Machine learning the Higgs boson-top quark *CP* phase

Rahool Kumar Barman,[1,*] Dorival Gonçalves,[1,†] and Felix Kling[2,3,‡]

[1]*Department of Physics, Oklahoma State University, Stillwater, Oklahoma 74078, USA*
[2]*SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA*
[3]*Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, Hamburg 22607, Germany*

(Received 2 December 2021; accepted 31 January 2022; published 22 February 2022)

We explore the direct Higgs boson-top *CP* measurement via the $pp \to t\bar{t}h$ channel at the high-luminosity LHC. We show that a combination of machine learning techniques and efficient kinematic reconstruction methods can boost new physics sensitivity, effectively probing the complex $t\bar{t}h$ multiparticle phase space. Special attention is devoted to top quark polarization observables, uplifting the analysis from a raw rate to a polarization study. Through a combination of hadronic, semileptonic, and dileptonic top pair final states in association with $h \to \gamma\gamma$, we obtain that the HL-LHC can probe the Higgs boson-top coupling modifier and *CP* phase, respectively, up to $|\kappa_t| \lesssim 8\%$ and $|\alpha| \lesssim 13°$ at 68% C.L.

DOI: 10.1103/PhysRevD.105.035023

I. INTRODUCTION

New sources of *CP* violation can be a key ingredient to explain the matter-antimatter asymmetry of the Universe [1–3]. Hence, the quest for new *CP* violating interactions is a clear target for beyond the Standard Model (SM) searches, being a critical component of the physics program of the LHC. A particularly interesting option is that the Higgs boson couplings present these new physics sources.

From the theoretical point of view, some Higgs interactions are more inclined to display *CP* violation effects than others. While the widely studied beyond the SM (BSM) *CP* structure for the Higgs to vector boson couplings are loop suppressed, arising only at dimension six or higher [4,5], *CP* violation in Higgs to fermion interactions can manifest already at the tree level [6], being naturally larger. Owning to its magnitude, the top quark Yukawa coupling can play a significant role in this context and be most sensitive to new physics.

Whereas it is possible to access the Higgs boson-top coupling through loop induced processes [7–14], the direct Higgs boson-top production via $pp \to t\bar{t}h$ is crucial to disentangle possible new physics effects [6,15–32]. This channel was observed in 2018 by both ATLAS and CMS with significances of 6.3 and 5.2σ, respectively [25,26]. The high-luminosity LHC (HL-LHC) studies indicate that the Higgs boson-top interaction will be probed to outstanding accuracy at the end of the LHC run, reaching $\delta y_t \lesssim 4\%$ when combining the HL-LHC ATLAS and CMS data [33]. The same projections indicate that the $t\bar{t}h$ channel in the $h \to \gamma\gamma$ final state will display dominant sensitivities. While the diphoton final state presents limited statistics, it highly benefits from controlled backgrounds from sidebands. Recently, ATLAS and CMS have reported the first experimental Higgs boson-top *CP* studies, exploring the $t\bar{t}h$ channel [34,35]. Both analyses focus on the diphoton final state, $h \to \gamma\gamma$. ATLAS and CMS exclude Higgs boson-top *CP*-mixing angles above 43° and 55° at 95% C.L., respectively.

In the present paper, we perform a detailed investigation of the Higgs boson-top *CP* sensitivity with the $pp \to t\bar{t}h$ channel at the HL-LHC, considering the most promising decay mode, $h \to \gamma\gamma$. We explore the complex multiparticle final state with a combination of machine learning techniques and efficient kinematic reconstruction methods. Since distinct Higgs boson-top *CP* phases affect the net top and antitop quark polarization, propagating the spin effects to the top quark final states, we devote special attention to include the top polarization observables in our study. In particular, these spin effects are used to define genuine *CP* observables. After motivating and constructing the relevant kinematic observables, we evaluate how much information can be extracted with them. The convenient metric adopted to quantify this is given by the Fisher information. We show that the ability of probing the $pp \to t\bar{t}h$ channel exploring the complex multiparticle final state not only in terms of a raw rate, but as a polarized process, can offer a crucial pathway to probe the underlying production dynamics, accessing possible new physics effects.

*rahool.barman@okstate.edu
†dorival@okstate.edu
‡felixk@slac.stanford.edu

The structure of this paper is as follows. In Sec. II, we present the theoretical parametrization for the top Yukawa coupling. We discuss the new physics effects to the top polarization, define the CP-sensitive observables, and quantify how much information on the CP phase can be extracted using distinct observables. In Sec. III, we present the kinematic reconstruction methods, which are relevant to build prominent observables to new physics and maximally explore the $t\bar{t}h$ final state. Next, in Sec. IV, we move on to the detailed analysis, where we derive the projected sensitivities for the Higgs boson-top CP phase at the HL-LHC. This study is inclusive in respect to the top pair final states, combining the leptonic, semileptonic, and hadronic channels. Finally, a summary of our key findings is delivered in Sec. V.

II. CP STRUCTURE AND OBSERVABLES

We parametrize the top quark Yukawa coupling with the following Lagrangian

$$\mathcal{L} = -\frac{m_t}{v}\kappa_t \bar{t}(\cos\alpha + i\gamma_5 \sin\alpha)th, \tag{1}$$

where m_t is the mass of the top quark, v is the vacuum expectation value in the SM ($v = 246$ GeV), κ_t is a real number, and α is the CP phase. The interaction between the CP-even Higgs boson and the top quark in the SM is represented by $(\kappa_t, \alpha) = (1, 0)$, while $\alpha = \pi/2$ results in a pure CP-odd Higgs boson-top interaction. New physics contributions in Eq. (1) will display effects both in the Higgs $t\bar{t}h$ production and decay, $h \to \gamma\gamma$. Whereas the Higgs decay will more relevantly change the total signal rate [15], we will devote special attention to probe the new physics effects in the Higgs production, exploring the top quarks' final state kinematics. This will be an essential ingredient to uplift the new physics sensitivity from CP-phase effects.

A. Top polarization

Among the observables sensitive to the structure of the top quark Yukawa interaction, the spin correlations between the top and antitop in $t\bar{t}h$ associated production offer a prominent pathway for precision studies [6,15–20,24,32,36–41]. Owing to its short lifetime ($\sim 10^{-25}$ s) [42], the top quark is expected to decay before hadronization occurs ($\sim 10^{-24}$ s) and spin decorrelation effects take place ($\sim 10^{-21}$ s) [43]. Thus, the spin-spin correlations between t and $\bar{t}$ can be traced back from the top quark decay products. In particular, it is possible to observe correlations between any two decay products, one from the top quark and the other from the antitop quark. The correlations scale with the spin analyzing power associated with each top decay product.

More accurately, the top quark final states in the leptonic $t \to W^+ b \to \ell^+ \nu b$ and hadronic $t \to W^+ b \to \bar{d} u b$ channels are correlated with the top quark spin axis as follows:

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\xi_i} = \frac{1}{2}(1 + \beta_i P_t \cos\xi_i), \tag{2}$$

where Γ is the partial decay width, ξ_i is the angle between the ith decay product and the top quark spin axis in the top quark rest frame, P_t is the polarization of the decaying top ($-1 \le P_t \le 1$), and β_i is the spin analyzing power of the final state particle i [44]. At leading order, the coefficient β_i is $+1$ for charged lepton ℓ^+ and $\bar{d}$ quark, -0.3 for $\bar{\nu}$ and u quark, -0.4 for the b quark, and 0.4 for W boson. The sign of the coefficient β_i is flipped for antitop decays.

Granted by the $V - A$ current structure of the weak interaction, the charged lepton will be a prominent spin analyzer, favoring studies with dileptonic top pairs. Exploring this phenomenology, the $\Delta\phi^{\rm lab}_{\ell\ell}$ observable, which is the azimuthal angle difference between the two charged leptons in the lab frame, is a good example of probe that has been found effective in accessing the Higgs boson-top CP properties [6,24]. Remarkably, the sensitivity of $\Delta\phi^{\rm lab}_{\ell\ell}$ improves further in the boosted Higgs regime due to the change in the net polarization for the top-pair at high energies.

Analogously to the charged lepton, the d quark also presents maximal spin analyzing power. However, it is a challenging task to tag a d-quark jet in a collider environment. An efficient solution is to select the softest of the two light-quark jets, $j_{\rm soft}$, in the top quark rest frame. This choice uplifts the spin analyzing power of $j_{\rm soft}$ to 50% of the lepton's [45]. This approach boosts the spin correlation analyses for the semileptonic and hadronic top quark pairs. Several observables can be defined exploiting this fact, a particularly relevant example, which we will explore in this paper, is the azimuthal angle difference between the charged lepton and softest light jet in the top pair rest frame, $\Delta\phi^{t\bar{t}}_{\ell j_{\rm soft}}$.

B. CP-sensitive observables

Various kinematic observables have been studied in the literature to access the Higgs boson-top CP structure in the $pp \to t\bar{t}h$ channel. Some illustrative distributions are presented in Fig. 1, such as the transverse momentum for the Higgs boson p_{Th} (top left) [46,47], the invariant mass for the top pair m_{tt} (top center), the product of projections of top and antitop momentum $b_4 = p_t^z p_{\bar{t}}^z / p_t p_{\bar{t}}$ (top right) [37], and the angle between the top quark and the beam direction in the $t\bar{t}$ center of mass (c.m.) frame θ^* which is also known as Collins-Soper angle (bottom left) [24]. These observables result in distinct profiles for different Higgs boson-top CP phases. The pure CP-odd phase, $\alpha = \pm\pi/2$, leads to a shift to higher energies in the peak of the distributions compared to the SM scenario, $\alpha = 0$. Different CP phases interpolate between these two profiles without sensitivity for the sign of the phase.

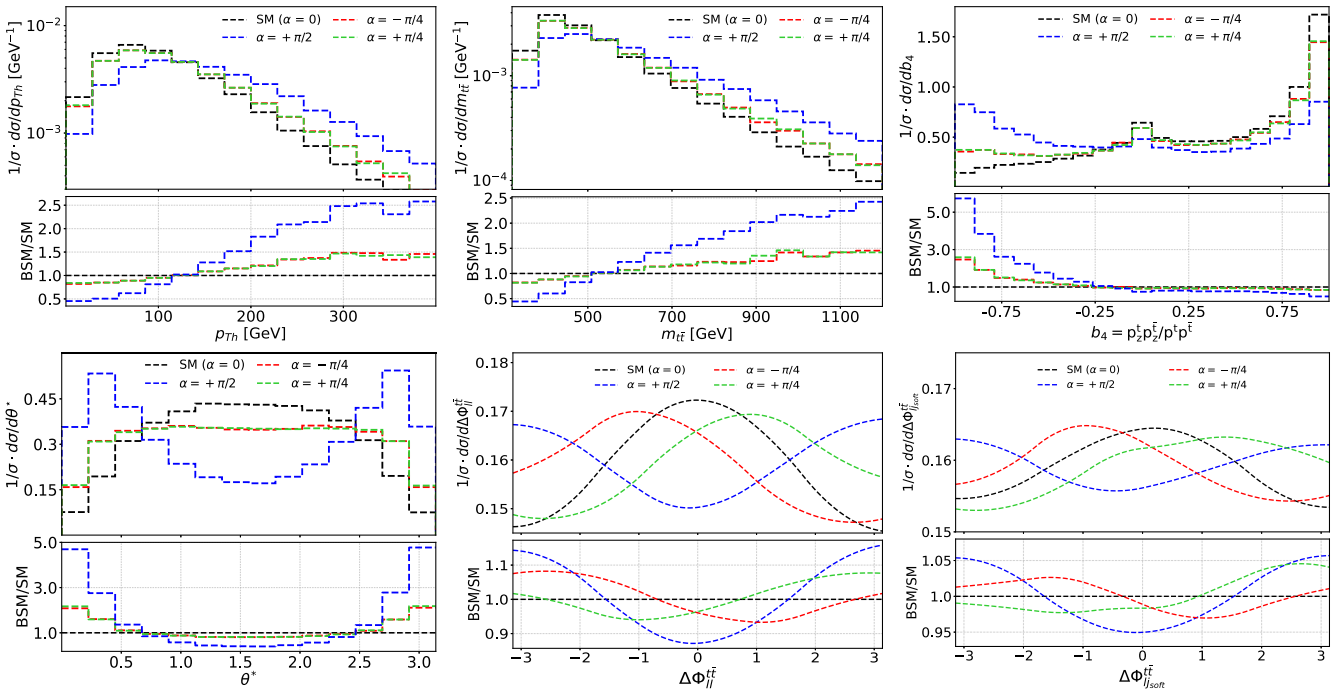

FIG. 1. Top panels: distributions for the transverse momentum for the Higgs boson p_{Th} (left), invariant mass for the top pair $m_{t\bar{t}}$ (center), and the product of the projections of the top and antitop momentum $b_4 = p_t^z p_{\bar{t}}^z / p_t p_{\bar{t}}$ (right). Bottom panels: distributions for the Collins-Soper angle θ^* (left), the azimuthal angle between the two charged leptons in the top pair rest frame $\Delta\phi_{\ell\ell}^{t\bar{t}}$ for fully leptonic $t\bar{t}h$ events (center), and the same angle between the charged lepton and the softest light jet in the top rest frame $\Delta\phi_{\ell j_{\rm soft}}^{t\bar{t}}$ for semileptonic $t\bar{t}h$ events (right). Each panel shows parton level results for the $t\bar{t}h$ sample for the SM Higgs ($\alpha = 0$), a *CP*-odd Higgs ($\alpha = \pi/2$) and mixed hypotheses ($\alpha = \pm\pi/4$). We also present the ratio between new physics and SM scenarios on the bottom panel of each figure. The results are presented for the 14 TeV LHC.

The variables p_{Th}, $m_{t\bar{t}}$, b_4, and θ^* are *CP*-even observables, being sensitive to the squared terms: $\cos^2\alpha$ and $\sin^2\alpha$. Thus, these probes are indifferent to the *CP*-even and *CP*-odd Higgs boson-top interference terms, which are proportional to $\cos\alpha \sin\alpha$. In particular, they are not sensitive to variations from a relative sign difference in the *CP* phase. Genuine *CP*-sensitive observables can be constructed from antisymmetric tensor products that require four linearly independent four-momenta. Owing to the top polarization being carried out to the decays, it is possible to construct such observable using, for instance, the top, antitop, and their decay products [16,20,24]. In general, the antisymmetric tensor product can be expressed as

$$\epsilon(p_t, p_{\bar{t}}, p_i, p_k) \equiv \epsilon_{\mu\nu\rho\sigma} p_t^\mu p_{\bar{t}}^\nu p_i^\rho p_k^\sigma, \tag{3}$$

where $\epsilon_{0123} = 1$, and $\{i, k\}$ represent the final state particles produced from the top and the antitop decays, respectively.

In the $t\bar{t}$ c.m. frame, Eq. (3) can be fortuitously simplified to $p_t \cdot (p_i \times p_k)$. This mathematical relation can be used to define azimuthal angle differences between the decay products in the $t\bar{t}$ c.m. frame that are odd under *CP* transformations [24]:

$$\Delta\phi_{ik}^{t\bar{t}} = \text{sgn}[\vec{p}_t \cdot (\vec{p}_i \times \vec{p}_k)] \arccos\left[\frac{\vec{p}_t \times \vec{p}_i}{|\vec{p}_t \times \vec{p}_i|} \cdot \frac{\vec{p}_t \times \vec{p}_k}{|\vec{p}_t \times \vec{p}_k|}\right]. \tag{4}$$

For illustration, we present in Fig. 1 the azimuthal angle between the two charged leptons $\Delta\phi_{\ell\ell}^{t\bar{t}}$ in the fully leptonic case (bottom center) and between the charged lepton and the softest light jet in the top rest frame $\Delta\phi_{\ell j_{\rm soft}}^{t\bar{t}}$ in the semileptonic case (bottom right). Two comments are in order. First, we notice that $\Delta\phi_{ik}^{t\bar{t}}$ is indeed sensitive to the sign of the *CP* phase, as illustrated in a comparison between the distribution profiles for $\alpha = \pi/4$ against $-\pi/4$. Second, in light of the spin analyzing power of the charged lepton in relation to $j_{\rm soft}$, the relative *CP* sensitivity of the dileptonic against the semileptonic correlation follows our expectation. Namely, the BSM effects in the $\Delta\phi_{\ell j_{\rm soft}}^{t\bar{t}}$ observable are ~50% weaker in respect to $\Delta\phi_{\ell\ell}^{t\bar{t}}$. This can be observed by comparing the bottom panel of these plots, where we display the BSM/SM ratio.

C. Observable information

Before proceeding to a full analysis, let us pause for a moment to better understand which distributions and

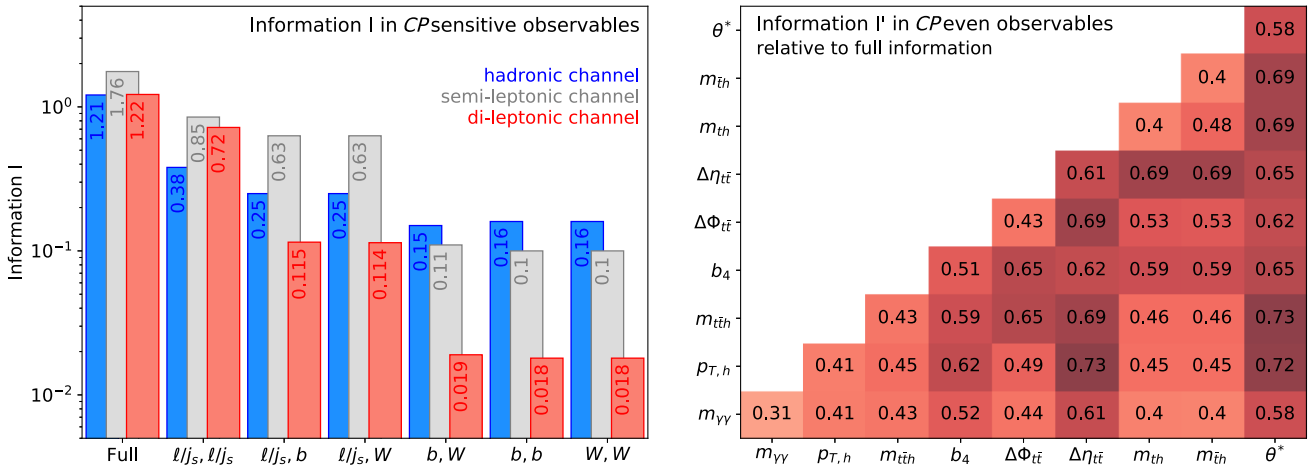

FIG. 2. Comparison of sensitivity on the CP phase α arising from different observables in terms of the Fisher information I for CP-odd observables probing linear new physics effects (left) and in terms of the modified Fisher information I' for CP-even observables probing the nonlinear new physics effects (right).

channels are sensitive to the CP phase α. In particular, we would like to quantify and compare how much information on the CP phase is available using the different observables in a parton level setup. This will provide some benchmarks and highlight the main ingredients required for an efficient analysis strategy that will be presented in Sec. IV.

Let us first consider the spin correlation observables $\Delta\phi_{ik}^{t\bar{t}}$ between two decay products from the top and antitop, which probe the new physics effects linear in α. A convenient metric to quantify the sensitivity of these observables to constrain the parameters of our model is given by the Fisher information [48,49]. Its component describing the sensitivity to the CP phase α is defined as

$$I = \mathbb{E}\left[\frac{\partial \log p(x|\kappa_t,\alpha)}{d\alpha}\frac{\partial \log p(x|\kappa_t,\alpha)}{d\alpha}\right]. \tag{5}$$

Here, $p(x|\kappa_t,\alpha)$ is the likelihood function, which describes the probability to observe a set of events with corresponding observables x as a function of the model parameter κ_t and α. $\mathbb{E}[\cdot]$ denotes the expectation value evaluated at the SM point, $(\kappa_t,\alpha)_{\rm SM} = (1,0)$. In the following we use the MadMiner package to calculate the Fisher information [50].

In the left panel of Fig. 2, we show the Fisher information associated with the CP-sensitive spin correlation observables for the dileptonic (red), semileptonic (gray), and hadronic (blue) channels. The bars on the left show the full information, i.e., the information that could be accessed via a comprehensive multivariate analysis. This was estimated using the machine learning method based on the SALLY algorithm [51–53] trained with all possible spin correlation observables. The remaining bars show the information in individual observables $\Delta\phi_{ik}^{t\bar{t}}$, which were estimated using a histogram based approach.

Focusing first on the dileptonic channel, the most sensitive among these observables is the spin correlation between the leptons, $\Delta\phi_{\ell\ell}^{t\bar{t}}$, since the spin analyzing power for the charged leptons are maximal. The next most sensitive observables are those where a charged lepton has been replaced with a b jet or a W boson. We observe that the corresponding Fisher information in $\Delta\phi_{\ell b}^{t\bar{t}}$ and $\Delta\phi_{\ell W}^{t\bar{t}}$ are suppressed compared to $\Delta\phi_{\ell\ell}^{t\bar{t}}$ by the square of the spin analyzing power $\beta_{b/W}^2 \sim 0.4^2$, as expected. The information in the spin correlations observables between a pair of b jet(s) and/or W boson(s) is further suppressed by an additional factor of $\beta_{b/W}^2$.

Let us now also consider the other top decay channels. As the Fisher information is proportional to the rate [48], we expect it to increase relative to the fully leptonic channel by a factor $2 \times \mathrm{BR}_{W\to\mathrm{had}}/\mathrm{BR}_{W\to\mathrm{lep}} \sim 6$ for the semi-leptonic channel and $(\mathrm{BR}_{W\to\mathrm{had}}/\mathrm{BR}_{W\to\mathrm{lep}})^2 \sim 9$ for the hadronic channel. Looking at the last three observables involving b jets and W bosons, this is indeed the case. For the other observables, we notice an additional loss of about a factor 2 in spin analyzing power, and hence a factor 4 in the Fisher information, which is caused by probing $j_{\rm soft}$ instead of the d quark.

Overall, we see that the different observables have distinct overall importance in the three channels. For dileptonic top decays, most of the information is contained in the spin correlation between the leptons, while the information in other observables is significantly suppressed. In contrast, for the hadronic decay channel, all shown observables have almost similar information. In this case, the resulting full information, which can be obtained by combining the different spin correlation observables, significantly exceeds the information of any individual observable. Overall, all three channels carry a similar amount of information on the CP phase α, which suggest performing a combined analysis.

Due to the limited $t\bar{t}h$ event rate at the LHC, we expect the nonlinear new physics effects to dominate over the

linear ones. We therefore expect most of the sensitivity on the *CP* structure of the top Yukawa coupling to arise from these nonlinear terms, despite the fact that the corresponding observables are not genuinely *CP* sensitive. To quantify the sensitivity of these *CP*-even observables to the squared terms, we use modified version of the Fisher information that was introduced in Ref. [54]. In this approach, we simply consider the square of the coupling as our new model parameter and define

$$I' = \mathbb{E}\left[\frac{\partial \log p(x|\kappa_t^2,\alpha^2)}{d\alpha^2}\frac{\partial \log p(x|\kappa_t^2,\alpha^2)}{d\alpha^2}\right]. \quad (6)$$

The result is shown in the right panel of Fig. 2. Here, we show the information associated with a two-dimensional distribution of two observables, relative to the full information associated with a multivariate analysis using all observables. As none of the presented observables relies on the top quark final state kinematics, the results are identical for all three top quark decay channels.

The distribution of the invariant mass of the photon pair, $m_{\gamma\gamma}$, is only sensitive to the theory parameters through its normalization. Correlating it with itself, we obtain the information associated with the signal strength measurements, which accounts for 31% of the information on the *CP* phase. In the absence of background, the correlation of $m_{\gamma\gamma}$ and any other observable is equivalent to the information in a single differential distribution of that observable. This is shown in the bottom row. As expected, it is also identical to the information for the correlation of an observable with itself, which is shown in the diagonal. We can identify $\Delta\eta_{t\bar{t}}$ and θ^* as the two most sensitive observables, which individually carry about 60% of the full information.

Combing two different observables further increases the information. The two most promising combinations are $\Delta\eta_{t\bar{t}}$ vs p_{Th} as well as θ^* vs $m_{t\bar{t}h}$, which carry about 73% of the full information. Successively adding more observables further increases the information. This shows that a multivariate analysis is vital to maximize the sensitivity on the *CP* phase α.

III. KINEMATIC RECONSTRUCTION

Most of the new physics probes discussed so far, viz. m_{tt}, θ^*, b_4, and $\Delta\phi_{ik}^{t\bar{t}}$, require a full reconstruction of the top and antitop momenta. This is a challenging task at the LHC due to combinatorial ambiguities and the presence of up to two neutrinos in the $t\bar{t}(h \to \gamma\gamma)$ final state. In this section, we discuss the strategies adopted for the kinematic reconstruction of the semileptonic and hadronic channels and the more complex dileptonic mode.

A. Semileptonic channel

In the semileptonic channel, the full reconstruction of the $t\bar{t}$ system requires the determination of the longitudinal momentum of the missing particle ν. We compute it by constraining the invariant mass of the lepton and the neutrino to the W-boson mass. Typically, either two solutions or zero solutions are obtained. Around 35% of events give zero solutions, and discarding all such events would lead to a significant reduction in event statistics. Therefore, in such events, we vary the transverse momentum of the missing system (at most by $\pm 10\%$) while keeping the azimuth angle of ν unchanged until physical solutions of $p_{z,\nu}$ are obtained. Events which give zero solutions even after the aforesaid variation are ignored. We perform the reconstruction for the top quarks iterating over all possible partitions of light jets (j) and b jet forming the hadronic top (jjb) and leptons and b jet for the leptonic top ($\ell\nu b$). The two possible neutrino solutions are separately accounted for, forming different partitions. We select the combination that minimizes

$$(m_{jjb} - m_t)^2 + (m_{\ell\nu b} - m_t)^2, \quad (7)$$

where m_t is the on shell mass of the top quark.

B. Hadronic channel

We follow a similar mass minimization strategy in the hadronic channel. We reconstruct the two top quarks, t_1 and t_2, by iterating over all possible combinations of light jets and b jets. The combination that minimizes

$$(m_{t_1} - m_t)^2 + (m_{t_2} - m_t)^2 \quad (8)$$

is chosen.

C. Dileptonic channel

In the more complex dileptonic $t\bar{t}h$ channel, the invisible system is constituted by two neutrinos. Therefore, in addition to determining the unknown longitudinal momentum of the missing particles, it is also indispensable to partition the four-momentum of the missing system into the two neutrinos in order to fully reconstruct the top and the antitop. An additional combinatorial ambiguity arises from the tandem b jet and ℓ pairing. The study in Ref. [24] reconstructed the $t\bar{t}(h \to b\bar{b})$ system in dileptonic mode using a M_2 assisted reconstruction algorithm and a boosted $h \to b\bar{b}$, with jet substructure techniques, to suppress the additional combinatorics between the Higgs boson and top quark decays. In contrast, the present analysis reconstructs the $t\bar{t}(h \to \gamma\gamma)$ system following the recursive jigsaw reconstruction algorithm presented in Ref. [55]. The recursive jigsaw reconstruction approach utilizes a series of jigsaw rules optimized to estimate the unknown kinematic degrees of freedom in an event topology and resolve the combinatorial ambiguities between/within the final state

visible and invisible objects. It results is a complete kinematic basis that can be used to define the four-momenta of all the final state and intermediate objects in an event decay tree.

The first step involves the resolution of combinatorial ambiguity between the b jets and the leptons by using the "combinatorial minimization" jigsaw rule (JR) [55], identifying the (b-jet, ℓ) pairs by minimizing

$$(m^2_{b_j\ell^+} + m^2_{b_k\ell^-}), \qquad j,k = 1,2, \qquad j \neq k. \tag{9}$$

After establishing the two visible hemispheres corresponding to the top and the antitop, we apply the "invisible mass" JR to estimate the invariant mass of the invisible system (m_I) [55] defined as

$$m_I^2 = m_V^2 - 4m^2_{V_a} m^2_{V_b}, \tag{10}$$

where m_V is the invariant mass of all the two b-tagged jets and the two leptons in the final state. m_{V_a} and m_{V_b} correspond to the invariant mass of the two visible hemispheres associated with the top and the antitop that were reconstructed in the previous step. m_I is chosen such that it is the smallest Lorentz invariant mass that ensures a nontachyonic four-momenta for the individual neutrinos upon partitioning the invisible system. Next, we determine the longitudinal momentum of the invisible system, $\not{p}_z$, using the following relation given by the "invisible rapidity" JR [55]:

$$\not{p}_z = p_z^V \frac{\sqrt{|\not{p}_T|^2 + m_I^2}}{\sqrt{|p_T^V|^2 + m_V^2}}. \tag{11}$$

Here, p_z^V and p_T^V represent the longitudinal and transverse momenta, respectively, of the visible system constituted by the two b jets and the two leptons, while $\not{p}_T$ is the missing transverse momentum.

At this point, we have all the ingredients required to reconstruct the $t\bar{t}$ system. However, in order to reconstruct the top and the antitop individually, the invisible four-momentum has to be correctly partitioned into the two neutrinos. This is achieved by using the "contraboost invariant" JR specified in Ref. [55] that estimates the four-momenta of the neutrinos produced from top and antitop decay in the $t\bar{t}$ c.m. frame under the assumption that both t and $\bar{t}$ have the same invariant mass. The resolved four-momenta of the neutrinos along with the correctly paired b jets and leptons allows defining the t and the $\bar{t}$ systems independently. The reconstruction efficiency of this method is about 80%, which is comparable with M_2 assisted reconstruction algorithm [24].

With the fully resolved $t\bar{t}h$ system, we can reconstruct a multitude of CP-even and CP-odd spin correlation observables defined in the $t\bar{t}$ c.m. frame and the lab frame. Several observables that do not depend on the spin-polarization of $t\bar{t}$ pair are also considered. Our goal here is to maximally explore the $t\bar{t}h$ multiparticle final state, augmenting the CP sensitivity of the $pp \to t\bar{t}h$ channel at the HL-LHC.

IV. ANALYSIS

A. Simulation and event selection

In this section, we explore the direct Higgs boson-top CP measurement combining machine learning techniques and efficient kinematic reconstruction methods. We consider $t\bar{t}h$ signal with $h \to \gamma\gamma$ in the dileptonic, semileptonic, and hadronic top decay modes at the HL-LHC. The dominant background to this process is given by continuum $t\bar{t}\gamma\gamma$ production. We simulate both the signal and background event samples with MadGraph5_aMC@NLO [56] within the MadMiner framework [50] at leading order with a center-of-mass energy of $\sqrt{s} = 14$ TeV. Higher order effects to the signal rate are included via a flat next-to-leading order k factor [57,58]. We use the NNPDF2.3QED parton distribution function [59]. No generation-level cuts have been applied for the signal events, while the backgrounds have been generated in the mass window $105\ \text{GeV} < m_{\gamma\gamma} < 145\ \text{GeV}$. Parton shower and hadronization effects have been included with PYTHIA 8 [60] and fast detector simulation with the DELPHES3 package [61], using the default HL-LHC detector card [33,62].

To obtain the cross section and likelihood function as a function of the theory parameters, we use the morphing technique that is already implemented in MadMiner. Here, we take into account the dependence of new physics theory parameters at both $t\bar{t}h$ production and $h \to \gamma\gamma$ decay, and therefore choose a quartic ansatz in the morphing setup, which is used to interpolate the event weights as a function of $\kappa_H = \kappa_t \cos\alpha$ and $\kappa_A = \kappa_t \sin\alpha$.

We start our analysis by selecting events consisting of two photons and at least two b-tagged jets. In addition, we require the final state to contain exactly two opposite-sign leptons for the dileptonic channel, exactly one lepton and at least two light jets for the semileptonic channel, and at least four light jets for the hadronic channel. We demand the individual particles to pass the following identification cuts:

$$\begin{aligned} p_{T\ell} &> 15\ \text{GeV}, & |\eta_\ell| &< 4, \\ p_{T\gamma} &> 15\ \text{GeV}, & |\eta_\gamma| &< 4, \\ p_{Tb} &> 25\ \text{GeV}, & |\eta_b| &< 4, \\ p_{Tj} &> 25\ \text{GeV}, & |\eta_j| &< 5. \end{aligned} \tag{12}$$

In addition, we require the diphoton invariant mass to satisfy $|m_{\gamma\gamma} - 125| < 10$ GeV.

We fully reconstruct the $t\bar{t}h$ system following the strategy described in Sec. III. In particular, this allows us to obtain both the lab frame and the $t\bar{t}$ c.m. frame observables. As an example for an observable that

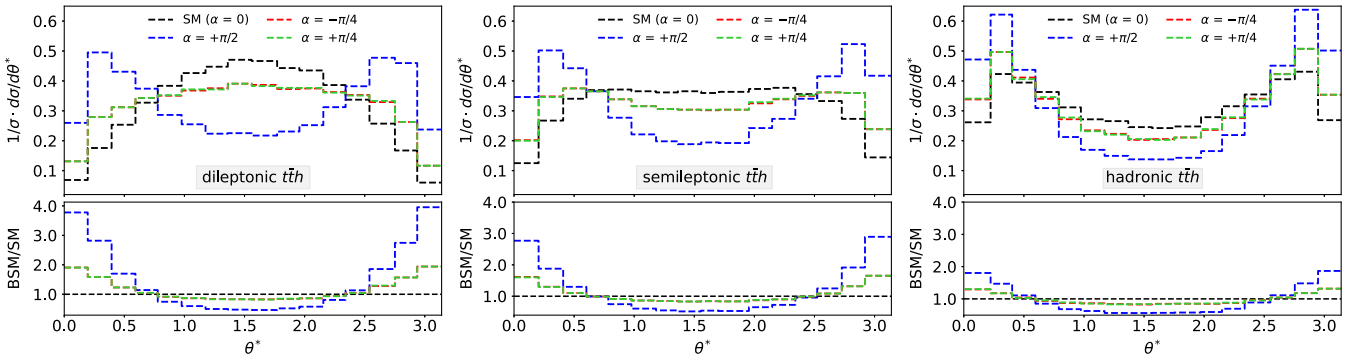

FIG. 3. Reconstructed detector level distributions for the Collins-Soper angle θ^* for the dileptonic channel (left), semileptonic channel (center), and hadronic channel (right). We present for the SM ($\alpha = 0$) and several beyond the SM Higgs boson-top *CP* hypotheses ($\alpha = \pi/2, \pm\pi/4$).

requires the top reconstruction, we present the distribution of the Collins-Soper angle θ^* in Fig. 3. When comparing these detector level distributions to the result at parton level, presented in Fig. 1, we observe the robustness of our analysis in respect to the reconstruction strategy and detector effects. The distributions are found to retain the *CP* sensitivity at the detector level, albeit a reduction of about 20% for the dileptonic channel, 40% for the semileptonic channel and 50% for the hadronic channel, compared to parton level.

B. Analysis methodology

As we have seen in Sec. II, there is no single observable that carries all the information on the *CP* structure of the top quark Yukawa. Instead, there is a variety of sensitive observables. Hence, a multivariate analysis is needed to extract the maximal information on the theory parameters from the data. In the following, we will summarize the adopted observables and the analysis methodology.

In this analysis, we consider the following list of 80 observables to describe the kinematics of signal and background events.

$$\begin{aligned}\textbf{Observables}:\ & \Delta\phi^{t\bar{t}}_{ik}, \quad \Delta\phi^{t\bar{t}}_{hi(k)}, \quad \Delta\phi^{\mathrm{lab}}_{ik}, \\ & \Delta\phi_{ht/h\bar{t}}, \quad \Delta\phi_{t\bar{t}}, \theta^*, \quad b_4, \\ & m_{\gamma\gamma}, \quad m_{t\bar{t}}, \quad m_{t\bar{t}h}, \quad \Delta R^{\mathrm{min}}_{\gamma j}, \\ & \Delta R_{j_{\mathrm{soft}}j_{\mathrm{hard}}}, \quad \Delta R_{\ell\nu}, \quad \Delta R_{Wb}, \quad \Delta R^{\mathrm{min}}_{\gamma j}, \\ & \Delta\eta_{t\bar{t}}, \quad m_{ht/h\bar{t}}, \quad H_T, \quad \not{E}_T/\sqrt{H_T}, \\ & \{p_T, \eta\}_X \quad \text{for } X = i, k, t, \bar{t}, h. \end{aligned} \tag{13}$$

We include the complete set of observables used by the ATLAS collaboration in a recent Higgs boson-top *CP* study [34] and complement this set with additional *CP*-even observables that show strong sensitivity to the *CP* phase $(\theta^*, b_4, m_{t\bar{t}}, m_{t\bar{t}h})$ together with the transverse momentum and pseudorapidity of all final state and reconstructed objects. We also incorporate a comprehensive list of spin correlations, as introduced in Eq. (4), which are constructed between all possible final state pairs. We include both observables constructed in the $t\bar{t}$ rest frame, $\Delta\phi^{t\bar{t}}_{ik}$, and in the lab frame, $\Delta\phi^{\mathrm{lab}}_{ik}$. Finally, we account for the correlation observables $\Delta\phi^{t\bar{t}}_{hi(k)}$ that arise from the tensor products involving the Higgs boson momentum, $\epsilon(p_t, p_{\bar{t}}, p_h, p_{i(k)})$. The following pairs $\{i, k\}$ are considered for the different channels

$$\begin{aligned} \text{dileptonic}:\ & i = \ell^+, \nu_t, b_t, W_t, \\ & k = \ell^-, \nu_{\bar{t}}, b_{\bar{t}}, W_{\bar{t}}, \\ \text{semileptonic}:\ & i = \ell, \nu, b_\ell, W_\ell, \\ & k = j_{\mathrm{soft}}, j_{\mathrm{hard}}, b_{\mathrm{had}}, W_{\mathrm{had}}, \\ \text{hadronic}:\ & i = j^{t_1}_{\mathrm{soft}}, j^{t_1}_{\mathrm{hard}}, b_{t_1}, W_{t_1}, \\ & k = j^{t_2}_{\mathrm{soft}}, j^{t_2}_{\mathrm{hard}}, b_{t_2}, W_{t_2}. \end{aligned} \tag{14}$$

In the semileptonic case, b_ℓ/W_ℓ and $b_{\mathrm{had}}/W_{\mathrm{had}}$ represent the b jets/W bosons produced from the leptonically and hadronically decaying top quarks, respectively. In events with more than two b-tagged jets, the hardest two are considered while reconstructing the top and the antitop quarks. j_{hard} corresponds to the hardest light jet, from the hadronic top quark, in the top rest frame.

To interpret the results of our analysis and obtain projected sensitivities, we follow a likelihood-based approach. According to the Neyman-Pearson lemma, the most powerful test statistic to discriminate between two hypotheses, in our case a new physics model parametrized by $\theta = (\kappa_t, \alpha)$ and the SM with $\theta_{\mathrm{SM}} = (1, 0)$, is the likelihood ratio $r(x|\theta; \theta_{\mathrm{SM}})$. Here, x denotes the set of reconstructed observables introduced above.

Whereas the likelihood ratio involving detector level observables is intractable, meaning that it cannot be computed directly, it can be estimated using simulations. To address this issue, we use the machine-learning-based

technique introduced in Refs. [51–53,63–66], which has been implemented in the MadMiner tool [50]. This approach uses both reconstructed observables and matrix-element information, which are then used to train neural network models that estimate the likelihood ratio. It therefore accounts for parton shower, hadronization, and detector effects, while the matrix-element information helps to significantly improve the performance of the neural network training. Using the estimated likelihood ratio function $r(x|\theta;\theta_{\rm SM})$, which describes both the linear and nonlinear new physics effects, we then perform a likelihood ratio test to obtain our projected sensitivities.

We simulate 10^6 signal and 10^6 background events before event selection. Using MadMiner, we train neural networks to estimate the likelihood ratio using the ALICES algorithm with its hyperparameter set to unity [63]. We use fully connected neural networks with three hidden layers, each containing 100 nodes and tanh activation function. The neural network training is performed over 100 epochs using the Adam optimizer. To avoid overtraining, we evaluate the loss function on an independent validation set and employ an early stopping procedure. We use a batch size of 128, and an exponentially decaying learning rate (from 10^{-4} to 10^{-5}). The limit setting is performed with MadMiner's LIKELIHOOD class.

C. Results

Let us now turn to the results of our study. In Fig. 4 we show the projected sensitivity on the top Yukawa coupling in terms of κ_t and α using the $t\bar{t}(h \to \gamma\gamma)$ measurement. In the left panel, we present the 68% C.L. contours for the individual top decay channels as colored dashed lines. A combination of all channels is shown in the black solid line. The studied channels can be organized in ascending order of sensitivity as dileptonic, semileptonic, and hadronic modes. Since the leading observables display efficient reconstruction for all channels, as illustrated in Fig. 3, the order of sensitivity among the final state modes follow their correspondent event rate.

In the right panel, we show the 68% and 95% C.L. contours as dashed and solid lines, respectively. The p values in the (κ_t, α) parameter space are presented through the color palette. We observe that $|\kappa_t|$ can be constrained within $\mathcal{O}(8\%)$ of the SM value at 68% C.L. through a combination of direct searches in the $t\bar{t}(h \to \gamma\gamma)$ channel at the HL-LHC. Assuming $\kappa_t = 1$, the combined search would be able to probe the Higgs boson-top CP phase up to $|\alpha| \lesssim 13°$ at 68% C.L.

We also perform a separate analysis in which we train a neural network exclusively with the CP-even observables shown in the right panel of Fig. 2. We observe that the projected sensitivity of such an analysis, using this smaller set of CP-even observables that are most sensitive to the nonlinear new physics effects, is almost comparable to the projected sensitivity of the combination study performed using the full set of observables. Overall, almost all the sensitivity to the Higgs boson-top CP structure is provided by the nonlinear terms in α. The limited $t\bar{t}(h \to \gamma\gamma)$ event statistics renders subleading sensitivity to the observables that probe the linear terms.

D. Systematic effects

In this section, we explore the implications from systematic uncertainties on the projected sensitivity of κ_t and

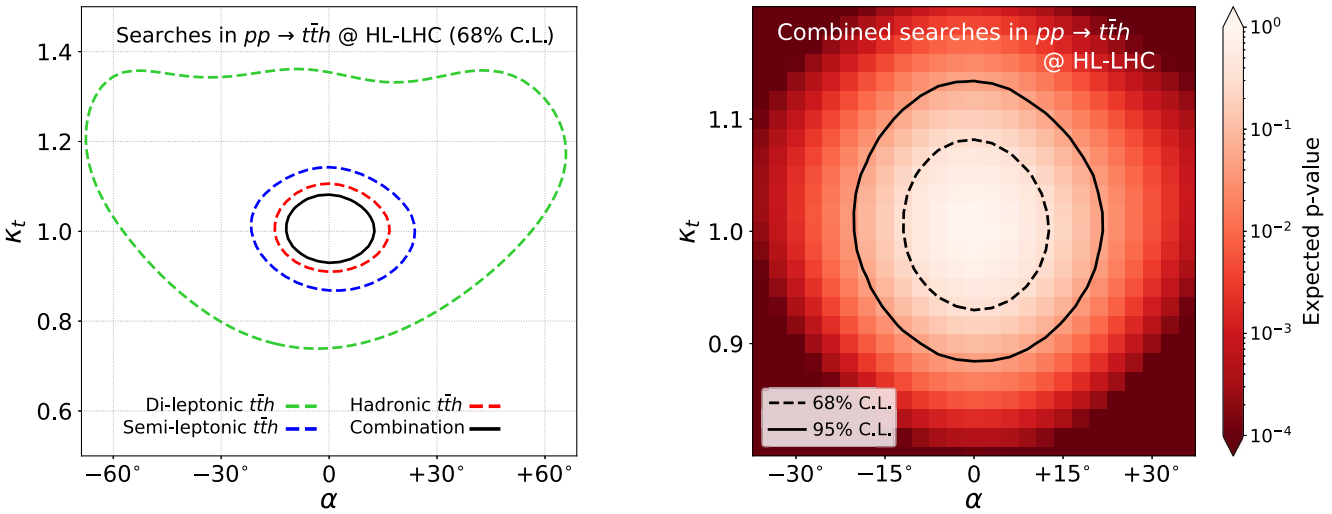

FIG. 4. Projected sensitivity in the (α, κ_t) plane. In the left panel we show the projected 68% C.L. contours from direct Higgs boson-top searches in the semileptonic (blue), dileptonic (green), hadronic (red) $t\bar{t}h$ channels, and their combination (black), considering all input observables. The right panel shows the projected 68% C.L. (dashed) and 95% C.L. (solid) contours from the combination of the three channels considering all input observables. The color palette illustrates the expected p value of the estimated log-likelihood ratio. The projections are derived for 14 TeV LHC assuming an integrated luminosity of 3 ab^{-1}.

α. In particular, we will consider two sources of uncertainty associated with the normalization of both signal and background.

In the statistical analysis, these uncertainties are parametrized through nuisance parameters ν_s and ν_b for the signal and background normalization, respectively. These nuisance parameters encode theoretical and experimental uncertainties on the normalization of distributions, neglecting possible shape uncertainties. As before, we train a neural network using the ALICES method in MadMiner to estimate the likelihood ratio $r(x|\theta,\nu;\theta_{\rm SM},\nu_{\rm SM})$. This is now a function of both the model parameters $\theta=(\kappa_t,\alpha)$ and the nuisance parameters $\nu=(\nu_s,\nu_b)$ that have a nominal value $\nu_{\rm SM}=(0,0)$. Before setting limits, a constraint term describing our prior knowledge on the nuisance parameter is added. Adopting a conservative approach, we assume a prior constraint of 20% and 50% in the $t\bar{t}(h\to\gamma\gamma)$ signal and the $t\bar{t}\gamma\gamma$ background, respectively. Finally, we profile over the nuisance parameters following the procedure described in Ref. [51].

Before turning to the sensitivity contours, let us remind ourselves that the presented results are based on a multivariate analysis. In particular, this includes the invariant mass of the diphoton pair. The considered range, $115\ {\rm GeV}<m_{\gamma\gamma}<135\ {\rm GeV}$, was chosen sufficiently wide to contain both a signal dominated region at the Higgs resonance and a background dominated region around it. MadMiner uses this background dominated region to constrain the nuisance parameter associated with the background normalization ν_b, and therefore effectively performs a data-driven sideband analysis. As we will see in a moment, the effective uncertainty of the background normalization is therefore significantly smaller than the 50% which we assumed as a prior.

In the following, we analyze three scenarios to study the impact of systematic uncertainties on the projected sensitivity in the (α,κ_t) plane. In the first scenario, we study the impact associated with only the uncertainty on the background normalization. To do so, we fix the $\nu_s=0$ in the estimated likelihood ratio and profile over ν_b. Similarly, in a second scenario, we fix $\nu_b=0$ and we profile over ν_s to study the impact of the signal uncertainty. Finally, in a third scenario, we obtain the limits after profiling the likelihood ratio over both ν_s and ν_b. In Fig. 5 we present the projected sensitivity on α and κ_t for all scenarios. The blue, green, and red contours correspond to the first, second, and third scenarios, respectively. The black contour represents the sensitivity assuming no systematic uncertainty and corresponds to the black-solid contours in Fig. 4.

At first, we observe that the sensitivity in α remains unaffected from systematic uncertainties [32]. This stems from the reason that at $\kappa_t=1$, the sensitivity in α is dominantly controlled by the shape information from kinematic distributions and is largely independent of the event rate due to the combination of two competing effects.

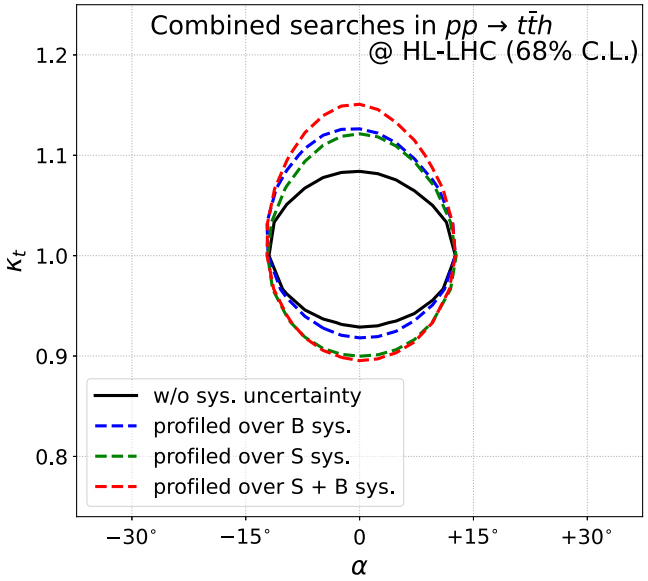

FIG. 5. Projected 68% C.L. contours on (α,κ_t) from the combination of the three channels considering all input observables, profiled over background (B) uncertainties (blue), signal (S) uncertainties (green), and both signal and background systematic uncertainties (red). The projections are derived for 14 TeV LHC assuming $\mathcal{L}=3\ {\rm ab}^{-1}$.

On the one hand, the signal cross section $\sigma_{t\bar{t}(h\to\gamma\gamma)}$ decreases with α: for example at $\kappa_t=1$ the cross section $\sigma_{t\bar{t}(h\to\gamma\gamma)}$ falls by $\mathcal{O}(25\%)$ from $\alpha=0$ until $\alpha\sim\pi/3$ and then remains roughly unchanged until $\alpha=\pi/2$. On the other hand, the signal efficiency also improves with α. These two effects roughly offset any overall dependence on the event rate, thereby leading to unchanged projection contours in the direction of α even after profiling over the nuisance parameters.

The situation is qualitatively different in the κ_t direction. When $\alpha=0$, the measurement is purely based on a rate information, implying that the Higgs coupling strength κ_t and the signal normalization, as parametrized by ν_s, are essentially degenerate. Therefore, our prior uncertainty of the signal normalization will directly propagate into a systematic uncertainty on κ_t. The new physics effects in the Higgs boson-top coupling manifest as $\sim\kappa_t^2$ at the $t\bar{t}h$ production level and as $\sim(1.28-0.28\kappa_t)^2$ in $h\to\gamma\gamma$ decay [7]. After combining these two factors, an uncertainty of 20% in the $pp\to t\bar{t}(h\to\gamma\gamma)$ cross section translates to roughly 12% uncertainty in κ_t. We observe this effect in Fig. 5: for $\alpha=0$ the projected sensitivity falls from $|\kappa_t|\lesssim 8\%$ in the absence of systematic uncertainties to $|\kappa_t|\lesssim 13\%$ on profiling over ν_s. We observe that, despite a prior 50% uncertainty in the background normalization compared to 20% in the signal, its impact on the projection contours in the (κ_t,α) plane is milder. As discussed above, this is a consequence of the sideband measurement and illustrates the robustness of our multivariate analysis.

V. SUMMARY

In this study, we derived the prospects to *direct* measure the Higgs boson-top CP structure in the $t\bar{t}(h \to \gamma\gamma)$ channel at the HL-LHC. We show that a combination of machine learning techniques and efficient kinematic reconstruction methods can boost new physics sensitivity, effectively exploring the complex $t\bar{t}h$ multiparticle phase space.

Among the several probes included in our machine learning analysis, this study encompass a comprehensive set of spin correlation observables. Beyond the SM CP phases steer the spin polarization of the top pair, and the spin correlations are carried forward by their decay products. We harness the potential of the spin correlation observables via the full reconstruction of the top and antitop, evaluating these particular observables in the $t\bar{t}$ c.m. frame, where the correlations are maximal. In the hadronic and semileptonic $t\bar{t}h$ channels, we used mass minimization to fully reconstruct the $t\bar{t}h$ system. In the more complex dileptonic channel, we employed the recursive jigsaw reconstruction technique to resolve the combinatorial ambiguities and determine the unknown degrees of freedom. In all channels, the effects of parton showering, hadronization, and detector resolution were included.

Exploring the intricate $t\bar{t}h$ multiparticle phase space with CP-odd and even observables defined in the laboratory frame and the $t\bar{t}$ c.m. frame, we obtain strong projections for the Higgs boson-top CP phase. Through a combined semileptonic, hadronic, and dileptonic $t\bar{t}(h \to \gamma\gamma)$ search, the HL-LHC can directly probe the Higgs boson-top coupling modifier and CP phase, respectively, up to $|\kappa_t| \lesssim 8\%$ and $|\alpha| \lesssim 13°$ at 68% C.L.

Possible improvements can be expected by including other relevant channels, such as $t\bar{t}(h \to b\bar{b})$ [6,21,24,32]. While this channel displays the bulk of the Higgs decay, $\mathcal{BR}(h \to b\bar{b}) \sim 58\%$, it results in subleading limits in comparison to $t\bar{t}(h \to \gamma\gamma)$ as it endures a substantial QCD background that is associated with sizable uncertainties [67,68]. Fast-moving precision calculations [69–71] and possible combination of sideband analysis with $t\bar{t}h/t\bar{t}Z$ ratios [32,72] may change this scenario, controlling the respective uncertainties, and pushing further forward the sensitivity with this extra channel in the near future.

ACKNOWLEDGMENTS

We thank Johann Brehmer and Sam Homiller for helpful discussions. R. K. B. and D. G. thank the U.S. Department of Energy for the financial support, under Grant No. DE-SC 0016013. The work of F. K. is supported by the U.S. Department of Energy under Grant No. DE-AC02-76SF00515 and by the Deutsche Forschungsgemeinschaft under Germanys Excellence Strategy—EXC 2121 Quantum Universe—Grant No. 390833306. Part of this work was performed at the Aspen Center for Physics, which is supported by National Science Foundation Grant No. PHY-1607611. Some computing for this project was performed at the High Performance Computing Center at Oklahoma State University, supported in part through the National Science Foundation Grant No. OAC-1531128.

[1] A. D. Sakharov, Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe, Sov. Phys. Usp. **34**, 392 (1991).
[2] K. Kajantie, M. Laine, K. Rummukainen, and M. E. Shaposhnikov, Is There a Hot Electroweak Phase Transition at m(H) Larger or Equal to m(W)?, Phys. Rev. Lett. **77**, 2887 (1996).
[3] P. Huet and E. Sather, Electroweak baryogenesis and standard model CP violation, Phys. Rev. D **51**, 379 (1995).
[4] W. Buchmuller and D. Wyler, Effective lagrangian analysis of new interactions and flavor conservation, Nucl. Phys. **B268**, 621 (1986).
[5] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, Dimension-six terms in the standard model lagrangian, J. High Energy Phys. 10 (2010) 085.
[6] M. R. Buckley and D. Goncalves, Boosting the Direct CP Measurement of the Higgs-Top Coupling, Phys. Rev. Lett. **116**, 091801 (2016).
[7] J. Brod, U. Haisch, and J. Zupan, Constraints on CP-violating Higgs couplings to the third generation, J. High Energy Phys. 11 (2013) 180.
[8] M. J. Dolan, P. Harris, M. Jankowiak, and M. Spannowsky, Constraining CP-violating Higgs sectors at the LHC using gluon fusion, Phys. Rev. D **90**, 073008 (2014).
[9] C. Englert, D. Goncalves-Netto, K. Mawatari, and T. Plehn, Higgs quantum numbers in weak boson fusion, J. High Energy Phys. 01 (2013) 148.
[10] A. Kobakhidze, N. Liu, L. Wu, and J. Yue, Implications of CP-violating top-Higgs couplings at LHC and Higgs factories, Phys. Rev. D **95**, 015016 (2017).
[11] F. U. Bernlochner, C. Englert, C. Hays, K. Lohwasser, H. Mildner, A. Pilkington, D. D. Price, and M. Spannowsky, Angles on CP-violation in Higgs boson interactions, Phys. Lett. B **790**, 372 (2019).
[12] C. Englert, P. Galler, A. Pilkington, and M. Spannowsky, Approaching robust EFT limits for CP-violation in the Higgs sector, Phys. Rev. D **99**, 095007 (2019).
[13] A. V. Gritsan, J. Roskes, U. Sarica, M. Schulze, M. Xiao, and Y. Zhou, New features in the JHU generator framework: Constraining Higgs boson properties from on-shell and off-shell production, Phys. Rev. D **102**, 056022 (2020).

[14] H. Bahl, P. Bechtle, S. Heinemeyer, J. Katzy, T. Klingl, K. Peters, M. Saimpert, T. Stefaniak, and G. Weiglein, Indirect $\mathcal{CP}$ probes of the Higgs-top-quark interaction: Current LHC constraints and future opportunities, J. High Energy Phys. 11 (2020) 127.

[15] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi, Disentangling Higgs-top couplings in associated production, J. High Energy Phys. 04 (2014) 004.

[16] F. Boudjema, R. M. Godbole, D. Guadagnoli, and K. A. Mohan, Lab-frame observables for probing the top-Higgs interaction, Phys. Rev. D **92**, 015019 (2015).

[17] M. R. Buckley and D. Goncalves, Constraining the strength and *CP* structure of dark production at the LHC: The associated top-pair channel, Phys. Rev. D **93**, 034003 (2016).

[18] A. V. Gritsan, R. Röntsch, M. Schulze, and M. Xiao, Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques, Phys. Rev. D **94**, 055023 (2016).

[19] D. Goncalves and D. Lopez-Val, Pseudoscalar searches with dileptonic tops and jet substructure, Phys. Rev. D **94**, 095005 (2016).

[20] N. Mileo, K. Kiers, A. Szynkman, D. Crane, and E. Gegner, Pseudoscalar top-Higgs coupling: Exploration of *CP*-odd observables to resolve the sign ambiguity, J. High Energy Phys. 07 (2016) 056.

[21] S. A. Dos Santos *et al.*, Probing the *CP* nature of the Higgs coupling in $t\bar{t}h$ events at the LHC, Phys. Rev. D **96**, 013004 (2017).

[22] D. Azevedo, A. Onofre, F. Filthaut, and R. Gonçalo, *CP* tests of Higgs couplings in $t\bar{t}h$ semileptonic events at the LHC, Phys. Rev. D **98**, 033004 (2018).

[23] J. Li, Z.-G. Si, L. Wu, and J. Yue, Central-edge asymmetry as a probe of Higgs-top coupling in $t\bar{t}h$ production at the LHC, Phys. Lett. B **779**, 72 (2018).

[24] D. Gonçalves, K. Kong, and J. H. Kim, Probing the top-Higgs Yukawa *CP* structure in dileptonic $t\bar{t}h$ with M_2-assisted reconstruction, J. High Energy Phys. 06 (2018) 079.

[25] M. Aaboud *et al.* (ATLAS Collaboration), Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector, Phys. Lett. B **784**, 173 (2018).

[26] A. M. Sirunyan *et al.* (CMS Collaboration), Observation of $t\bar{t}$H Production, Phys. Rev. Lett. **120**, 231801 (2018).

[27] J. Ren, L. Wu, and J. M. Yang, Unveiling *CP* property of top-Higgs coupling with graph neural networks at the LHC, Phys. Lett. B **802**, 135198 (2020).

[28] B. Bortolato, J. F. Kamenik, N. Košnik, and A. Smolkovič, Optimized probes of *CP* -odd effects in the $t\bar{t}h$ process at hadron colliders, Nucl. Phys. **B964**, 115328 (2021).

[29] Q.-H. Cao, K.-P. Xie, H. Zhang, and R. Zhang, A new observable for measuring *CP* property of top-Higgs interaction, Chin. Phys. C **45**, 023117 (2021).

[30] R. M. Abraham, D. Gonçalves, T. Han, S. C. I. Leung, and H. Qin, Directly probing the Higgs-top coupling at high scales, Phys. Lett. B **825**, 136839 (2022).

[31] T. Martini, R.-Q. Pan, M. Schulze, and M. Xiao, Probing the *CP* structure of the top quark Yukawa coupling: Loop sensitivity versus on-shell sensitivity, Phys. Rev. D **104**, 055045 (2021).

[32] D. Gonçalves, J. H. Kim, K. Kong, and Y. Wu, Direct Higgs-top *CP*-phase measurement with $t\bar{t}h$ at the 14 TeV LHC and 100 TeV FCC, arXiv:2108.01083.

[33] M. Cepeda *et al.*, Report from working group 2: Higgs physics at the HL-LHC and HE-LHC, CERN Yellow Rep. Monogr. **7**, 221 (2019).

[34] G. Aad *et al.* (ATLAS Collaboration), *CP* Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \to \gamma\gamma$ with the ATLAS Detector, Phys. Rev. Lett. **125**, 061802 (2020).

[35] A. M. Sirunyan *et al.* (CMS Collaboration), Measurements of $t\bar{t}H$ Production and the *CP* Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel, Phys. Rev. Lett. **125**, 061801 (2020).

[36] S. Bar-Shalom, D. Atwood, G. Eilam, R. R. Mendel, and A. Soni, Large tree level *CP* violation in $e^+e^- \to t\bar{t}H^0$ in the two Higgs doublet model, Phys. Rev. D **53**, 1162 (1996).

[37] J. F. Gunion and X.-G. He, Determining the *CP* Nature of a Neutral Higgs Boson at the LHC, Phys. Rev. Lett. **76**, 4468 (1996).

[38] D. Atwood, S. Bar-Shalom, G. Eilam, and A. Soni, *CP* violation in top physics, Phys. Rep. **347**, 1 (2001).

[39] S. Berge, W. Bernreuther, and J. Ziethe, Determining the *CP* Parity of Higgs Bosons at the LHC in Their τ Decay Channels, Phys. Rev. Lett. **100**, 171605 (2008).

[40] G. Aad *et al.* (ATLAS Collaboration), *CP* Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \to \gamma\gamma$ with the ATLAS Detector, Phys. Rev. Lett. **125**, 061802 (2020).

[41] C. Degrande and J. Touchèque, A reduced basis for *CP* violation in SMEFT at colliders and its application to Diboson production, arXiv:2110.02993.

[42] P. A. Zyla *et al.* (Particle Data Group Collaboration), Review of particle physics, Prog. Theor. Exp. Phys. **2020**, 083C01 (2020).

[43] G. Mahlon and S. J. Parke, Spin correlation effects in top quark pair production at the LHC, Phys. Rev. D **81**, 074024 (2010).

[44] W. Bernreuther and Z.-G. Si, Distributions and correlations for top quark pair production and decay at the Tevatron and LHC, Nucl. Phys. **B837**, 90 (2010).

[45] M. Jezabek, Top quark physics, Nucl. Phys. B Proc. Suppl. **37**, 197 (1994).

[46] F. Demartin, F. Maltoni, K. Mawatari, B. Page, and M. Zaro, Higgs characterisation at NLO in QCD: *CP* properties of the top-quark Yukawa interaction, Eur. Phys. J. C **74**, 3065 (2014).

[47] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, Higgs production in association with a single top quark at the LHC, Eur. Phys. J. C **75**, 267 (2015).

[48] J. Brehmer, K. Cranmer, F. Kling, and T. Plehn, Better Higgs boson measurements through information geometry, Phys. Rev. D **95**, 073002 (2017).

[49] J. Brehmer, F. Kling, T. Plehn, and T. M. P. Tait, Better Higgs-CP tests through information geometry, Phys. Rev. D **97**, 095017 (2018).

[50] J. Brehmer, F. Kling, I. Espejo, and K. Cranmer, MadMiner: Machine learning-based inference for particle physics, Comput. Softw. Big Sci. **4**, 3 (2020).

[51] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, A guide to constraining effective field theories with machine learning, Phys. Rev. D **98**, 052004 (2018).
[52] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, Constraining Effective Field Theories with Machine Learning, Phys. Rev. Lett. **121**, 111801 (2018).
[53] J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, Mining gold from implicit models to improve likelihood-free inference, Proc. Natl. Acad. Sci. U.S.A. **117**, 5242 (2020).
[54] J. Brehmer, S. Dawson, S. Homiller, F. Kling, and T. Plehn, Benchmarking simplified template cross sections in WH production, J. High Energy Phys. 11 (2019) 034.
[55] P. Jackson and C. Rogan, Recursive jigsaw reconstruction: HEP event analysis in the presence of kinematic and combinatoric ambiguities, Phys. Rev. D **96**, 112007 (2017).
[56] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, J. High Energy Phys. 07 (2014) 079.
[57] D. de Florian *et al.* (LHC Higgs Cross Section Working Group Collaboration), Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector, arXiv:1610.07922.
[58] https://twiki.cern.ch/twiki/bin/view/LHCPhysics.
[59] R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo (NNPDF Collaboration), Parton distributions with QED corrections, Nucl. Phys. **B877**, 290 (2013).
[60] T. Sjostrand, S. Mrenna, and P. Z. Skands, A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. **178**, 852 (2008).
[61] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.
[62] https://github.com/delphes/delphes/releases/tag/3.4.2pre16.
[63] M. Stoye, J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, Likelihood-free inference with an improved cross-entropy estimator, arXiv:1808.00973.
[64] J. Brehmer, K. Cranmer, I. Espejo, F. Kling, G. Louppe, and J. Pavez, Effective LHC measurements with matrix elements and machine learning, J. Phys. Conf. Ser. **1525**, 012022 (2020).
[65] J. Brehmer, K. Cranmer, I. Espejo, A. Held, F. Kling, G. Louppe, and J. Pavez, Constraining effective field theories with machine learning, EPJ Web Conf. **245**, 06026 (2020).
[66] J. Brehmer, K. Cranmer, and F. Kling, Improving inference with matrix elements and machine learning, Int. J. Mod. Phys. A **35**, 2041008 (2020).
[67] M. Aaboud *et al.* (ATLAS Collaboration), Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Phys. Rev. D **97**, 072016 (2018).
[68] A. M. Sirunyan *et al.* (CMS Collaboration), Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 03 (2019) 026.
[69] T. Ježo, J. M. Lindert, N. Moretti, and S. Pozzorini, New NLOPS predictions for $t\bar{t} + b$ -jet production at the LHC, Eur. Phys. J. C **78**, 502 (2018).
[70] A. Denner, J.-N. Lang, and M. Pellen, Full NLO QCD corrections to off-shell ttbb production, Phys. Rev. D **104**, 056018 (2021).
[71] G. Bevilacqua, H.-Y. Bi, H. B. Hartanto, M. Kraus, M. Lupattelli, and M. Worek, $t\bar{t}b\bar{b}$ at the LHC: On the size of corrections and b-jet definitions, J. High Energy Phys. 08 (2021) 008.
[72] M. L. Mangano, T. Plehn, P. Reimitz, T. Schell, and H.-S. Shao, Measuring the top yukawa coupling at 100 TeV, J. Phys. G **43**, 035001 (2016).