DEUTSCHES ELEKTRONEN-SYNCHROTRON
Ein Forschungszentrum der Helmholtz-Gemeinschaft

DESY 21-160
EFI 21-6
arXiv:2110.10177
October 2021

Constraining $\mathcal{CP}$-Violation in the Higgs-Top-Quark Interaction Using Machine-Learning-Based Inference

H. Bahl

Department of Physics, University of Chicago, USA

and

Deutsches Elektronen-Synchrotron DESY, Hamburg

S. Brass

Deutsches Elektronen-Synchrotron DESY, Hamburg

ISSN 0418-9833

NOTKESTRASSE 85 - 22607 HAMBURG

To be sure that your reports and preprints are promptly included in the HEP literature database
send them to (if possible by air mail):

DESY Zentralbibliothek Notkestraße 85 22607 Hamburg Germany	DESY Bibliothek Platanenallee 6 15738 Zeuthen Germany

DESY 21-160
EFI 21-6

Constraining $\mathcal{CP}$-violation in the Higgs–top-quark interaction using machine-learning-based inference

Henning Bahl[*1,2] and Simon Brass[†2]

[1] *University of Chicago, Department of Physics, 5720 South Ellis Avenue, Chicago, IL 60637 USA*
[2] *Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany*

While $\mathcal{CP}$ violation in the Higgs interactions with massive vector boson is already tightly constrained, the $\mathcal{CP}$ nature of the Higgs interactions with fermions is far less constrained. In this work, we assess the potential of machine-learning-based inference methods to constrain $\mathcal{CP}$ violation in the Higgs top-Yukawa coupling. This approach enables the use of the full available kinematic information. Concentrating on top-associated Higgs production with the Higgs decaying to two photons, we derive expected exclusion bounds for the LHC and the high-luminosity phase of the LHC. We also study the dependence of these bounds on the Higgs interaction with massive vector bosons and their robustness against theoretical uncertainties. In addition to deriving expected exclusion bounds, we discuss at which level a non-zero $\mathcal{CP}$-violating top-Yukawa coupling can be distinguished from the SM. Moreover, we analyze which kinematic distributions are most sensitive to a $\mathcal{CP}$-violating top-Yukawa coupling.

[*]hbahl@uchicago.edu
[†]simon.brass@desy.de

1 Introduction

After the discovery of a particle consistent with the Standard Model (SM) Higgs boson — within the current theoretical and experimental uncertainties — by the ATLAS and CMS collaborations at the LHC [1, 2], the determination of its couplings and quantum numbers is one of the most important task for future LHC runs (as well as future colliders). One especially interesting property is the $\mathcal{CP}$ nature of the Higgs boson interactions. While the amount of $\mathcal{CP}$ violation in the SM is not sufficient to explain the baryon asymmetry of the Universe [3, 4], the interactions of the Higgs boson can provide additional sources of $\mathcal{CP}$ violation in many beyond SM (BSM) theories.

$\mathcal{CP}$ violation in the Higgs sector can be constrained in two different ways: In the direct approach, $\mathcal{CP}$-odd observables are measured. For collider studies, these are typically constructed using angular distributions (e.g. for the Higgs decay into two tau leptons [5, 6]). Measuring a non-zero value for a $\mathcal{CP}$-odd observable is an unambiguous sign of $\mathcal{CP}$ violation. Therefore, such measurements, which are often challenging experimentally, are typically comparably model independent. As an alternative approach, the indirect approach can provide important complementary information. In the indirect approach, the effect of $\mathcal{CP}$-odd interactions on $\mathcal{CP}$-even observables is considered (e.g. the effect of a $\mathcal{CP}$-odd top-Yukawa coupling on the total top-associated Higgs production cross section). This indirect approach can yield strong constraints on $\mathcal{CP}$-violating interactions. It is, however, not guaranteed that deviations from the SM originate from $\mathcal{CP}$-violating interactions but they can in principle also be caused by $\mathcal{CP}$-even interactions. Consequently, these type of constraints are often associated with a larger model dependence.

Electric dipole moments (EDMs) are prominent examples for $\mathcal{CP}$-odd observables. While recent EDM measurements (see e.g. Refs. [7, 8]) severely constrain $\mathcal{CP}$ violation in the Higgs sector [9–12], different contributions to the electric dipole moments — originating e.g. from $\mathcal{CP}$ violation in more than one Higgs coupling — can cancel each other [13–15]. Therefore, searching for $\mathcal{CP}$ violation in the Higgs sector at colliders is an important complementary approach, especially since collider studies allow to disentangle the different couplings.

LHC studies have already ruled out the hypothesis of a pure $\mathcal{CP}$-odd Higgs boson [16, 17]. A $\mathcal{CP}$-mixed Higgs boson is, however, far less constrained. Most existing experimental studies concentrate on the Higgs interaction with massive vector bosons [17–23] (investigating Higgs production via vector-boson fusion and/or the Higgs decay into W or Z bosons). $\mathcal{CP}$ violation in the Higgs interaction with massive vector boson can, however, only be induced at the loop level in a generic BSM theory and is, therefore, expected to be small. In contrast, $\mathcal{CP}$ violation in the Higgs interaction with fermions can occur unsuppressed making an investigation of the Higgs–fermion–fermion interactions especially interesting from a phenomenological point of view. Existing experimental studies target $\mathcal{CP}$ violation in the Higgs coupling to tau leptons [5, 6] and the Higgs coupling to top quarks [23–26]. In addition, also $\mathcal{CP}$ violation in the effective Higgs coupling to gluons has been constrained experimentally [27].

In the present study, we focus on $\mathcal{CP}$ violation in the Higgs interaction with top quarks, a possibility which has received considerable attention in the literature [28–47]. While

a $\mathcal{CP}$-violating top-Yukawa coupling affects many processes (e.g. Higgs production via gluon fusion, Higgs decay into two photons, $t\bar{t}$ production, etc.), top-associated Higgs production is the most direct probe since the top-Yukawa coupling appears at the tree level. Even though $\mathcal{CP}$-odd variables for top-associated Higgs production have been proposed [48–51], the current experimental studies [23–26] — targeting the Higgs decay to two photons — exploit the effects of the $\mathcal{CP}$-odd Yukawa coupling on the kinematic distributions — effectively mixing $\mathcal{CP}$-even and $\mathcal{CP}$-odd observables.

From a technical point of view, the studies presented in Refs. [23–26] rely on the use of boosted decision trees (BDTs) in order to separate the signal process from background as well as to distinguish a $\mathcal{CP}$-conserving from a $\mathcal{CP}$-violating top-Yukawa interaction. While the use of BDTs is a well-established tool for collider analyses, this method relies on binning the events in a discrimination observable, which is optimized using the BDTs. This reduction of the high-dimensional data into a low-dimensional summary statistics is unavoidably associated with a loss of information.

Novel analysis techniques promise to use the full information available in the event data. The main goal of the present paper to evaluate the potential of machine-learning-based inference methods to probe a $\mathcal{CP}$-odd top-Yukawa coupling. This approach was developed in Refs. [52–56]. It uses machine learning to approximate the full likelihood fully taking into account parton shower as well as detector effects, which are only approximated in similar approaches like the matrix element method (see e.g. Refs. [36, 57–69]) or the optimal observable approach [70–72].

From a practical point of view, we use the implementation of machine-learning-based inference in the tool `Madminer` [73]. Similar to Refs. [23–25], we focus on the Higgs decay to photons. We, however, only consider events containing at least one lepton. While this choice reduces the expected number of events, it significantly simplifies our analysis. In principle, it is, however, straightforward to also take into account events with no lepton as well as other Higgs decay channels. In addition to a modified Higgs–top-quark interaction, we also allow for deviations of the Higgs interaction to massive vector bosons (HVV coupling) with respect to the SM, which is the most relevant Higgs coupling in top-associated Higgs production besides the top-Yukawa coupling. This allows us to study the dependence of the top-Yukawa coupling constraints on the HVV coupling. Moreover, we study the impact of theoretical uncertainties by treating the renormalization scale as a nuisance parameter.

Using this setup, we derive expected limits on a $\mathcal{CP}$-violating Yukawa coupling using the currently available luminosity. In addition, we derive projections using the full LHC and High-Luminosity LHC (HL-LHC) data. In addition to assuming SM data, we also investigate for an exemplary case of a $\mathcal{CP}$-admixed Higgs boson what amount of data is needed to establish a deviation from the SM. Moreover, we evaluate which observables are most sensitive to the $\mathcal{CP}$ character of the Higgs top-Yukawa coupling by evaluating the Fisher information.

Our paper is organized as follows. In Section 2, we introduce the effective model used for our study. The employed methods are reviewed in Section 3. We give details on the event simulation and selection in Section 4. The results are presented in Section 5. We conclude in Section 6.

2 Effective model

For the present study, we employ an effective Higgs model (based on the SM), which is closely related to the "Higgs characterisation" model defined in Refs. [30, 74, 75].

We are most interested in the top-Yukawa part of the Lagrangian which is given by

$$\mathcal{L}_{\text{top-yuk}} = -\frac{y_t^{\text{SM}}}{\sqrt{2}}\bar{t}\,(c_t + i\gamma_5\tilde{c}_t)\,tH. \tag{1}$$

Here, H is the Higgs field and t is the top-quark field. The prefactor y_t^{SM} is the SM top-Yukwawa coupling. Deviations from the SM are parameterized in terms of c_t and $\tilde{c}_t$. c_t modifies the $\mathcal{CP}$-even part of the top-Yukawa interaction, whereas $\tilde{c}_t$ induces a $\mathcal{CP}$-odd top-Yukawa interaction. The SM is recovered for $c_t = 1$ and $\tilde{c}_t = 0$. In an effective field theory framework, the deviations from SM can be thought of to be induced by dimension-6 operators (see e.g. Ref. [31]). c_t and $\tilde{c}_t$ correspond to κ_t and $\tilde{\kappa}_t$ as used in Ref. [24], respectively.

Instead of the coupling modifiers c_t and $\tilde{c}_t$, also an absolute value, denoted by $|g_t|$ and a $\mathcal{CP}$-violating phase α are often used to parameterize the top-Yukawa interaction. They are related to c_t and $\tilde{c}_t$ via

$$|g_t| \equiv \sqrt{c_t^2 + \tilde{c}_t^2}, \qquad \tan\alpha = \frac{\tilde{c}_t}{c_t}. \tag{2}$$

The quantities $|g_t|$ and α correspond to κ_t and α as used in Ref. [25].

Top-associated Higgs production — the target of this study —, however, depends not only on the Higgs–top-quark interaction but also on the Higgs interaction with massive vector bosons. We take into account a $SU(2)_L$ preserving modification of the SM interaction,

$$\mathcal{L}_V = c_V H\left(\frac{M_Z^2}{v}Z_\mu Z^\mu + 2\frac{M_W^2}{v}W_\mu^+ W^{-\mu}\right), \tag{3}$$

where Z and W denote the massive vector boson fields and $M_{Z,W}$ their respective masses ($v \simeq 246$ GeV is the Higgs vacuum expectation value). The Higgs interaction with massive vector bosons is rescaled by the common factor c_V. In principle, it is also possible to include additional operators going beyond the form of the SM operators (e.g. operators of the form $HZ_{\mu\nu}Z^{\mu\nu}$ with $Z_{\mu\nu}$ being the Z boson field strengths). Since the main focus of the present study is the Higgs–top-quark interaction, we omit these additional operators. Moreover, we do not expect that their presence would be distinguishable from a deviation in the SM-like HVV couplings (see Eq. (3)) when investigating top-associated Higgs production — the target process of the present study.

Note that the modified Higgs–top-quark interaction of Eq. (1) also affects the Higgs decay into two photons, which will be our target decay process in the next Section. The modification of the Higgs decay rate to two photons can result in stringent constraints on the parameter space (see e.g. Ref. [46]). In the present work, we will, however, assume the Higgs decay rate to two photons to be SM-like, which can e.g. be achieved by the presence of at least one electrically charged BSM particle decorrelating the Higgs decay to two photons from the Higgs–top-quark interaction.

3 Methodology

In this Section, we discuss the concept of machine-learning-based inference [52–56] (as implemented in the tool `Madminer`). The discussion follows closely Refs. [54, 55, 73], in which more details can be found.

3.1 Estimating the likelihood

When performing LHC measurements, a key object is the likelihood function $p_{\text{full}}(\{x_i\}|\theta)$ giving the probability of observing a set of events with the observables x_i for a given model with parameters θ. It can be written as

$$p_{\text{full}}(\{x_i\}|\theta) = \text{Pois}(n|L\sigma(\theta)) \prod_i p(x_i|\theta), \tag{4}$$

where n is the number of events, L the integrated luminosity, $\sigma(\theta)$ is the cross section as a function of the model parameters, and $\text{Pois}(n|\lambda) = \lambda^n e^{-\lambda}/n!$ is the probability mass function of the Poisson distribution. $p(x|\theta)$ denotes the probability density of observing a single event with observables x for a given model with parameters θ,

$$p(x|\theta) = \frac{1}{\sigma(x)} \frac{d^d\sigma(x|\theta)}{dx^d}, \tag{5}$$

where d is the dimension of the vector x. We can sample the distribution $p(x|\theta)$ by using Monte-Carlo (MC) simulators. Typically, these work in three steps: First, parton-level events are generated according to the matrix element of the process; second, the parton-level events are processed by a parton shower accounting for the effects of soft radiation and hadronization; third, the detector response is simulated. These steps can be written symbolically in the form

$$p(x|\theta) = \int dz_d \int dz_s \int dz_p \underbrace{p(x|z_d)p(z_d|z_s)p(z_s|z_p)p(z_p|\theta)}_{=p(x,z|\theta)} \tag{6}$$

with the latent variables z_d, z_p, and z_d. The variables z_p describe the parton-level observables; z_s, the parton-shower history; and z_d, the response of the detector. We denote the integrand as $p(x, z|\theta)$. Due to the large number of latent variables, the integral cannot be computed. Therefore, while $p(x|\theta)$ can be sampled, it cannot be computed directly.

Traditionally, this problem is circumvented by restricing the analysis to a low-dimensional summary statistics (e.g. the invariant mass of a tentative resonance). The likelihood $p(x|\theta)$ can then be estimated using histograms. Other approaches, like the matrix element method (see e.g. Refs. [36, 60, 61, 63, 65]) or the optimal observable approach [70–72] try to approximate the integrals over z_d and z_p in Eq. (6) by suitable transfer functions.

Machine-learning-based inference avoids using low-dimensional summary statistics or approximating the effects of the parton shower or the detector. Instead, the likelihood or

the likelihood ratio

$$r(x|\theta_0, \theta_1) = \frac{p(x|\theta_0)}{p(x|\theta_1)}, \tag{7}$$

where θ_0 and θ_1 represent two different parameter points, is estimated directly.

The estimation process can be improved by directly taking into account information from the MC simulator. This can be understood if looking at the joint likelihood ratio,

$$\begin{aligned} r(x, z|\theta_0, \theta_1) &\equiv \frac{p(x, z|\theta_0)}{p(x, z|\theta_1)} = \\ &= \frac{p(x|z_d)p(z_d|z_s)p(z_s|z_p)p(z_p|\theta_0)}{p(x|z_d)p(z_d|z_s)p(z_s|z_p)p(z_p|\theta_1)} = \\ &= \frac{p(z_p|\theta_0)}{p(z_p|\theta_1)} = \\ &= \frac{d\sigma(z_p|\theta_0)}{d\sigma(z_p|\theta_1)} \frac{\sigma(\theta_1)}{\sigma(\theta_0)}, \end{aligned} \tag{8}$$

where $d\sigma(z_p|\theta)$ are the parton-level event weights which can be calculated via

$$d\sigma(z_p|\theta) = \frac{(2\pi)^4 f_1(x_1, Q^2) f_2(x_2, Q^2)}{8 x_1 x_2 s} |\mathcal{M}|^2(z_p|\theta) d\Phi(z_p). \tag{9}$$

Here, f_1 and f_2 are the parton-distribution functions (PDF) depending on the momentum fractions $x_{1,2}$ and the momentum transfer Q. s is the square of the center-of-mass energy; $\mathcal{M}$, the matrix element of the considered process; and Φ, the corresponding phase space. The parton-shower and detector effects cancel in Eq. (8). Since we can calculate the parton-level event weights reasonably fast (using the "morphing" reweighting technique [55, 73, 76]), the joint likelihood ratio is accessible.

Analogously, one can obtain the joint score,

$$\begin{aligned} t(x, z|\theta) &\equiv \nabla_\theta \log p(x, z|\theta) = \\ &= \frac{p(x|z_d)p(z_d|z_s)p(z_s|z_p)\nabla_\theta p(z_p|\theta)}{p(x|z_d)p(z_d|z_s)p(z_s|z_p)p(z_p|\theta)} = \\ &= \frac{\nabla_\theta d\sigma(z_p|\theta)}{d\sigma(z_p|\theta)} - \frac{\nabla_\theta \sigma(\theta)}{\sigma(\theta)}, \end{aligned} \tag{10}$$

which can be calculated by evaluating the derivatives of the parton-level event weights and the total cross section.

As shown in Refs. [53, 55], the joint likelihood ratio (and the joint score) can be used to define suitable loss functions whose minimizing function is the true likelihood ratio $r(x|\theta_0, \theta_1)$ (or the true score $t(x|\theta) \equiv \nabla_\theta \log p(x|\theta)$). A simple example for such a loss function is

$$L[\hat{r}(x|\theta_0, \theta_1)] = \frac{1}{N} \sum_{(x_i, z_i) \sim p(x,z|\theta_1)} |r(x_i, z_i|\theta_0, \theta_1) - \hat{r}(x_i|\theta_0, \theta_1)|^2, \tag{11}$$

where $\hat{r}(x|\theta_0,\theta_1)$ is the estimator for $r(x|\theta_0,\theta_1)$. The sum runs of events sampled according to the probability distribution $p(x,z|\theta_1)$ (using MC simulators). An analogous loss function can be defined for the score.

The minimization procedure is perfectly suited to be tackled by machine learning. The estimator function can be expressed in terms of a neural network (NN). The NN can then be trained with the loss function given in Eq. (11) using standard techniques. In order to reduce the uncertainty associated with the training of the NN, an ensemble of NNs with different random seeds can be trained [77]. For the results presented in Section 5, we found this averaging to be essential in order to smoothen out fluctuations of the individual networks.

Note that in practice more involved loss functions, featuring faster and more stable convergence, are used. For our numerical study, we will employ the ALICES [56] and SALLY [53] methods.

The ALICES loss functional is given by

$$\begin{aligned}\mathcal{L}_{\text{ALICES}}[\hat{s}(x|\theta_0,\theta_1)] &= \\ = -\frac{1}{N}\sum_{(x_i,z_i)\sim p(x_i,z_i)} \Bigg[& s(x_i,z_i|\theta_0,\theta_1)\log\hat{s}(x_i) + (1-s(x_i,z_i|\theta_0,\theta_1))\log(1-\hat{s}(x_i)) \\ & + \alpha(1-y_i)\left| t(x_i,z_i|\theta_0,\theta_1) - \nabla_\theta \log\left(\frac{1-\hat{s}(x_i|\theta,\theta_1)}{\hat{s}(x_i|\theta,\theta_1)}\right)\Big|_{\theta_0}\right|^2 \Bigg], \end{aligned} \tag{12}$$

where the function s is defined via

$$\hat{r}(x|\theta_0,\theta_1) = \frac{1-\hat{s}(x|\theta_0,\theta_1)}{\hat{s}(x|\theta_0,\theta_1)}. \tag{13}$$

The first term in the square bracket parametrizes the deviation of the joint likelihood score $\hat{r}$ from the true likelihood ratio r. The second term in the square bracket — proportional to the hyperparameter α — includes an additional mean-squared error loss for the joint score.

The simpler SALLY loss functional — with the goal to estimate the score — reads

$$\mathcal{L}_{\text{SALLY}}[\hat{s}(x|\theta_0,\theta_1)] = -\frac{1}{N}\sum_{(x_i,z_i)\sim p(x_i,z_i)} \left| t(x,z|\theta_0) - \hat{t}(x|\theta_0)\right|^2 \tag{14}$$

and basically corresponds to the second line of Eq. (12).

3.2 Fisher information

While the likelihood is crucial to derive exclusion limits, it is not the most useful quantity to quickly assess the sensitivity of a certain measurement. For the purpose of optimizing an analysis or for comparing the sensitivity of different processes, often the Fisher information matrix is more useful [78, 79]. It is defined via

$$I_{ij}(\theta) = \mathbb{E}\left[\frac{\partial \log p_{\text{full}}(\{x\}|\theta)}{\partial\theta_i}\frac{\partial \log p_{\text{full}}(\{x\}|\theta)}{\partial\theta_j}\bigg|_\theta\right], \tag{15}$$

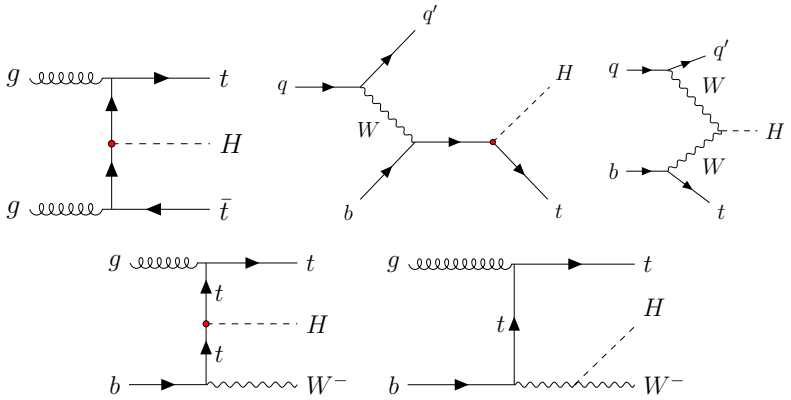

Figure 1: Exemplary Feynman diagrams for the $t\bar{t}H$ (upper left diagram), tH (upper middle and right diagrams), and tWH (lower diagrams) signal processes.

where the symbol "$\mathbb{E}$" is used to denote the expectation value.

According to the Cramér-Rao bound [80, 81], the minimal covariance of an estimator $\hat{\theta}$ is given by the inverse of the Fisher matrix,

$$\mathrm{cov}(\hat{\theta}|\theta)_{ij} \geq I^{-1}_{ij}(\theta), \tag{16}$$

where we assumed that our estimator is unbiased (i.e. that its expectation value $\bar{\theta}$ is equal to the true parameter point θ). $\mathrm{cov}(\hat{\theta}|\theta)_{ij} \equiv \mathbb{E}\left[(\hat{\theta}_i - \bar{\theta})(\hat{\theta}_i - \bar{\theta})|\theta)\right]$ is the covariance matrix. In other words, the size of the Fisher information is directly related to the reachable precision level of the respective parameter. For example, the measurement error for a parameter θ in the one-dimensional case is given by $\Delta\theta \geq 1/\sqrt{I(\theta)}$.

For our purposes, we can calculate the Fisher matrix via

$$I_{ij}(\theta) \simeq \frac{L}{\sigma(\theta)} \frac{\partial\sigma(\theta)}{\partial\theta_i} \frac{\partial\sigma(\theta)}{\partial\theta_j} + \frac{L\sigma(\theta)}{n} \sum_{x\sim p(x|\theta)} t_i(x|\theta) t_j(x|\theta), \tag{17}$$

where the first term encodes the information in the total rate of a process and the second term the information contained in the kinematic distributions.

4 Event simulation and selection

The target of our analysis is top-associated Higgs production consisting out of the sub-channels $t\bar{t}H$, tH, and tWH production. We focus on the $H \to \gamma\gamma$ channel due to its sharp peak structure around the measured Higgs mass in the invariant mass spectrum of the photons. This feature allows for a subtraction of the background event rates originating from non-Higgs processes, whose invariant mass spectrum typically falls smoothly with increasing invariant mass. We assume that the non-Higgs background is already subtracted by a fit to the $m_{\gamma\gamma}$ distribution in the experimental data.

In order to simplify the analysis further, we also require at least one prompt lepton in the final state produced via a Z or W boson decay. Correspondingly, we consider $t\bar{t}H$,

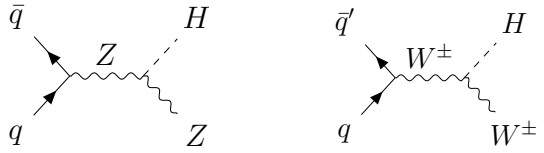

Figure 2: Exemplary Feynman diagrams for the ZH and WH background processes.

tH, and tWH production as signal processes (see Fig. 1); ZH and WH production are considered as background processes (see Fig. 2).[1]

We use `MadGraph5_aMC@NLO 2.8.2` [82] to generate MC event samples for these processes with `Pythia 8.244` [83] as parton shower (employing the A14 set of tuned parameters [84]). As parton-distribution function (PDF), we use the `MSTW2008LO` [85] PDF set assessed via the `LHAPDF 6.2.3` interface [86]. The detector response is simulated using `Delphes 3.4.2` [87] employing either the ATLAS-LHC or the HL-LHC [88, 89] configuration card provided by `Delphes`.[2] We assume the center-of-mass energy to be $\sqrt{s} = 13$ TeV for both the LHC and the HL-LHC.[3] We generate events for these processes at the leading-order (LO) level employing the "Higgs characterization model" [30, 74, 75]. The overall event rates for each subprocess are rescaled to state-of-the-art SM predictions [90] using flat K factors.

In order to account for theoretical uncertainties, we vary the renormalization and factorization scale in the interval $[1/2, 2]$ times the central scale, which is the central squared transverse mass after k_T clustering, using the built-in `MadGraph` reweighting functionality.[4] For the estimate of the likelihood, the renormalization scale is treated as an additional free parameter. After deriving the likelihood, we profile over the renormalization scale without preferring specific values. From a conceptional point of view, also the PDF uncertainty can be taken into account in exactly the same way. With the current version of `MadMiner` (version `0.9`), we, however, found the profiling over the additional PDF nuisance parameters to significantly increase the computational cost to evaluate the final profiled likelihood. Therefore, we do not include this source of uncertainty in the present study. We expect that the run time can be reduced by further optimizations of `MadMiner`.

Using the setup described above, we generate MC event samples for 10 different benchmark points, which are distributed throughout our three-dimensional parameter space. One of these benchmark points represents the SM ($c_V = c_t = 1$ and $\tilde{c}_t = 0$). For the SM point, we generated 5×10^5 events for each of the five different subprocesses

[1] We neglect the s-channel contribution to tH production as well as the gluon-induced ZH production since their cross sections are small in comparison to the other involved processes.

[2] In the ATLAS-LHC card, we have modified the default setting for the radius parameter used by the anti-k_t jet clustering algorithm to $R = 0.4$ instead of $R = 0.6$. This choice is more often used in ATLAS analyses and also corresponds to the value used in the HL-LHC card.

[3] The possible increase in the center-of-mass energy to $\sqrt{s} = 14$ TeV at the HL-LHC [88, 89] would lead to an increase of the cross section for the top-associated Higgs production by about 20%.

[4] For top-associated Higgs production, the variation of the renormalization scale at LO results in a shift of the total cross section similar in size as the shift between the LO and NLO cross sections (see e.g. Refs. [30–32]).

observable	condition
N_γ	≥ 2 (with $\lvert\eta\rvert < 2.5$ and $p_T > 25$ GeV)
$(p^\gamma_{T,1}, p^\gamma_{T,2})$	$\geq (35, 25)$ GeV
$m_{\gamma\gamma}$	$[105 - 160]$ GeV
$(p^\gamma_{T,1}/m_{\gamma\gamma}, p^\gamma_{T,2}/m_{\gamma\gamma})$	$\geq (0.35, 0.25)$
N_ℓ	≥ 1 (with $\lvert\eta\rvert < 2.5$ and $p_T > 15$ GeV)
$m_{\ell\ell}$	$[80, 100]$ GeV vetoed if same flavour
N_{jet}	≥ 1 (with $\lvert\eta\rvert < 2.5$ and $p_T > 25$ GeV)

Table 1: Summary of preselection cuts.

($t\bar{t}H$, tH, tWH, ZH and WH); for the other benchmark points, we generated 1×10^5 events per subprocess. Event weights for other parameter points, which are not identical to one of the benchmark points, are calculated using the morphing technique described in Refs. [55, 73, 76]. As input for the neural network, a set of events is drawn from MC events with probabilities given by the respective event weights (a given MC event can appear multiple times). In a second step, these unweighted events are augmented by calculating the joint likelihood ratio $r(x, z)$ and the joint score $t(x, z)$ (see Section 3).

Following closely the experimental analyses, we impose preselection cuts on our event samples, which correspond to a simplified version of the typical preselection cuts as used by ATLAS and CMS for $t\bar{t}H, H \to \gamma\gamma$ measurements [24, 25]. They are summarized in Table 1. A very similar preselection was used in Ref. [46].

We require all events to feature at least two photons with a pseudo-rapidity $|\eta| < 2.5$ and $p_T > 25$ GeV fulfilling the default isolation criteria of the ATLAS and HL-LHC `Delphes` cards. The two highest-p_T photons form the Higgs boson candidate. The highest-p_T photon is required to have a transverse momentum, $p^\gamma_{T,1}$, larger than 35 GeV; the second-highest-p_T photon, to have a transverse momentum, $p^\gamma_{T,2}$, larger than 25 GeV. The invariant mass of the di-photon system must be close to the Higgs boson mass (i.e. within in the range of $[105 - 160]$ GeV). Moreover, we impose that $p^\gamma_{T,1}/m_{\gamma\gamma} \geq 0.35$ and $p^\gamma_{T,2}/m_{\gamma\gamma} \geq 0.35$ in order to suppress photon radiation from initial-state partons.

As mentioned above, we also require at least one lepton with $p_{T,l} > 15$ GeV and $|\eta_l| < 2.5$ in the final state. For an event containing two leptons, the invariant mass of leptons must not lie within in the range $[80, 100]$ GeV if the leptons have opposite charge and the same flavour in order to suppress Z-boson induced background (i.e. from ZH production). Besides, we also require at least one jet with a $p_T > 20$ GeV and $|\eta_j| < 2.5$ in the final state.

After event generation and imposing the cuts listed above, a set of observables is calculated for each remaining event. This observable set is then used as input for the NN (see Section 3). The observable set includes the number of leptons (n_l), the number of photons (n_γ), the number of jets (n_j), the number of b jets (n_b), the missing transverse energy ($E_{T,\text{miss}}$), the azimuth angle of the missing transverse energy vector ($\phi_{E_{T,\text{miss}}}$),

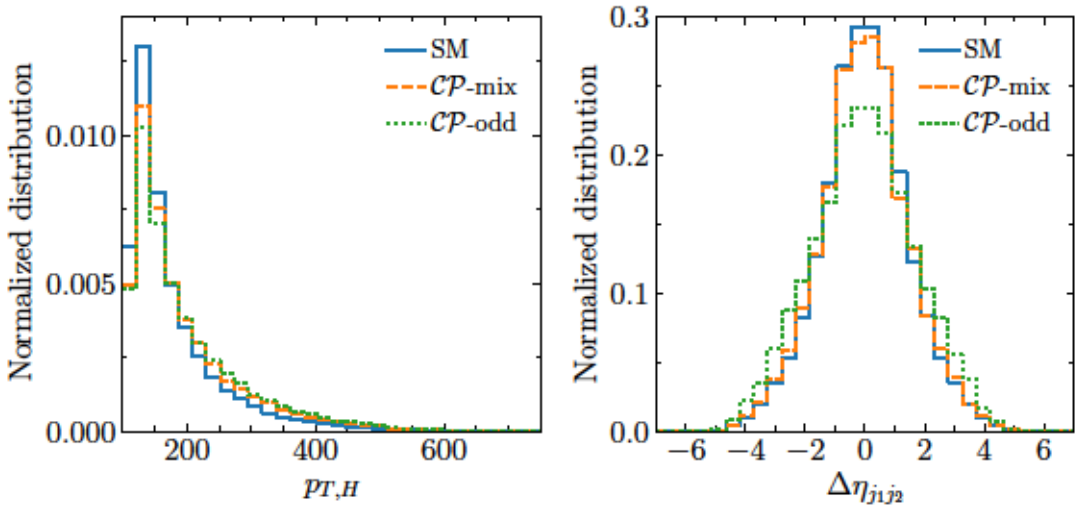

Figure 3: Normalized distribution of $p_{T,H}$ and $\Delta\eta_{j_1j_2}$ for three different parameter points: SM with $c_t = 1$, $\tilde{c}_t = 0$ and $c_V = 1$ (blue solid), $\mathcal{CP}$-mix with $c_t = \tilde{c}_t = 1.2$ and $c_V = 1$ (orange dashed), as well as $\mathcal{CP}$-odd with $c_t = 0, \tilde{c}_t = c_V = 1$ (green dotted).

the visible energy (E_{vis}), the pseudorapidity of the visible energy vector ($\eta_{E_{\text{vis}}}$), the transverse momenta of the two leading jets (p_{T,j_1} and p_{T,j_2}), the azimuth angles of the two leading jets (ϕ_{j_1} and ϕ_{j_2}), the pseudorapidities of the two leading jets (η_{j_1} and η_{j_2}), the transverse momenta of the two leading photons (p_{T,γ_1} and p_{T,γ_2}), the azimuth angles of the two leading photons (ϕ_{γ_1} and ϕ_{γ_2}), the pseudorapidities of the two leading photons (η_{γ_1} and η_{γ_2}), the transverse momenta of the two leading leptons (p_{T,ℓ_1} and p_{T,ℓ_2}), the azimuth angles of the two leading leptons (ϕ_{ℓ_1} and ϕ_{ℓ_2}), the pseudorapidities of the two leading leptons (η_{ℓ_1} and η_{ℓ_2}), and the charges of the two leading leptons (Q_{ℓ_1} and Q_{ℓ_2}). In addition to these low-level inputs, we also include several high-level observables which are computed out of the low-level observables: the rapidity difference of the two leading jets ($\Delta\eta_{j_1j_2}$), the azimuth angle difference of the two leading jets ($\Delta\phi_{j_1j_2}$), the invariant mass of the two leading jets (m_{jj}), the rapidity difference of the leading b jet and the leading non-b jet (y_{bj}), the invariant mass of the two leading photons ($m_{\gamma\gamma}$), the transverse momenta of the two leading photon system ($p_{T,H}$), the invariant mass of the two leading leptons ($m_{\ell\ell}$), and the transverse mass of leading b jet and leading lepton system (m_T^{top}). We then use these 36 observables as input for the NN.

Some of these observables are only very weakly sensitive to the $\mathcal{CP}$ character of the top-Yukawa coupling. We, nevertheless, take these observables into account in order to provide the NN with as much information as possible. The invariant masses of the two-leading photons, $m_{\gamma\gamma}$, as well as of the two leading leptons, $m_{\ell\ell}$ are included in order to impose the cuts summarized in Table 1.

As examples for observables strongly sensitive to the $\mathcal{CP}$ nature of the top-Yukawa coupling (see also e.g. Ref. [30]), we show in Fig. 3 the normalized distributions of the Higgs transverse momentum, $p_{T,H}$, and the pseudorapidity difference of the two leading

jets, $\Delta\eta_{j_1 j2}$ (events with only one jet are not shown in the histogram). The distributions are shown for three different parameter points: SM with $c_t = 1$, $\tilde{c}_t = 0$ and $c_V = 1$ (blue solid), $\mathcal{CP}$-mix with $c_t = \tilde{c}_t = 1.2$ and $c_V = 1$ (orange dashed), as well as $\mathcal{CP}$-odd with $c_t = 0, \tilde{c}_t = c_V = 1$ (green dotted). For the left panel of Fig. 3, showing the distribution of $p_{T,H}$, we observe that the Higgs transverse momentum is on average harder in presence of a non-zero $\mathcal{CP}$-odd top-Yukawa coupling. For the right panel of Fig. 3, showing the distribution of $\Delta\eta_{j_1 j_2}$, we notice that for a $\mathcal{CP}$-odd top-Yukawa coupling the two leading jets tend to have a larger angular separation. We perform a more detailed analysis regarding the sensitivity of the different observables to the $\mathcal{CP}$ character of the top-Yukawa coupling in Section 5.3.

5 Results

In this Section, we present the tentative constraints on c_t and $\tilde{c}_t$ using the methodology described in Section 3.

The limits shown in Figs. 4 to 6 are derived using the ALICES method as described in Section 3.1. For the LHC results, the likelihood is constructed using four different NNs, with two hidden layers, trained with a sample of 10^6 unweighted events, over which we average by forming an ensemble [55]; for the HL-LHC results, which cover a smaller parameter region, we use samples of 10^5 unweighted events and average over six different neural networks.

5.1 Expected limits at the LHC and the HL-LHC

First, we look at the limits on c_t and $\tilde{c}_t$ expected for SM data at the LHC using 139 fb^{-1} (corresponding roughly to the currently recorded event data) and 300 fb^{-1} (corresponding to event data recorded after LHC Run 3) as well the limits expected for SM data at the HL-LHC using 3000 fb^{-1}.

The resulting limits in the $(c_t, \tilde{c}_t)$ plane are shown in the upper left, the upper right, and the lower left panels of Fig. 4. The colour coding indicates the p-value associated with each parameter point. The white and black dashed contours define the 68.3%, 95.4%, and 99.7% confidence level (CL) limits, respectively. For these plots, we project the three-dimensional likelihood to the $(c_t, \tilde{c}_t)$ plane by setting $c_V = 1$ and we do not perform any variation of the renormalization scale here.

In the upper left plot of Fig. 4, the constraints are shown for a luminosity of 139 fb^{-1}. With this amount of data, the negative c_t range cannot be completely excluded at the 99.7% CL. The form of the 99.7% CL region is a consequence of the dependence of the top-associated Higgs production cross section on c_t and $\tilde{c}_t$ (see e.g. Ref. [46]). With 139 fb^{-1}, $\tilde{c}_t$ is constrained to the interval $\sim [-0.8, 0.8]$ at the 68.3% CL level. Increasing the luminosity to 300 fb^{-1} (see upper right plot of Fig. 4) tightens the bounds on $\tilde{c}_t$ to $\sim [-0.5, 0.5]$ and excludes the negative c_t region at the 99.7% CL level. Even stronger constraints are possible at the HL-LHC (see bottom left plot of Fig. 4; note the reduced parameter range) tentatively constraining $\tilde{c}_t$ to the interval $\sim [-0.25, 0.25]$. At

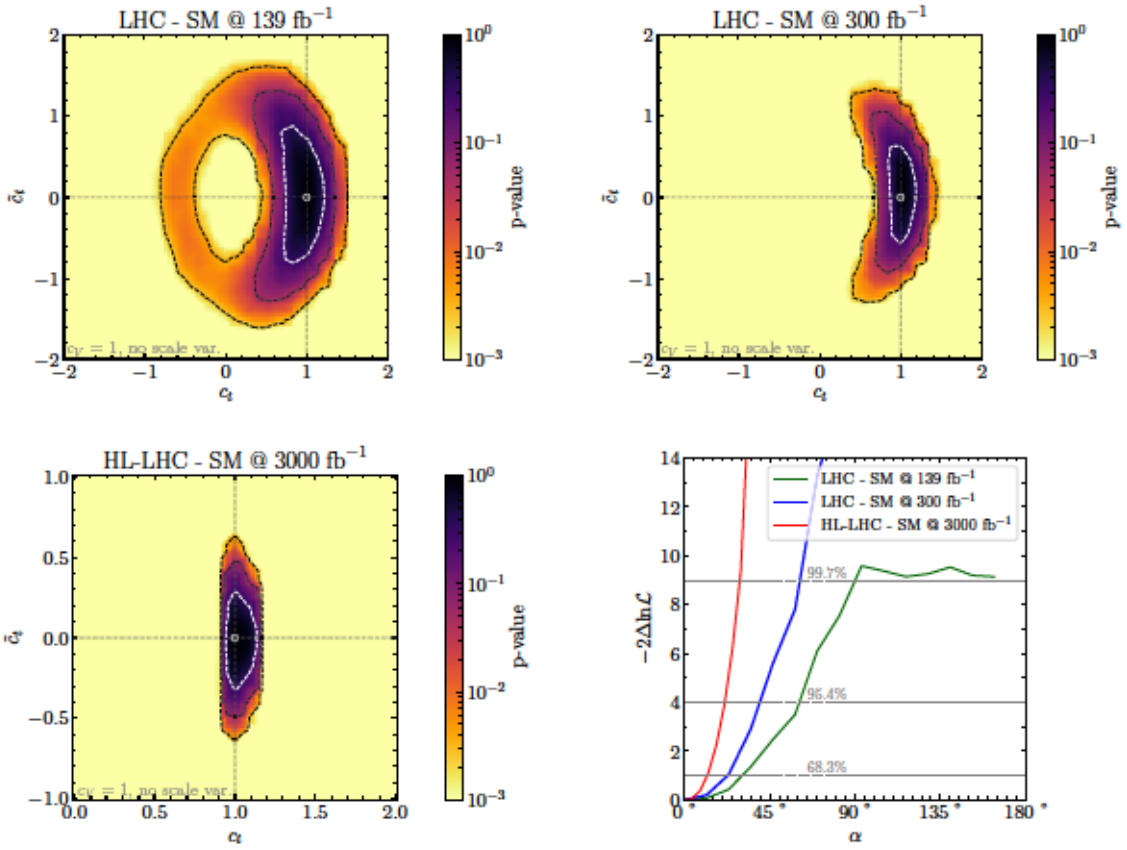

Figure 4: *Upper left:* Expected limits (assuming SM data) on c_t and $\tilde{c}_t$ using a luminosity of 139 fb^{-1}. We assume that $c_V = 1$ and do not vary the renormalization scale. *Upper right:* same as upper left panel but for a luminosity of 300 fb^{-1}. *Bottom left:* Same as upper left panel, but the constraints are evaluated for the HL-LHC with a luminosity of 3000 fb^{-1}. *Bottom right:* Constraints on the $\mathcal{CP}$-violating phase α for a luminosity of 139 fb^{-1} (green), 300 fb^{-1} (blue), and 3000 fb^{-1} (red) after profiling of $|g_t|$.

the HL-LHC, the analysis does not only profit from the increased luminosity but also from the improved detector coverage in the forward region. While the bound on $\tilde{c}_t$ for $\mathcal{L} = 300$ fb^{-1} corresponds roughly to a rescaling of the bound for $\mathcal{L} = 139$ fb^{-1} by the increased luminosity, the improvement at the HL-LHC is weaker than expected by a simple luminosity rescaling. This is a consequence of the small dependence of the total rate on $\tilde{c}_t$ close to the SM point. In this situation, the constraints on $\tilde{c}_t$ rely increasingly on the information encoded in the kinematic distributions.

These results can also be presented in terms of a $\mathcal{CP}$-violating phase α and an absolute value $|g_t|$ of the top-Yukawa coupling (see Eq. (2)). Profiling over $|g_t|$, we show the one-dimensional profiles for α in the lower right panel of Fig. 4 for the LHC using 139 fb^{-1} (green line), the LHC using 300 fb^{-1} (blue line), and the HL-LHC using 3000 fb^{-1} (red line).[5] With a luminosity of 139 fb^{-1}, α is hardly constrained α at the 99.7% confidence level. At 95.4% confidence level, α is constrained to be below $\sim 60°$, $\sim 40°$, and $\sim 22°$ using 139 fb^{-1}, 300 fb^{-1}, and 3000 fb^{-1} respectively.

The expected bound on α using 139 fb^{-1} are close to the precision level reached in Refs. [23–26], which also focused on top-associated Higgs production. The experimental analyses, however, also consider events which do not contain a lepton.[6] Note also that for a luminosity of 139 fb^{-1}, the constraints on $\tilde{c}_t$ are mainly due to measurements of the total rate (as indicated by the ellipse-shaped 99.7% CL allowed region in the upper left plot of Fig. 4). With an increased luminosity, kinematic constraints become increasingly important. Correspondingly, we expect machine-learning-based inference to demonstrate its advantages more clearly if the luminosity is increased.

So far, we have assumed that the Higgs coupling to massive vector bosons is SM-like (i.e., $c_V = 1$). In order to assess the impact of c_V on the $(c_t, \tilde{c}_t)$ constraints, we now include c_V fully in our likelihood estimation (floating it freely in the interval $[0.5, 1.5]$). The constraints on c_t and $\tilde{c}_t$ (using a luminosity of 300 fb^{-1}) after profiling over c_V are shown in the left panel of Fig. 5. In comparison to the upper right panel of Fig. 4, for which $c_V = 1$ is fixed, the constraints in the $(c_t, \tilde{c}_t)$ plane are only slightly weaker (i.e., the 95.4% C.L. limit on α is weakened to $\sim 45°$). This indicates that even without a precise knowledge of the Higgs coupling to massive vector boson, top-associated Higgs production allows for a precise determination of the top-Yukawa coupling.

An additional source of uncertainty not considered so far are uncertainties of theoretical nature. In order to estimate them, we consider a variation of the renormalization (and factorization) scale entering the MC event generation. The resulting constraints on c_t and $\tilde{c}_t$ (using a luminosity of 300 fb^{-1}) after profiling over the scale nuisance parameter (and c_V) are shown in the right panel of Fig. 5. Also here, the constraints are only slightly weaker than those presented in the upper right panel of Fig. 4 and the left panel

[5] Here, we directly show the negative likelihood. In the other Figures, we display the p-value obtained using Wilks' theorem.

[6] Note moreover that the effective model used in Ref. [25] is not directly comparable to the model used in this work (see Section 2), since MC events have been generated at next-to-leading order (NLO) implying an interdependence of Higgs production via gluon fusion and top-associated Higgs production. Moreover, the results of Refs. [23, 26] are not directly comparable, since two Higgs decay channels ($H \to \gamma\gamma$ and $H \to 4\ell$) have been combined.

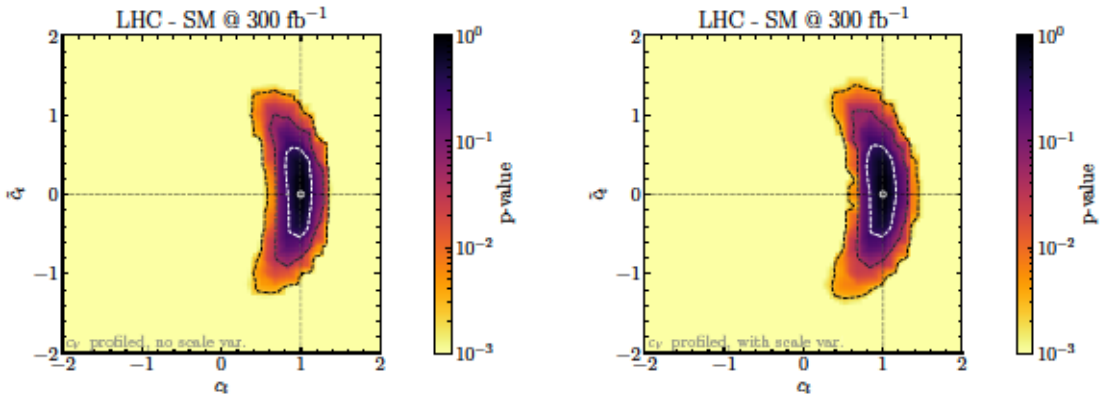

Figure 5: *Left:* Expected limits on c_t and $\tilde{c}_t$ with a luminosity of 300 fb^{-1} for a freely floating c_V. *Right:* Same as left, but also the renormalization scale is floated.

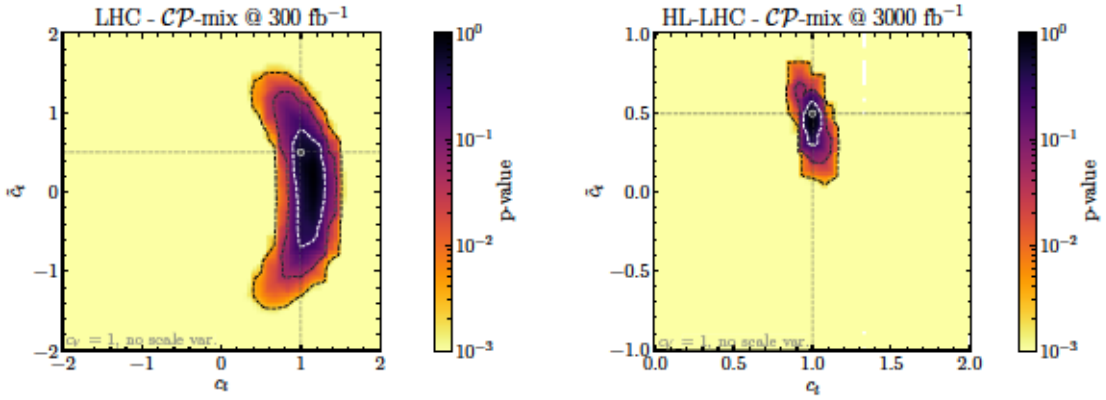

Figure 6: Observed limit on c_t and $\tilde{c}_t$ for a $\mathcal{CP}$-mixed Higgs boson (with $c_t = 1$ and $\tilde{c}_t = 0.5$) using a luminosity of 300 fb^{-1} (left) and 3000 fb^{-1} (right).

of Fig. 5 (finding a 95.4% CL limit of $\sim 45°$ on α). This result indicates that our result is relatively robust against theoretical uncertainties even though a more detailed study taking into account e.g. PDF uncertainties as well as next-to-leading order effects, which is beyond the scope of the present paper, would be necessary to answer this question more quantitatively.

5.2 Observed limits in case of a deviation from the SM

All results presented above show expected limits assuming SM data. It is, however, also relevant to ask how well c_t and $\tilde{c}_t$ can be constrained if the data is not SM like.

We address this question in Fig. 6 where we assume the observed data to follow the predictions for a $\mathcal{CP}$-admixed Higgs boson. For this $\mathcal{CP}$-admixed Higgs boson, we fix $c_t = 1$ and $\tilde{c}_t = 0.5$ as an example. Moreover, we assume that $c_V = 1$.

In the left panel of Fig. 6, showing the exclusion boundaries for a luminosity of 300 fb^{-1}

collected at the LHC, we observe that the SM hypothesis cannot be excluded even at the 68.3% CL level. In comparison to the expected exclusion boundaries in case of SM data (see Fig. 4), the 68.3% CL allowed $\tilde{c}_t$ interval weakens by $\sim 50\%$.

The situation is completely different for a luminosity of 3000 fb^{-1} collected at HL-LHC (see right panel of Fig. 6). Here, the SM point is excluded at the 99.7% CL level. The HL-LHC precision level allows to quite precisely pinpoint c_t (to $\sim [0.9, 1.1]$) and $\tilde{c}_t$ (to $\sim [0.3, 0, 6]$) at the 68.3% CL level.

5.3 Most sensitive observables

After discussing the expected constraints on c_t and $\tilde{c}_t$, it is a relevant question which observables are most important for setting these limits. We approach this question by evaluating the Fisher information for different sets of input observables for a luminosity of 300 fb^{-1}.

As a first step, we evaluate the Fisher information matrix for the SM parameter point ($c_V = c_t = 1$, $\tilde{c}_t = 0$) taking into account the full kinematic and cross-section information (labelled by "full"),[7]

$$I_{ij}^{\text{full}}(\text{SM}) \simeq \begin{pmatrix} 91.4 & 13.7 & 0.1 \\ 13.7 & 108.2 & -0.1 \\ 0.1 & -0.1 & 0.004 \end{pmatrix}, \tag{18}$$

where we span our parameter space by the vector $(c_V, c_t, \tilde{c}_t)^T$. As eigenvalues of $I_{ij}^{\text{full}}(\text{SM})$, we obtain 115.9, 83.7, and 0.001 with the respective normalized eigenvectors

$$v_1 \simeq \begin{pmatrix} 0.49 \\ 0.87 \\ -0.001 \end{pmatrix}, \quad v_2 \simeq \begin{pmatrix} 0.87 \\ -0.49 \\ 0.001 \end{pmatrix}, \quad v_3 \simeq \begin{pmatrix} -0.0001 \\ 0.0013 \\ 1.00 \end{pmatrix}. \tag{19}$$

As explained in Section 3.2, the Fisher information matrix is directly related to the reachable precision level for the various theory parameters: the higher the information, the more precise the corresponding coupling can be measured. Correspondingly, Eqs. (18) and (19) imply that close to the SM point c_t can be measured most precisely. There, however, is a strong correlation with c_V (as indicated by the comparably large off-diagonal entries), which can be constrained a little less. In contrast, the sensitivity to $\tilde{c}_t$ is much lower. This finding is in agreement with the shape of the exclusion boundaries found in Fig. 4 which are almost flat in the $\tilde{c}_t$ direction close to the SM point.

As a second step, we recompute the Fisher information based on a selection of one- and two-dimensional kinematic distributions in order to understand which observables contain the most information about the various parameters.[8] Moreover, we evaluate

[7] For evaluating the kinematic information, we rely on an ensemble of three `SALLY` estimators trained with a sample of 10^5 unweighted events at the SM and the $\mathcal{CP}$-mixed benchmark points.

[8] We also evaluated the information contained in all other evaluated observables (see Section 4) and their pair-wise combinations. In Figs. 7 and 8, we only display those observables with the highest amount of information or those of special experimental interest. The one-dimensional distributions are evaluated by filling a histogram with 20 bins; the two-dimensional distributions, by filling a two-dimensional histogram with 10×10 bins.

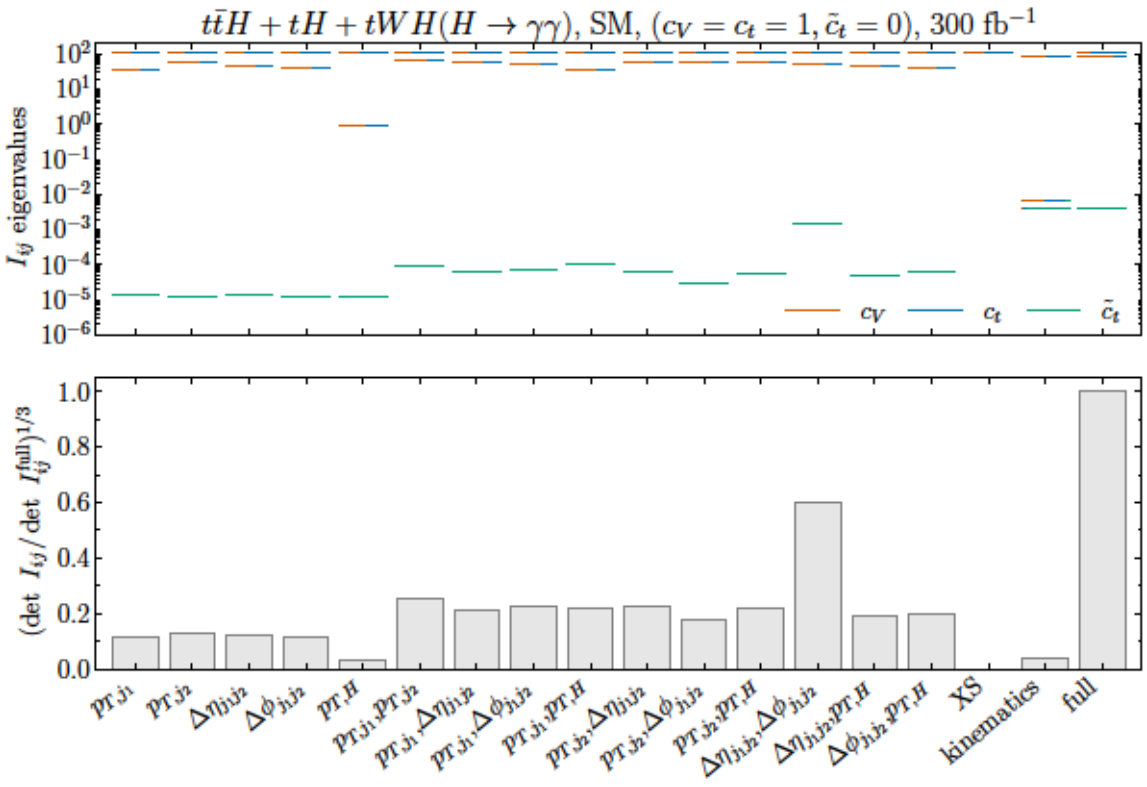

Figure 7: Fisher information for top-associated Higgs production with $H \to \gamma\gamma$ (for a luminosity of 300 fb^{-1}). The full information is compared to the information contained in several one- and two-dimensional distributions. The Fisher information is evaluated for the SM benchmark point ($c_V = c_t = 1$, $\tilde{c}_t = 0$).

the Fisher information using only the kinematic information (labelled by "kinematics") and using only the cross-section information (labelled by "XS"). Note that the one- and two-dimensional kinematic distributions always implicitly contain the cross section information.

In Fig. 7, we show the results of the different evaluations of the Fisher information at the SM point. The panel is divided in two sub-panels. In the upper sub-panels, the eigenvalues of the Fisher matrix are shown. As explained for Eqs. (18) and (19), the size and decomposition — as indicated by the colour coding (orange for c_V, blue for c_t, green for $\tilde{c}_t$) — of these eigenvalues is a measure which combinations of parameters can be constrained most strongly to a certain precision. In the lower sub-panels, the determinant of the Fisher matrix, which can be used as a measure of the overall reachable precision level, is shown. We normalize it to the determinant of the Fisher matrix calculated taking into account the full kinematic and cross-section information and take the third root to obtain a measure for the average precision level reachable for each of three parameters.

As already discussed above, the highest precision level is reachable for c_t. The precision level reachable for c_V is slightly smaller. As visible by comparing the different ways of

evaluating the Fisher information, the information about c_t and c_V is to a large extent contained in the total rate. Taking into account only kinematic information results in significantly weaker constraints. Note, however, that the cross section alone is only sensitive to a specific combination of c_t and c_V, since the cross section is only a single number. Consequently, the determinant of the Fisher matrix calculated based only on the cross section is zero. If, however, the kinematic distribution of at least one observable is taken into account in addition to the total rate, the constraints on c_t and c_V are almost as strong as if the full information is considered. The situation is, however, quite different for constraints on $\tilde{c}_t$, which are in general much weaker than the constraints on c_t and c_V. The information on $\tilde{c}_t$ is almost exclusively contained in the kinematic distributions as evident when comparing the eigenvalues of the Fisher matrix calculated based on the full and on only the kinematic information. This kinematic information on $\tilde{c}_t$ can also hardly be captured by one- or two-dimensional kinematic distributions. Only the two-dimensional histogram of the pseudo-rapidity and angular difference of the leading and the subleading jet is able to capture a significant part of the information on $\tilde{c}_t$. In summary, it seems to be necessary to fully exploit all available information to constrain $\tilde{c}_t$ at the SM point.

The situation is different if the Fisher information is evaluated at a $\mathcal{CP}$-mixed benchmark point ($c_V = c_t = 1$, $\tilde{c}_t = 0.5$, as for the $\mathcal{CP}$-admixed Higgs boson in Section 5.2). Using the full kinematic and cross section information, we obtain for the Fisher information matrix

$$I_{ij}^{\text{full}}(\mathcal{CP}\text{-mixed}) \simeq \begin{pmatrix} 91.3 & 10.6 & 6.4 \\ 10.6 & 98.2 & 26.3 \\ 6.4 & 26.3 & 11.7 \end{pmatrix} \tag{20}$$

with the eigenvalues 112.4, 84.6, and 4.2 as well as the corresponding eigenvectors

$$v_1 \simeq \begin{pmatrix} 0.50 \\ 0.87 \\ -0.04 \end{pmatrix}, \quad v_2 \simeq \begin{pmatrix} 0.83 \\ -0.49 \\ -0.27 \end{pmatrix}, \quad v_3 \simeq \begin{pmatrix} 0.25 \\ -0.10 \\ 0.96 \end{pmatrix}. \tag{21}$$

While c_t can still be constrained most precisely (and still is strongly correlated with c_V), $\tilde{c}_t$ can be constrained more precisely in comparison to the SM point (see Eq. (18)) and is now weakly correlated with both c_t (and too lesser extent with c_V).

The information obtained at the $\mathcal{CP}$-mixed benchmark point using the various kinematic observables (and their combination) is visualized in Fig. 8. As visible in the upper subpanel, the information on $\tilde{c}_t$ is to a large extent contained in the kinematic information as for the SM point. In contrast to the SM point, however, the one- and two-dimensional kinematic distributions are able to capture this information. For the $\mathcal{CP}$-mixed benchmark point, the one-dimensional transverse momentum distribution of the Higgs boson is able to capture the information on $\tilde{c}_t$ almost completely. This is especially interesting in the light of recent simplified-template cross section (STXS) measurements of the Higgs transverse momentum for top-associated Higgs production [91].[9]

[9] In Ref. [91], $t\bar{t}H$ and tH events are, however, disentangled based upon the assumption of SM-like

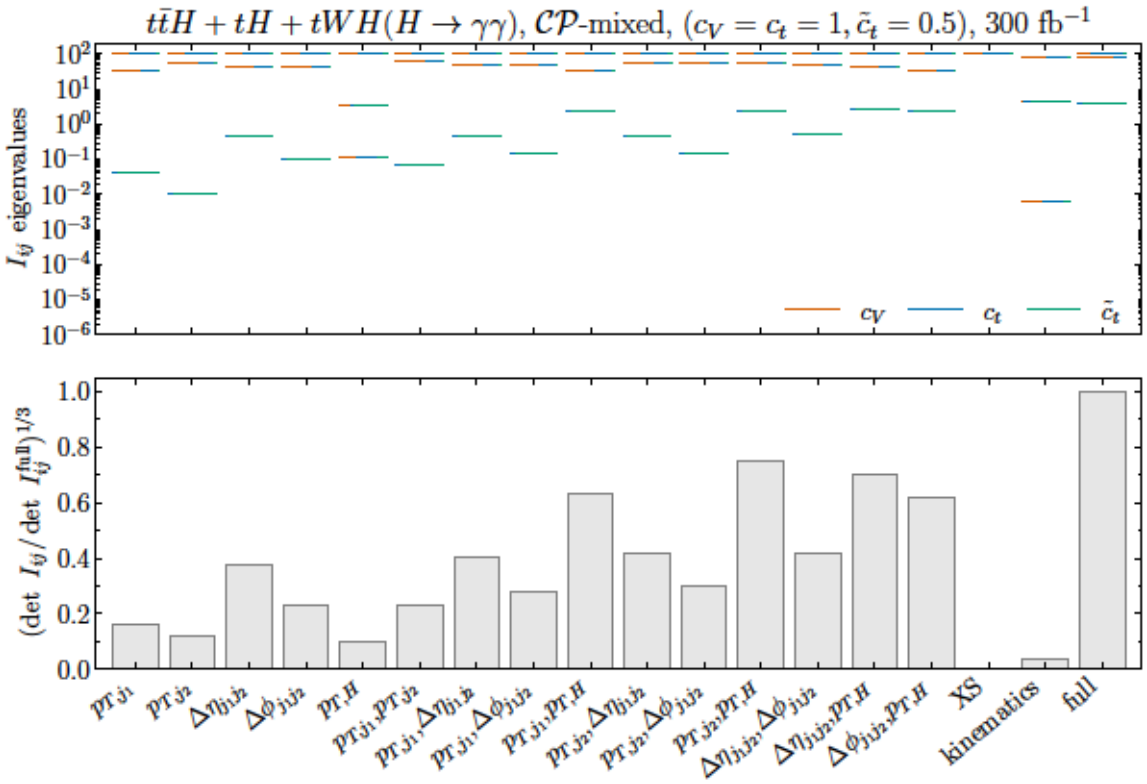

Figure 8: Same as Fig. 7 but the Fisher information is evaluated for a $\mathcal{CP}$-mixed benchmark point $(c_V = c_t = 1,\ \tilde{c}_t = 0.5)$.

6 Conclusions

A detailed investigation of the top-Yukawa coupling is crucial to probe the $\mathcal{CP}$ nature of the Higgs boson discovered at the LHC. In this work, we assessed the potential of machine-learning-based inference to constrain a possible $\mathcal{CP}$-odd component of the Higgs–top-quark interaction.

Machine-learning-based inference allows to approximate the full likelihood fully exploiting the kinematic information contained in the event data. It, moreover, fully takes into account parton shower and detector effects without relying on simplifying assumptions like e.g. the matrix element method.

We performed our analysis in a simplified model framework in which not only the top-Yukawa coupling is allowed to vary freely but also a global rescaling factor is introduced for the Higgs interactions with massive vector bosons.

As physical target process, we focused on top-associated Higgs production with the Higgs decaying to two photons. In order to suppress background, we demanded at least one lepton in the final state. The method can, however, also be used for more inclusive final states.

Applying machine-learning-based inference to top-associated Higgs production — using the tool `Madminer` —, we derived expected bounds on a $\mathcal{CP}$-violating top-Yukawa coupling for the LHC and HL-LHC. Assuming SM data, we found that a $\mathcal{CP}$-odd top Yukawa coupling can be constrained at the 68.3% CL level to lie within $\sim [-0.8, 0.8]$ using a luminosity of 139 fb^{-1} at the LHC; with a luminosity of 300 fb^{-1}, this bound is tightened to $\sim [-0.5, 0.5]$; and, with a luminosity of 3000 fb^{-1} at the HL-LHC, a bound of $\sim [-0.25, 0.25]$ can be obtained. The constraints for a luminosity of 139 fb^{-1} are of a similar precision as current experimental studies, which use a larger data set (i.e. they also include events without a lepton in the final state). Assuming the existence of a small $\mathcal{CP}$-odd Yukawa coupling ($\tilde{c}_t = 0.5$), we found that such a deviation from the SM can not be distinguished from the SM at LHC. To establish a deviation at 99.7% CL level, the full data set to be accumulated at the HL-LHC is needed.

In addition to deriving expected bounds, we also studied the impact of a non-SM-like Higgs–vector-boson coupling on the top-Yukawa coupling constraints. We found deviations of the Higgs–vector-boson couplings from the SM to have negligible impact on the top-Yukawa coupling constraints showing the promising potential of top-associated Higgs production for precision constraints on the top-Yukawa coupling. Moreover, we studied the impact of theoretical uncertainties by treating the renormalization scale as a nuisance parameter. We found also the variation of the renormalization scale to have small effect on our results indicating that our result is robust against theoretical uncertainties.

Aside of direct constraints on the $\mathcal{CP}$ nature of the top-Yukawa coupling, we investigated which kinematic observables are most sensitive to a $\mathcal{CP}$-odd top-Yukawa coupling by extracting the Fisher information. While it is in general very hard to constrain a $\mathcal{CP}$-odd

kinematics. As indicated e.g. in Ref. [46], this assumption does not hold in the case of a non-vanishing $\mathcal{CP}$-odd component of the top-Yukawa coupling.

top-Yukawa coupling if the SM is realized in nature, we found the transverse momentum distribution of the Higgs boson to have a high sensitivity to a $\mathcal{CP}$-odd top-Yukawa coupling if this coupling is non-zero.

Our analysis shows the potential of machine-learning based inference to probe the $\mathcal{CP}$ character of the Higgs–top-quark interaction. We hope that the present study triggers future work aiming a exploiting the full kinematic information of top-quark-associated Higgs production.

Note added

During the final stage of this project, Ref. [92] appeared following a similar approach to constrain $\mathcal{CP}$ violation in the top-Yukawa interaction.

Acknowledgments

We thank Tim Stefaniak for collaboration in an early stage of the project. We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2121 "Quantum Universe" – 390833306. HB acknowledges support by the Alexander von Humboldt foundation.

References

[1] ATLAS, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Phys. Lett. **B716** (2012) 1, 1207.7214.

[2] CMS, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Phys. Lett. **B716** (2012) 30, 1207.7235.

[3] M. Gavela et al., "Standard model CP violation and baryon asymmetry", Mod. Phys. Lett. A **9** (1994) 795, hep-ph/9312215.

[4] P. Huet and E. Sather, "Electroweak baryogenesis and standard model CP violation", Phys. Rev. D **51** (1995) 379, hep-ph/9404302.

[5] CMS, "Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV", (2020), CMS-PAS-HIG-20-006.

[6] CMS, "Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV", (2021), 2110.04836.

[7] ACME, "Improved limit on the electric dipole moment of the electron", Nature **562** (2018) 355.

[8] nEDM, "Measurement of the permanent electric dipole moment of the neutron", Phys. Rev. Lett. **124** (2020) 081803, 2001.11966.

[9] J. Brod, U. Haisch, and J. Zupan, "Constraints on CP-violating Higgs couplings to the third generation", JHEP **11** (2013) 180, 1310.1385.

[10] J. de Vries et al., "Electroweak Baryogenesis and the Standard Model Effective Field Theory", JHEP **01** (2018) 089, 1710.04061.

[11] J. De Vries, M. Postma, and J. van de Vis, "The role of leptons in electroweak baryogenesis", JHEP **04** (2019) 024, 1811.11104.

[12] E. Fuchs et al., "CP violation from τ, t and b dimension-6 Yukawa couplings - interplay of baryogenesis, EDM and Higgs physics", JHEP **05** (2020) 056, 2003.00099.

[13] Y. Chien et al., "Direct and indirect constraints on CP-violating Higgs-quark and Higgs-gluon interactions", JHEP **02** (2016) 011, 1510.00725.

[14] V. Cirigliano et al., "Constraining the top-Higgs sector of the Standard Model Effective Field Theory", Phys. Rev. D **94** (2016) 034031, 1605.04311.

[15] G. Panico, A. Pomarol, and M. Riembau, "EFT approach to the electron Electric Dipole Moment at the two-loop level", JHEP **04** (2019) 090, 1810.09413.

[16] CMS, "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV", Phys. Rev. D **92** (2015) 012004, 1411.3441.

[17] ATLAS, "Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector", Eur. Phys. J. C **75** (2015) 476, 1506.05669.

[18] ATLAS, "Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector", Eur. Phys. J. C **76** (2016) 658, 1602.04516.

[19] CMS, "Constraints on anomalous Higgs boson couplings using production and decay information in the four-lepton final state", Phys. Lett. B **775** (2017) 1, 1707.00541.

[20] CMS, "Constraints on anomalous HVV couplings from the production of Higgs bosons decaying to τ lepton pairs", Phys. Rev. D **100** (2019) 112002, 1903.06973.

[21] CMS, "Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state", Phys. Rev. D **99** (2019) 112003, 1901.00174.

[22] ATLAS, "Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \to \tau\tau$ channel in proton−proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", Phys. Lett. B **805** (2020) 135426, 2002.05315.

[23] CMS, "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in production and decay in the $H \to 4\ell$ channel", (2020), CMS-PAS-HIG-19-009.

[24] CMS, "Measurements of $\mathrm{t\bar{t}}H$ 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** (2020) 061801, 2003.10866.

[25] ATLAS, "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** (2020) 061802, 2004.04545.

[26] CMS, "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state", (2021), 2104.12152.

[27] ATLAS, "Constraints on Higgs boson properties using $WW^*(\to e\nu\mu\nu)jj$ production in 36.1 fb^{-1} of $\sqrt{s}$=13 TeV pp collisions with the ATLAS detector", (2021), 2109.13808.

[28] P. Agrawal, S. Mitra, and A. Shivaji, "Effect of Anomalous Couplings on the Associated Production of a Single Top Quark and a Higgs Boson at the LHC", JHEP **12** (2013) 077, 1211.4362.

[29] J. Ellis et al., "Disentangling Higgs-Top Couplings in Associated Production", JHEP **04** (2014) 004, 1312.5736.

[30] F. Demartin et al., "Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction", Eur. Phys. J. **C74** (2014) 3065, 1407.5089.

[31] F. Demartin et al., "Higgs production in association with a single top quark at the LHC", Eur. Phys. J. **C75** (2015) 267, 1504.00611.

[32] F. Demartin et al., "tWH associated production at the LHC", Eur. Phys. J. **C77** (2017) 34, 1607.05862.

[33] A. Kobakhidze et al., "Implications of CP-violating Top-Higgs Couplings at LHC and Higgs Factories", Phys. Rev. D **95** (2017) 015016, 1610.06676.

[34] Q.-H. Cao et al., "Limiting top quark-Higgs boson interaction and Higgs-boson width from multitop productions", Phys. Rev. **D99** (2019) 113003, 1901.04567.

[35] M. R. Buckley and D. Goncalves, "Boosting the Direct CP Measurement of the Higgs-Top Coupling", Phys. Rev. Lett. **116** (2016) 091801, 1507.07926.

[36] A. V. Gritsan et al., "Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques", Phys. Rev. D **94** (2016) 055023, 1606.03107.

[37] D. Azevedo et al., "CP tests of Higgs couplings in $t\bar{t}h$ semileptonic events at the LHC", Phys. Rev. **D98** (2018) 033004, 1711.05292.

[38] V. Barger, K. Hagiwara, and Y.-J. Zheng, "Probing the Higgs Yukawa coupling to the top quark at the LHC via single top+Higgs production", Phys. Rev. **D99** (2019) 031701, 1807.00281.

[39] 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", JHEP **06** (2018) 079, 1804.05874.

[40] A. Freitas and P. Schwaller, "Higgs CP Properties From Early LHC Data", Phys. Rev. **D87** (2013) 055014, 1211.1980.

[41] A. Djouadi and G. Moreau, "The couplings of the Higgs boson and its CP properties from fits of the signal strengths and their ratios at the 7+8 TeV LHC", Eur. Phys. J. **C73** (2013) 2512, 1303.6591.

[42] F. Boudjema et al., "Lab-frame observables for probing the top-Higgs interaction", Phys. Rev. **D92** (2015) 015019, 1501.03157.

[43] W.-S. Hou, M. Kohda, and T. Modak, "Probing for extra top Yukawa couplings in light of $t\bar{t}h(125)$ observation", Phys. Rev. D **98** (2018) 075007, 1806.06018.

[44] 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** (2020) 135198, 1901.05627.

[45] M. Kraus et al., "Exploring BSM Higgs couplings in single top-quark production", (2019), 1908.09100.

[46] H. Bahl et al., "Indirect $\mathcal{CP}$ probes of the Higgs-top-quark interaction: current LHC constraints and future opportunities", JHEP **11** (2020) 127, 2007.08542.

[47] T. Martini et al., "Probing the CP structure of the top quark Yukawa coupling: Loop sensitivity versus on-shell sensitivity", Phys. Rev. D **104** (2021) 055045, 2104.04277.

[48] N. Mileo et al., "Pseudoscalar top-Higgs coupling: exploration of CP-odd observables to resolve the sign ambiguity", JHEP **07** (2016) 056, 1603.03632.

[49] D. A. Faroughy et al., "Probing the CP nature of the top quark Yukawa at hadron colliders", JHEP **02** (2020) 085, 1909.00007.

[50] B. Bortolato et al., "Optimized probes of CP -odd effects in the $t\bar{t}h$ process at hadron colliders", Nucl. Phys. B **964** (2021) 115328, 2006.13110.

[51] D. Gonçalves et al., "Direct Higgs-top CP-phase measurement with $t\bar{t}h$ at the 14 TeV LHC and 100 TeV FCC", (2021), 2108.01083.

[52] J. Brehmer et al., "Effective LHC measurements with matrix elements and machine learning", J. Phys. Conf. Ser. **1525** (2020) 012022, 1906.01578.

[53] J. Brehmer et al., "Mining gold from implicit models to improve likelihood-free inference", Proc. Nat. Acad. Sci. **117** (2020) 5242, 1805.12244.

[54] J. Brehmer et al., "Constraining Effective Field Theories with Machine Learning", Phys. Rev. Lett. **121** (2018) 111801, 1805.00013.

[55] J. Brehmer et al., "A Guide to Constraining Effective Field Theories with Machine Learning", Phys. Rev. D **98** (2018) 052004, 1805.00020.

[56] M. Stoye et al., "Likelihood-free inference with an improved cross-entropy estimator", (2018), 1808.00973.

[57] D0, "A precision measurement of the mass of the top quark", Nature **429** (2004) 638, hep-ex/0406031.

[58] Y. Gao et al., "Spin Determination of Single-Produced Resonances at Hadron Colliders", Phys. Rev. D **81** (2010) 075022, 1001.3396.

[59] J. Alwall, A. Freitas, and O. Mattelaer, "The Matrix Element Method and QCD Radiation", Phys. Rev. D **83** (2011) 074010, 1010.2263.

[60] S. Bolognesi et al., "On the spin and parity of a single-produced resonance at the LHC", Phys. Rev. D **86** (2012) 095031, 1208.4018.

[61] P. Avery et al., "Precision studies of the Higgs boson decay channel H→ZZ→4ℓ with MEKD", Phys. Rev. D **87** (2013) 055006, 1210.0896.

[62] J. R. Andersen, C. Englert, and M. Spannowsky, "Extracting precise Higgs couplings by using the matrix element method", Phys. Rev. D **87** (2013) 015019, 1211.3011.

[63] P. Artoisenet et al., "Unravelling $t\bar{t}h$ via the Matrix Element Method", Phys. Rev. Lett. **111** (2013) 091802, 1304.6414.

[64] J. M. Campbell et al., "Finding the Higgs boson in decays to $Z\gamma$ using the matrix element method at Next-to-Leading Order", Phys. Rev. D **87** (2013) 073005, 1301.7086.

[65] J. S. Gainer et al., "The Matrix Element Method: Past, Present, and Future", in: *Community Summer Study 2013: Snowmass on the Mississippi*, 2013, 1307.3546.

[66] D. Schouten, A. DeAbreu, and B. Stelzer, "Accelerated Matrix Element Method with Parallel Computing", Comput. Phys. Commun. **192** (2015) 54, 1407.7595.

[67] T. Martini and P. Uwer, "Extending the Matrix Element Method beyond the Born approximation: Calculating event weights at next-to-leading order accuracy", JHEP **09** (2015) 083, 1506.08798.

[68] T. Martini and P. Uwer, "The Matrix Element Method at next-to-leading order QCD for hadronic collisions: Single top-quark production at the LHC as an example application", JHEP **05** (2018) 141, 1712.04527.

[69] M. Kraus, T. Martini, and P. Uwer, "Matrix Element Method at NLO for (anti-)$\mathbf{k_t}$-jet algorithms", Phys. Rev. D **100** (2019) 076010, 1901.08008.

[70] J. Chang et al., "Probing the Top-Yukawa Coupling in Associated Higgs production with a Single Top Quark", JHEP **05** (2014) 062, 1403.2053.

[71] J. Yue, "Enhanced thj signal at the LHC with $h \to \gamma\gamma$ decay and $\mathcal{CP}$-violating top-Higgs coupling", Phys. Lett. B **744** (2015) 131, 1410.2701.

[72] X.-G. He, G.-N. Li, and Y.-J. Zheng, "Probing Higgs boson CP Properties with $t\bar{t}H$ at the LHC and the 100 TeV pp collider", Int. J. Mod. Phys. **A30** (2015) 1550156, 1501.00012.

[73] J. Brehmer et al., "MadMiner: Machine learning-based inference for particle physics", Comput. Softw. Big Sci. **4** (2020) 3, 1907.10621.

[74] P. Artoisenet et al., "A framework for Higgs characterisation", JHEP **11** (2013) 043, 1306.6464.

[75] F. Maltoni, K. Mawatari, and M. Zaro, "Higgs characterisation via vector-boson fusion and associated production: NLO and parton-shower effects", Eur. Phys. J. **C74** (2014) 2710, 1311.1829.

[76] *A morphing technique for signal modelling in a multidimensional space of coupling parameters*, tech. rep. ATL-PHYS-PUB-2015-047, CERN, 2015, URL: http://cds.cern.ch/record/2066980, ATL-PHYS-PUB-2015-047.

[77] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles", (2017), 1612.01474.

[78] J. Brehmer et al., "Better Higgs boson measurements through information geometry", Phys. Rev. D **95** (2017) 073002, 1612.05261.

[79] J. Brehmer et al., "Better Higgs-CP Tests Through Information Geometry", Phys. Rev. D **97** (2018) 095017, 1712.02350.

[80] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters", Bull. Calcutta Math. Soc. **37** (1945) 81.

[81] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, ISBN: 0691080046.

[82] 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, 1405.0301.

[83] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", Comput. Phys. Commun. **178** (2008) 852, 0710.3820.

[84] ATLAS Collaboration, *ATLAS Pythia 8 tunes to* 7 *TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419.

[85] A. Martin et al., "Parton distributions for the LHC", Eur. Phys. J. C **63** (2009) 189, 0901.0002.

[86] M. Whalley, D. Bourilkov, and R. Group, "The Les Houches accord PDFs (LHAPDF) and LHAGLUE", in: *HERA and the LHC: A Workshop on the implications of HERA for LHC physics. Proceedings, Part B*, 2005 575, hep-ph/0508110.

[87] DELPHES 3, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", JHEP **02** (2014) 057, 1307.6346.

[88] M. Cepeda et al., "Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC", in: *Report on the Physics at the HL-LHC,and Perspectives for the HE-LHC*, vol. 7, 2019 221, 1902.00134.

[89] A. Dainese et al., "Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC", 10.23731/CYRM-2019-007 (2019), URL: https://cds.cern.ch/record/2703572.

[90] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", (2016), 1610.07922.

[91] ATLAS, "Measurement of the properties of Higgs boson production at $\sqrt{s}$=13 TeV in the $H \to \gamma\gamma$ channel using 139 fb^{-1} of pp collision data with the ATLAS experiment", (2020), ATLAS-CONF-2020-026.

[92] R. K. Barman, D. Gonçalves, and F. Kling, "Machine Learning the Higgs-Top CP Phase", (2021), 2110.07635.