IL NUOVO CIMENTO **46 C** (2023) 21
DOI 10.1393/ncc/i2023-23021-2

Colloquia: La Thuile 2022

Search for CP violation in $t\bar{t}H$ and tH production in multilepton channels at $\sqrt{s} = 13$ TeV

C. Ramon Alvarez on behalf of the CMS Collaboration
University of Oviedo - Oviedo, Spain

received 19 September 2022

Summary. — We measure the CP structure of the Yukawa interaction between the Higgs boson (H) and the top quark in a data sample enriched in $t\bar{t}H$ and tH associated production. Data corresponding to an integrated luminosity of $138\,\text{fb}^{-1}$ collected in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ by the CMS experiment at the CERN LHC is used. Events where the H decays via $H \to WW$ or $H \to \tau\tau$ and top quarks decay either leptonically or hadronically are used. We apply machine learning techniques to enhance the separation of CP-even from CP-odd scenarios. Two-dimensional confidence regions are set on κ_t and $\tilde{\kappa_t}$ which are respectively the CP-even and CP-odd top-Higgs Yukawa coupling modifiers. We do not observe a significant fractional CP-odd component $|f_{CP}^{Htt}|$, which is determined to be 0.59 within an interval of (0.24, 0.81) at 68% confidence level. The results are combined with previously published analyses covering the $H \to ZZ$ and $H \to \gamma\gamma$ decay modes, yielding one- and two-dimensional confidence regions for κ_t and $\tilde{\kappa_t}$, while $|f_{CP}^{Htt}|$ is determined to be $|f_{CP}^{Htt}| = 0.28$ with an interval of $|f_{CP}^{Htt}| < 0.55$ at 68% confidence level.

1. – Introduction

The observation of a scalar boson by the ATLAS and CMS Collaborations in 2012 [1-3] opened a new experimental program in the field of particle physics: the properties of the observed boson need to be measured in detail to check whether they are consistent with the ones of the Higgs boson (H) predicted by the standard model (SM).

One such properties is how this boson couples to other particles. The SM Yukawa couplings of the H to fermions (y_f) is proportional to the fermion mass m_f, namely $y_\text{f} = \sqrt{2}m_\text{f}/v$, where $v \approx 246\,\text{GeV}$ denotes the vacuum expectation value of the Higgs field. The top quark is the heaviest known fermion; its Yukawa coupling y_t is expected to be of order one, and deviations of y_t from the SM prediction would indicate the presence of new physics beyond the SM, hence the determination of y_t provides a good handle to study the H.

In particular, the SM H is even under charge-parity (CP) inversion. Determining the CP structure of the couplings of the observed boson is therefore very important, as the presence of a CP-odd term in the H Lagrangian would be a direct indication of the presence of new physics.

The production of the H in association with a top quark pair ($t\bar{t}H$) or one top quark (tH) provides direct access to the magnitude and sign of y_t. The measurement presented [4] is based on data recorded by the CMS experiment [5] in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ during the LHC Run 2, corresponding to an integrated luminosity of $138\,\text{fb}^{-1}$. This analysis includes the signatures $2\ell ss + 0\tau_h$, $2\ell ss + 1\tau_h$, and $3\ell + 0\tau_h$, which account for the H decay modes $H \to WW$ and $H \to \tau\tau$, targeting events in which at least one top quark decays leptonically and providing the highest sensitivity to possible CP violation effects. The symbol ℓ denotes light leptons (e, μ), and "ss" means *same-sign* charge. The symbol τ_h denotes hadronically decaying tau leptons. As in previous analyses [8], the separation of the $t\bar{t}H$ and tH signals from backgrounds is improved with machine learning techniques. Boosted decision trees (BDTs) and artificial deep neural networks (DNNs) are used. In particular, machine learning methods are employed to improve the separation between CP-odd and CP-even scenarios, both pure and mixed, for the $t\bar{t}H$ and tH signals.

The Lagrangian for the fermion-Higgs interaction can be written as a superposition of a CP-even and a CP-odd phase:

$$\mathcal{L}_{t\bar{t}H} = \frac{m_t}{v}\bar{\psi}_t(\kappa_t + i\gamma_5\widetilde{\kappa}_t)\psi_t H. \tag{1}$$

Here $\bar{\psi}_t$ and ψ_t are Dirac spinors, m_t is the top mass, v is the SM Higgs field vacuum expectation value, while κ_t and $\widetilde{\kappa}_t$ are respectively the CP-even and CP-odd top-Higgs Yukawa couplings. Different CP scenarios would yield a change in the $t\bar{t}H$ and tH cross sections as well as modifications in kinematic distributions.

2. – Event selection

Events are classified depending on the number of leptons in the final state in three orthogonal categories: $2\ell ss + 0\tau_h$, $2\ell ss + 1\tau_h$, and $3\ell + 0\tau_h$. In each category a dedicated selection based on jet and b-tag multiplicity is imposed: at least 2 (3) jets are required in $2\ell ss + 0\tau_h$, $2\ell ss + 1\tau_h$ ($3\ell + 0\tau_h$) and at least one medium b-tagged jet ([1]) or two loose b-tagged jets should be present on the event. In order to target tH events, if the event does not pass the previous selection but contains at least 1 light jet (that can be in the forward region) and at least 1 b-tagged medium jet, the event is also selected.

The main remaining backgrounds are nonprompt leptons, misidentified taus and electron flips which are estimated using data driven techniques. As well as $t\bar{t}W$, $t\bar{t}Z$ and (less importantly) diboson production; these are modelled from simulation and dedicated control regions are used in order to constrain its normalization.

Furthermore, machine learning techniques (DNNs) are used to discriminate $t\bar{t}H$ and tH signal from background. Events are classified depending on its most-probable output nodes. In the $2\ell ss + 1\tau_h$ and $3\ell + 0\tau_h$ three output nodes are defined, targeting: $t\bar{t}H$, tH and other backgrounds, respectively. In the $2\ell ss + 0\tau_h$ four nodes are defined, each of them dedicated to: $t\bar{t}H$, tH, $t\bar{t}W$ and other backgrounds.

The strategy summarized in this section follows the developments done in ref. [8].

([1]) In order to discriminate jets produced by heavy-flavour quarks from those coming from light-flavour quarks and gluons, we make use of the DeepJet discriminato [6, 7].

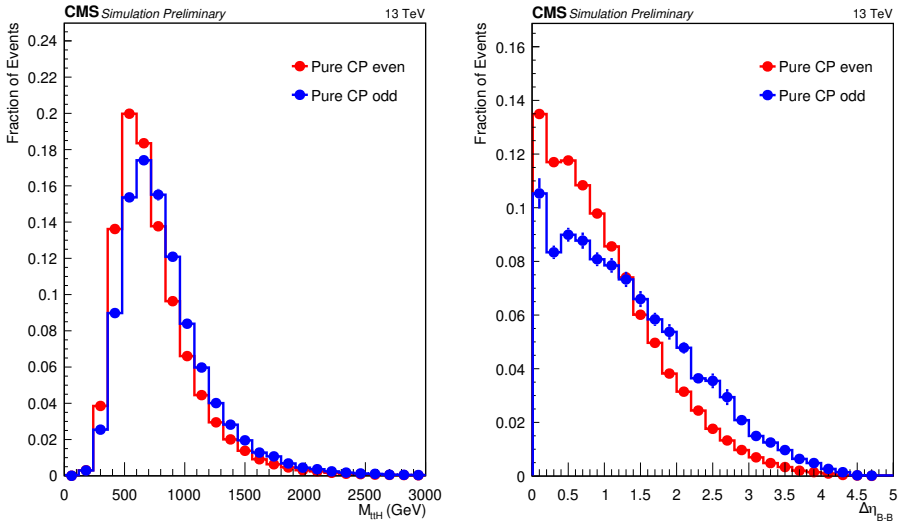

Fig. 1. – Two of the most important input variables to the BDT used for CP discrimination. The left plot shows the $M_{\mathrm{t\bar{t}H}}$ in the $2\ell ss + 1\tau_{\mathrm{h}}$ channel while the right one shows the $\Delta\eta_{B-B}$ distribution in the $3\ell + 0\tau_{\mathrm{h}}$ channel. The vertical bars represent the statistical uncertainty originating from the limited amount of simulated events. When not visible, the bars are smaller than the marker size [4].

3. – CP discrimination

In order to discriminate CP-even and CP-odd scenarios, three different BDT classifiers are trained, one for each of the channels: $2\ell ss + 0\tau_{\mathrm{h}}$, $2\ell ss + 1\tau_{\mathrm{h}}$, and $3\ell + 0\tau_{\mathrm{h}}$.

Each of the BDTs uses a different set of input features. The different kinematic characteristics of the t$\bar{\mathrm{t}}$H process depending on the CP scenario are exploited in the training. The three-momenta, pseudorapidity and ϕ values of the final state particles are used as input variables in all BDTs.

The invariant mass of the system composed by the leptons in the final state (including τ's), the missing transverse energy, and an appropriate number of jets (depending on the final state), show good discriminating power between CP-even and CP-odd scenarios as depicted in fig. 1 (left). This variable is defined as follows,

$$M_{\mathrm{t\bar{t}H}} = \sqrt{\left| \sum_i p^{\mathrm{lep}_i} + \vec{p_T}^{\,\mathrm{miss}} + \sum_{i \leq k} p^{\mathrm{jet}_i} \right|^2}, \tag{2}$$

where k is the number of jets used to compute the invariant mass of the system, which is taken according to the expected number of jets given by the t$\bar{\mathrm{t}}$H subsequent decays in the final state corresponding to the category, i.e. k = 6 (4) in the final state $2\ell ss + 0\tau_{\mathrm{h}}$ ($2\ell ss + 1\tau_{\mathrm{h}}$ and $3\ell + 0\tau_{\mathrm{h}}$).

As can be seen in fig. 1 (right) the $\Delta\eta$ ($\Delta\eta_{B-B}$) of two jets with highest score in the b-tagging discriminator (DeepJet) also shows good discrimination between the two scenarios.

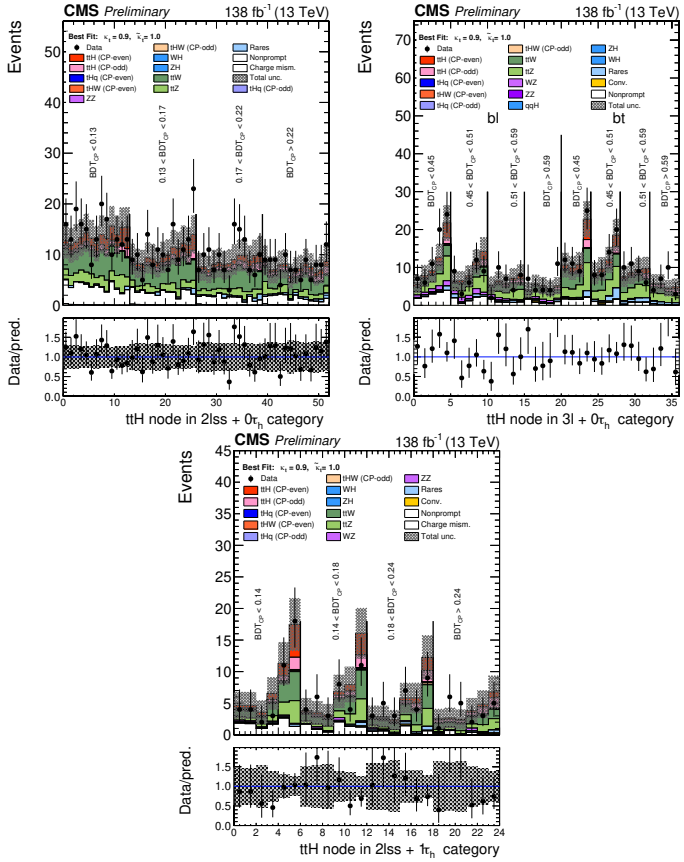

Fig. 2. – Postfit plots of the discriminating variables in the $t\bar{t}H$ SR used as input to the fit in each of the categories: $2\ell ss + 0\tau_h$ (upper left) $3\ell + 0\tau_h$ (upper right) and $2\ell ss + 1\tau_h$ (lower row) [4].

4. – Signal extraction

Events are classified in regions enriched in $t\bar{t}H$ and tH signals, and backgrounds, as described in sect. **2**. In the $2\ell ss + 0\tau_h$ category, events are further classified depending on the flavour of the leptons ($ee/e\mu/\mu\mu$), except in the $t\bar{t}H$ node where no flavour categorization is applied. In the $3\ell + 0\tau_h$ category, the tH and $t\bar{t}H$ nodes are separated according to the number of b-tagged jets. Events in the $t\bar{t}H$ node are also classified according to the CP discriminants described in the previous section. Four CP-bins are used as shown in fig. 2, where the three discriminating distributions in the $t\bar{t}H$ signal node are shown.

We perform a maximum likelihood fit including the distributions of selected observables in the signal regions defined for the $2\ell ss + 0\tau_h$, $2\ell ss + 1\tau_h$, and $3\ell + 0\tau_h$ categories as well as the control regions.

5. – Results

The ttH and tH yields are parametrized as a function of the coupling modifiers of the Higgs to the top. We keep other coupling modifiers and branching ratios fixed to its

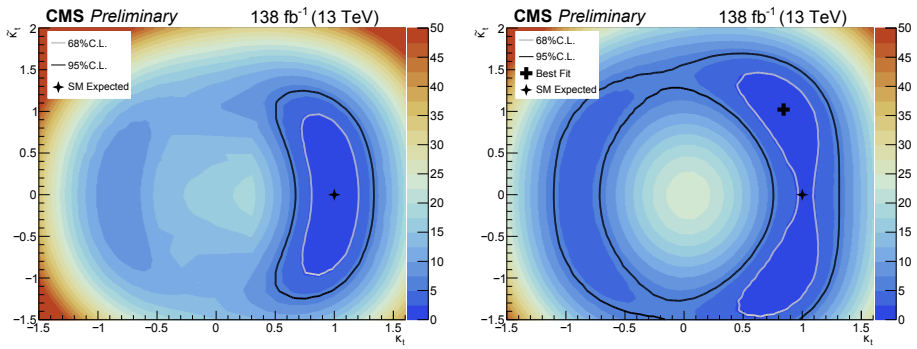

Fig. 3. – Likelihood scan as a function of κ_t and $\widetilde{\kappa}_t$: expected (left) and observed (right). The black diamond shows the best value for κ_t and $\widetilde{\kappa}_t$ given by the fit. The black cross shows the expected SM values for κ_t and $\widetilde{\kappa}_t$. Both 68 and 95% CL interval contours are shown [4].

standard model value. Confidence regions at 68% and 95% CL for the κ_t and $\widetilde{\kappa}_t$ couplings are obtained applying the asymptotic properties of the Likelihood ratio. Figure 3 (left) shows the expected likelihood scan as a function of κ_t and $\widetilde{\kappa}_t$. The observed likelihood scan as a function of κ_t and $\widetilde{\kappa}_t$ is shown in fig. 3 (right): the best fit is compatible with the SM within 68% CL.

An alternative interpretation is also provided using the $t\bar{t}H$ signal strength multiplier to the SM cross section, $\mu_{t\bar{t}H}$, and a parameter $|f_{CP}^{Htt}| = \frac{\widetilde{\kappa}_t^{\,2}}{\widetilde{\kappa}_t^{\,2}+\kappa_t^2}$, where the H couplings to other particles are constrained to their SM prediction. The likelihood as a function of $|f_{CP}^{Htt}|$, while profiling $\mu_{t\bar{t}H}$, is shown in fig. 4 for the expected and observed results.

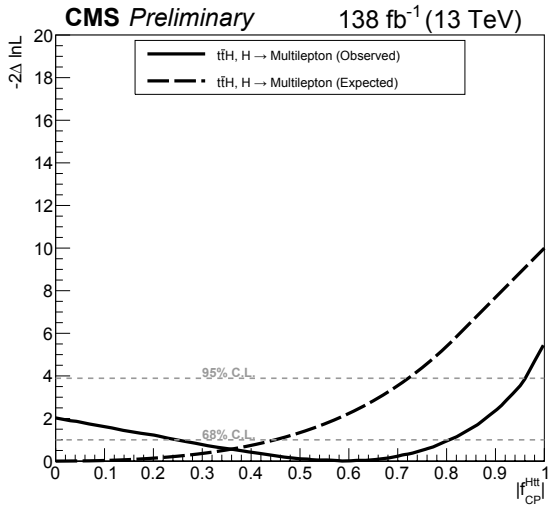

Fig. 4. – Likelihood scan as a function of $|f_{CP}^{\mathrm{Htt}}|$ for multilepton final states. The solid (dashed) line shows the observed (expected) scan [4].

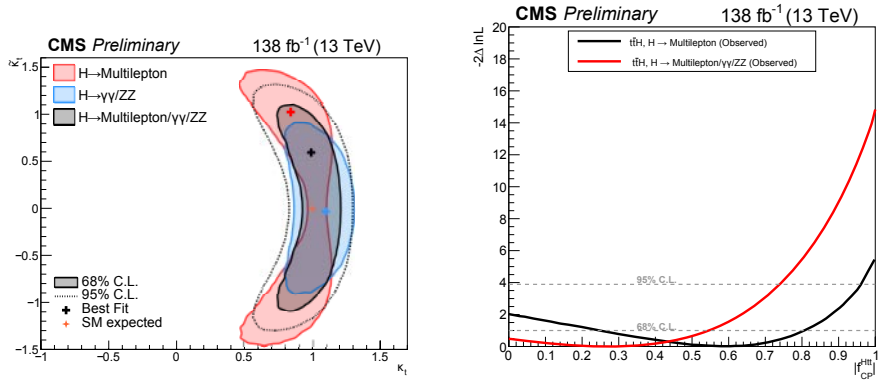

Fig. 5. – Left: Likelihood scan as a function of κ_t and $\widetilde{\kappa_t}$. Two-dimensional confidence intervals at 68% CL are depicted as shaded areas, for multilepton (red), the combination of H→ $\gamma\gamma$ and H→ZZ (blue), and the combination of the three channels (black). The 95% CL for the combination is shown as a dashed line. The best fit for each is shown as a cross of the corresponding color. Right: Observed likelihood scan as a function of $|f_{CP}^{Htt}|$ for multilepton final states and the combination of multilepton, H→ $\gamma\gamma$, and H→ZZ final states [4].

This parametrization results in a best fit value for the fractional contribution of $|f_{CP}^{Htt}| = 0.59$ and an interval of (0.24, 0.81) at 68% CL using multilepton final states only. This is in agreement with the SM within 95% CL. Besides, the pure CP-odd scenario is excluded at 68% CL.

6. – Combination with other channels

We combine the results described in the previous section for the decay channels H →WW and H → $\tau\tau$ (referred to as the *multilepton decay channel*) with previously published CMS results on the CP parameters in other $t\bar{t}H$ decay channels, namely the H→ $\gamma\gamma$ [9] and H→ ZZ final states [10].

The combination improves the sensitivity significantly. We show the results in fig. 5.

All results are in good agreement with the SM. We get a best fit value of $|f_{CP}^{Htt}| = 0.28$ and an interval of $|f_{CP}^{Htt}| < 0.55$ at 68% CL. In addition, the scenario with $|f_{CP}^{Htt}| = 1$ is excluded with 3.7 standard deviations.

7. – Summary

A measurement is presented of the CP structure of the Yukawa coupling between the Higgs boson and the top quark at tree level, when H is produced in association with one (tH) or two ($t\bar{t}H$) top quarks. This measurement is based on data collected in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ by the CMS experiment at the CERN LHC, corresponding to an integrated luminosity of $138\,\text{fb}^{-1}$. The analysis targets events with multiple leptons in the final state. Separation of CP-even from CP-odd scenarios is achieved by applying machine learning techniques. Two-dimensional confidence regions are derived for κ_t and $\widetilde{\kappa_t}$ which are respectively the CP-even and CP-odd top-Higgs Yukawa coupling modifiers. The fractional contribution observed is $|f_{CP}^{Htt}| = 0.59$ with an interval of $(0.24, 0.81)$ at 68% CL. The results are combined with previously published

analyses covering the H→ZZ and H→ $\gamma\gamma$ decay modes. One- and two-dimensional confidence regions are set on κ_t and $\widetilde{\kappa}_t$, constraining κ_t to be within $(0.86, 1.26)$ and $\widetilde{\kappa}_t$ to be within $(-1.07, 1.07)$ at 95% CL. The fractional contribution is also investigated in the combination, yielding a best fit of $|f_{CP}^{Htt}| = 0.28$ and an interval of $|f_{CP}^{Htt}| < 0.55$ at 68% CL. The results are compatible with predictions for the standard model H and the pure CP-odd scenario is excluded with 3.7 standard deviations.

REFERENCES

[1] ATLAS Collaboration, *Phys. Lett. B*, **716** (2012) 1.
[2] CMS Collaboration, *Phys. Lett. B*, **716** (2012) 30.
[3] CMS Collaboration, *JHEP*, **06** (2013) 081.
[4] CMS Collaboration, CMS-PAS-HIG-21-006, `http://cds.cern.ch/record/2803420?ln=en`.
[5] CMS Collaboration, *JINST*, **3** (2008) S08004.
[6] Bols E. *et al.*, *JINST*, **15** (2020) P12012.
[7] CMS Collaboration, *Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector*, CMS Detector Performance Note, CMS-DP-2018-058 (2018).
[8] CMS Collaboration, *Eur. Phys. J. C*, **378** (2021) 81.
[9] CMS Collaboration, *Phys. Rev. Lett.*, **125** (2020) 061801.
[10] CMS Collaboration, *Phys. Rev. D*, **104** (2021) 052004.