadronization products of the two charm quarks are reconstructed as a single large-radius jet (R=1.5). A novel powerful jet identification algorithm, ParticleNet [8], is used to identify large-radius jets arising from the Higgs boson decay into charm quarks. In order to further discriminate signal from the main backgrounds (V+jets and $t\bar{t}$) a separate BDT algorithm is trained for each category, taking as input kinematic variables not correlated with the H_{cand} mass ($m(H_{cand})$). As for the resolved-jet topology, background normalizations are obtained by defining dedicated CRs. The signal component is extracted from a fit of the Higgs boson candidate mass.

The combined $m(H_{cand})$ distribution in all the three categories is shown in fig. 1 (right).

2˙3. *Combined results*. – In order to improve the sensitivity of the search, a simultaneous fit of the two analysis results is performed. Figure 2 shows the 95% confidence level (CL) upper limits on the signal strength $\mu_{VH(H\to c\bar{c})}$ evaluated from the independent analyses and from their combination. The observed (expected) 95% CL upper limit on $\mu_{VH(H\to c\bar{c})}$ is $14(7.6^{+3.4}_{-2.3})$. Constraints on the Higgs-charm Yukawa coupling modifier k_c are then estimated, by reparametrizing $\mu_{VH(H\to c\bar{c})}$ in terms of k_c, according to eq. (1):

$$\mu_{VH(H\to c\bar{c})} = \frac{k_c^2}{1 + B_{SM}(H \to c\bar{c})(1 - k_c^2)} \quad (1)$$

The observed 95% CL interval is $1.1 < |k_c| < 5.5$, and the corresponding expected constraint is $k_c < 3.4$. The analysis is validated by searching for the Z decay to a charm

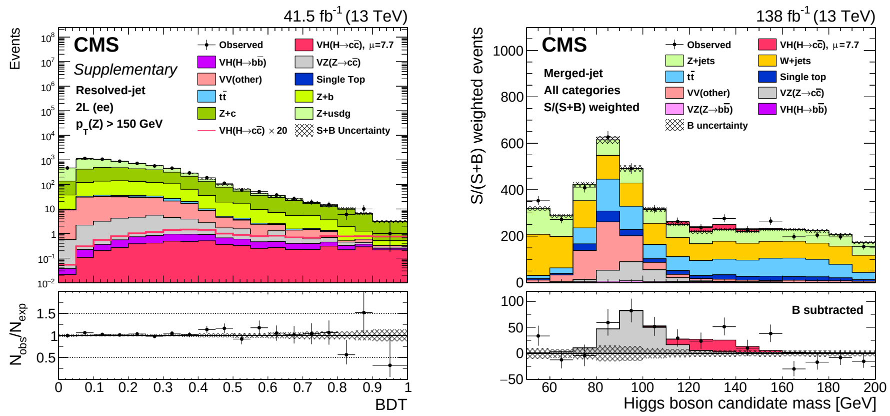

Fig. 1. – (Left) Post-fit distribution of the BDT discriminant in the high-pT(V) signal region of the resolved-jet topology in the 2L(ee) channel in 2017 data. The $VH(H \to c\bar{c})$ signal yield is scaled by the best fit signal strength, $VH(H \to c\bar{c}) = 7.7$, in the filled red histogram, and the red line represents the expected signal contribution multiplied by a factor of 20. (Right) Combined H_{cand} mass (m) distribution in all channels of the merged-jet analysis. The lower panel shows data (points) and the fitted VH($H \to c\bar{c}$) (red) and VZ($Z \to c\bar{c}$) (grey) distributions after subtracting all other processes [6].

quark-antiquark pair ($Z \to c\bar{c}$) in VZ production. For the first time at a hadron collider, this process was observed with a significance of 5.7 standard deviations.

3. – Search for $H \to c\bar{c}$ in the ggF production channel

A search for the Higgs boson produced with $p_T > 450$ GeV and decaying to a charm quark-antiquark pair was performed by the CMS Collaboration, targeting the ggF pro-

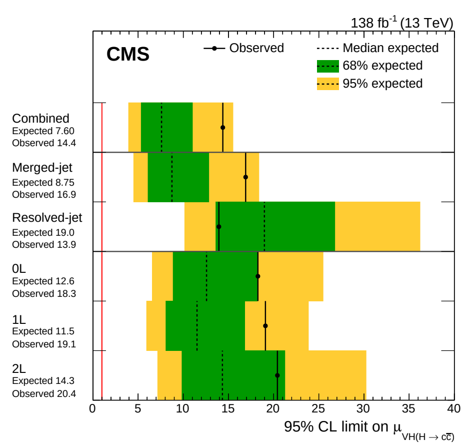

Fig. 2. – The 95% CL upper limits on $\mu_{VH(H\to c\bar{c})}$. Green and yellow bands indicate the 68 and 95% intervals on the expected limits, respectively. The vertical red line indicates the SM value $\mu_{VH(H\to c\bar{c})} = 1$ [6].

duction mode [9]. The H_{cand} is reconstructed as a single large-radius jet (R=0.8), tagged with the DeepDoubleX algorithm [10]. The soft-drop (SD) algorithm [11] is applied to the jet mass (m_{SD}) to remove soft and wide-angle radiation. The dominant QCD multi-jet background is estimated from data by using CRs. The V+jets processes are estimated from simulation, while the $t\bar{t}$ background is given by simulation and normalized with data in a dedicated CR. The signal strength μ_H is extracted from a binned (m_{SD}, p_T) maximum likelihood fit to data. The observed (expected) 95% CL upper limit on $\mu_{H\to c\bar{c}}$ is 47 (39). The search is validated by measuring the $Z \to c\bar{c}$ process, observed in association with high-p_T jets for the first time in this analysis, with signal strength $\mu = 1.00^{+0.17}_{-0.14}$ (syst) ± 0.08 (theo) ± 0.06 (stat).

4. – Conclusions and future developments

In this work, the searches for the Higgs boson decay in a charm quark-antiquark pair performed by the CMS Collaboration are summarized. The VH and ggF Higgs production channels have been explored, the former providing the tightest constraint on the Higgs-charm Yukawa coupling modifier to date: $1.1 < k_c < 5.5$ (expected: $|k_c| < 3.4$). During Run-3 of the LHC, new trigger and analysis methods are expected to further improve the constraints on the coupling between the Higgs boson and charm quarks.

REFERENCES

[1] ATLAS Collaboration, *Phys. Lett. B*, **716** (2012) 1, arXiv:1207.7214.
[2] CMS Collaboration, *Phys. Lett. B*, **716** (2012) 30, arXiv:1207.7235.
[3] CMS Collaboration, *JHEP*, **06** (2013) 081.
[4] CMS Collaboration, *JINST*, **3** (2008) S08004.
[5] Workman R. L. *et al.*, *PTEP*, **2022** (2022) 083C01.
[6] CMS Collaboration, *Phys. Rev. Lett.*, **131** (2023) 061801, arXiv:2205.05550 (additional figures at `http://cern.ch/go/Hdd9`).
[7] Bols E. *et al.*, *JINST*, **15** (2020) P12012.
[8] Qu H. and Gouskos L., *Phys. Rev. D*, **101** (2020) 056019.
[9] CMS Collaboration, *Phys. Rev. Lett.*, **131** (2023) 041801.
[10] CMS Collaboration, *Performance of the mass-decorrelated DeepDoubleX classifier for double-b and double-c large-radius jets with the CMS detector*, CERN-CMS-DP-2022-041.
[11] Larkoski A. J., Marzani S., Soyez G. and Thale J., *JHEP*, **05** (2014) 146, arXiv:1402.2657.