Optimized probes of $CP$-odd effects in the $t\bar{t}h$ process at hadron colliders

Blaž Bortolato$^a$ Jernej F. Kamenik$^{a,b}$ Nejc Košnik$^{a,b}$ Aleks Smolkovič$^a$

$^a$J. Stefan Institute, Jamova 39, P. O. Box 3000, 1001 Ljubljana, Slovenia
$^b$Department of Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

E-mail: blaz.bortolato@ijs.si, jernej.kamenik@ijs.si, nejc.kosnik@ijs.si, aleks.smolkovic@ijs.si

Abstract: We use machine learning (ML) and non-ML techniques to study optimized $CP$-odd observables, directly and maximally sensitive to the $CP$-odd $i\tilde{\kappa}\gamma^5th$ interaction at the LHC and prospective future hadron colliders using the final state with a Higgs boson and a top quark pair, $pp \rightarrow t\bar{t}h$, followed by semileptonic $t$ decays. We perform phase-space optimization of manifestly $CP$-odd observables ($\omega$), sensitive to the sign of $\tilde{\kappa}$, and constructed from experimentally accessible final state momenta. We identify a simple optimized linear combination $\alpha \cdot \omega$ that gives similar sensitivity as the studied fully fledged ML models. Using $\alpha \cdot \omega$ we project the expected sensitivities to $\tilde{\kappa}$ at HL-LHC, HE-LHC, and FCC-hh.
1 Introduction

The interaction between the heaviest particles of the Standard Model (SM), the top quark $t$ and the Higgs boson $h$, is well known in the SM. The measured top quark mass $m_t$ and the electroweak condensate value $v$ precisely determine the on-shell scalar coupling $-y_t h t$ to be $y_t = \sqrt{2m_t/v}$, while the $P$- and CP-odd interaction $i\gamma^5 h t h$ is absent. Beyond the SM, effective operators of dimension-6 can break this correlation and result in more general (pseudo)scalar $h t h$ couplings $\kappa (\tilde{\kappa})$ [1],

$$L_{ht} = -\frac{y_t}{\sqrt{2}} f(\kappa + i\tilde{\kappa}\gamma_5) t h,$$  \hspace{1cm} (1.1)

which reduce to the SM case at $\kappa = 1$, $\tilde{\kappa} = 0$. At the LHC it is possible to probe these couplings directly with two of the particles in Eq. (1.1) on-shell\footnote{Since $m_h < 2m_t$ one cannot probe these couplings with all the three particles on-shell.} in top-Higgs associated production processes $pp \to thj$ and $pp \to t\bar{t}h$ [2–21].\footnote{The loop induced partonic process $gg \to h \to t\bar{t}$ depends on $\kappa^2$, $\kappa$, and $\tilde{\kappa}^2$ already on the production side as it is dominated by the top quark loop [22].} The corresponding total cross sections scale as $\kappa^2$, $\tilde{\kappa}^2$ ($thj$ also as $\kappa$), and are thus poorly sensitive to small nonzero $\tilde{\kappa}$. Linear sensitivity to $\tilde{\kappa}$ on the other hand can be achieved by measuring $P$- and CP-odd observables.

In a recent paper we have proposed CP-odd probes of $\tilde{\kappa}$ in $thj$ and $t\bar{t}h$ final states at the LHC and prospective future hadron colliders [23]. The overwhelming irreducible backgrounds make the $thj$ channel impractical. For the $t\bar{t}h$ case we have identified 13 different CP-odd observables that can be constructed out of 5 measurable final state momenta and an additional triple-product asymmetry [24]. Namely, assuming $pp \to t\bar{t}h$ production with semileptonically decaying tops, we combined the final state lepton momenta $p_{l+}$, $p_{l-}$,
two $b$-jet momenta $p_b, p_{\bar{b}}$ (although without discriminating their charges) and the Higgs momentum $p_h$ in different ways to construct $C$-even, $P$-odd laboratory frame observables $\omega_i$ [23]. Note that the Higgs momentum $p_h$ can be reconstructed in any feasible final state in the approach we propose. For completeness we list again the 14 $\omega$’s in App. A. Finally, we have also singled out the observable with the largest individual sensitivity to $\tilde{\kappa}$, namely

$$\omega_6 \sim [(p_{\ell^-} \times p_{\ell^+}) \cdot (p_b + p_{\bar{b}})][(p_{\ell^-} - p_{\ell^+}) \cdot (p_b + p_{\bar{b}})].$$  

(1.2)

Due to the high dimensionality and complexity of the phase-space in this process with top quarks decaying semileptonically, in Ref. [23] we have not ventured further in the search for an optimal CP probe of the $t\bar{t}h$ interaction. The aim of the present paper is to finally tackle this problem and use the complete kinematical information accessible experimentally to construct an optimal $CP$-odd observable. To this end we rely on neural networks (NN) trained on Monte-Carlo generated samples to efficiently parametrize the weight function of events across the multi-dimensional phase-space in order to maximize the statistical sensitivity to $\tilde{\kappa}$. We show how the required $P$- and $CP$-symmetry properties of the NN-based observables can be imposed a priori. Finally, we compare in terms of optimality, a general $CP$-odd NN function of the phase-space to a linear combination of manifestly $CP$-odd variables.

The outline of this paper is as follows. In Sec. 2 we perform the phase-space optimization of $\omega_6$, analogous to the study of $th$ production in Ref. [23], but now applied to a multi-dimensional phase-space of semileptonic $t\bar{t}h$ parametrized through a NN. Next, as a generalization to other available $\omega$’s we consider a manifestly $CP$-odd ($C$-even and $P$-odd) observable completely parameterized by a NN. A non-negligible improvement can be achieved, however due to concerns about the complexity and stability with respect to the choice of initial random weights of the NN, we also consider a first order approximation of this observable. In this limit, the significance optimization can be performed without the need for advanced machine learning techniques. At the same time we show that it is just slightly suboptimal compared to the fully fledged NN. We can further simplify this observable by estimating the significance of each term in the linear expansion and keeping only the few most significant terms. We use this optimized observable in Sec. 3 to produce limits in the $\kappa - \tilde{\kappa}$ plane at HL-LHC [25–27], HE-LHC [28, 29], and FCC-hh [30–32]. We conclude in Sec. 4.

2 NN approach to the optimal $CP$-odd observable in $t\bar{t}h$

We implement the training and evaluation of neural networks using the TensorFlow framework [33]. In all cases we train on a sample of $10^7$ $pp \rightarrow h t(\rightarrow b\ell^+\nu)\bar{t}(\rightarrow \bar{b}\ell^-\nu)$ events generated using Madgraph5 [34] with $\tilde{\kappa} = 1$, and split into separate training (7.5M) and test (2.5M) samples. In the following we always set $\kappa = 1$ and only vary $\tilde{\kappa}$, without loss of generality, since the leading $CP$-odd differential rate is proportional to $\tilde{\kappa}\kappa$. Unless stated otherwise the results are shown for events in $pp$ collisions at 14 TeV. We randomly initialize the neural network weights using the default Glorot uniform initializer and use the Adam optimizer with a custom varying learning rate $l(e) = l(e - 1)/(1 + 0.8e)$ where $e$ is the
current epoch and the initial learning rate is set to 0.1. We train all networks using the loss function

$$\text{loss}(\alpha) = \left( \frac{\text{mean}(\mathcal{F}(X; \alpha))}{\text{std}(\mathcal{F}(X; \alpha))/\sqrt{N}} \right)^{-2},$$

where the mean() and the standard deviation std() are to be calculated over all events in the sample. The loss corresponds to the inverse of the significance-squared of the observable $\mathcal{F}(X; \alpha)$ that should be minimized in order to achieve optimal statistical sensitivity. Here $N$ is the size of the sample, $\alpha$ are the free neural network weights and biases and $X$ stands for the values of $CP$-even and/or $CP$-odd phase-space variables in the given event. We avoid over-fitting of the training sample by stopping the training when at least 30 epochs have passed and one of the following two criteria is satisfied: either the running average of 20 training losses saturates to 0.5\% or the running average of 20 test losses increases for 5 epochs in a row. We keep a model history and in the end choose the best model in terms of test loss. In practice we find that mostly the first condition terminates the training loop, and the best model is usually the model from the final epoch of training. In order to determine the optimal NN architecture we perform a scan over a set of possible NN configurations with up to 2 hidden layers and up to 9 nodes per NN layer.\(^3\)

### 2.1 Phase-space optimization of $\omega_6$

Here we study the optimization of the $\omega_6$ variable (1.2) based on phase-space averaging. We denote the $CP$-even phase-space variables with $x$ and a single $CP$-odd one with $\omega_6$. Using this notation we can write the $\bar{t}t$ production differential cross section with semileptonically decaying tops as

$$\frac{d\sigma}{dx d\omega_6} = A(x, |\omega_6|) + \tilde{\kappa} \kappa B(x, \omega_6).$$

where $A$ is manifestly $CP$-even and $B$ a $CP$-odd function of $\omega$ that stems from the interference of scalar and pseudoscalar amplitudes. We do not follow the optimization procedure based on separating $A$ and $B$ since this would require cumbersome multidimensional binning [37]. We use a vector of easily accessible $CP$-even Mandelstam variables $x$:

$$x = \begin{pmatrix} (p_{\ell^+} + p_{\ell^-}) \cdot p_h \\ (p_{\ell^+} + p_{\ell^-}) \cdot (p_b + p_\bar{b}) \\ (p_b + p_\bar{b}) \cdot p_h \\ p_{\ell^+} \cdot p_{\ell^-} \\ p_b \cdot p_\bar{b} \end{pmatrix}.$$  \hspace{1cm} (2.3)

Our goal is to find the optimal $CP$-even weight function $f(x; \alpha)$, which should be used to calculate the weighted average of $\omega_6$. The function $f$ takes $CP$-even quantities $x$ as inputs, therefore we expect its dependence on $\tilde{\kappa}$ to be of the form

$$f(x; \alpha) = C(x; \alpha) + \tilde{\kappa} \kappa^2 D(x; \alpha) + O(\tilde{\kappa}^4).$$  \hspace{1cm} (2.4)

\(^3\)In the initial stages of this study we have also employed an automated algorithm to determine the optimal NN architecture (i.e. Hyperopt [35], see [36] for one of its recent uses.), but we abandoned this approach and settled for manual scans over a set of possible NN configurations in order to have better control over the NN parameters.
Using (2.2) we can now express the observable as
\[
\langle f(x; \alpha) \omega_6 \rangle = \int \frac{d\sigma}{dxd\omega_6} f(x; \alpha) \omega_6 dxd\omega_6
\]
\[
= \tilde{\kappa} \kappa \int B(x, \omega_6) C(x; \alpha) \omega dxd\omega_6 + \tilde{\kappa}^3 \int B(x, \omega_6) D(x; \alpha) dxd\omega_6 + O(\tilde{\kappa}^5),
\]
where the definition of the average is \( \langle \# \rangle \equiv \int \frac{d\sigma}{dxd\omega} \# dxd\omega \). The presence of odd powers of \( \tilde{\kappa} \) reflects the CP-oddness of the observable. The large dimensionality of the phase-space suggests the parameterisation of the function \( f(x; \alpha) \) by means of an appropriate NN. In terms of the loss function (2.1) we have \( F(x, \omega_6; \alpha) = f(x; \alpha) \omega_6 \).

To understand the impact of using different possible neural network architectures, we have performed a manual scan over a set of neural network configurations. The input layer has 5 nodes (one per each \( x \) component) and the output layer has one node resulting in a scalar \( f(x; \alpha) \). We study networks with a single hidden layer of 1-9 nodes and double hidden layer networks with 1-9 nodes each, constraining the number of nodes on the second hidden layer to be smaller than or at most equal to the number of nodes on the first hidden layer. The results of the converged test losses of 50 different random weight initializations per configuration are shown on Fig. 1 in the purple box plot. The plain \( \omega_6 \)-based observable is shown in gray, with the dashed lines denoting its 1\( \sigma \) statistical uncertainty. We find that the phase-space optimization of \( \omega_6 \) gives a noticeable improvement over plain \( \omega_6 \) when using a large enough network. To test how well the resulting network generalizes to other values of \( \tilde{\kappa} \) we use the 50 converged \{9, 9\} models and calculate the dependence of the resulting observable significance with respect to \( \tilde{\kappa} \). This is shown on Fig. 2 where a consistent improvement over simple \( \langle \omega_6 \rangle \) can be seen at all considered \( \tilde{\kappa} \).

As the phase-space optimization of \( \omega_6 \) gives good results, we now turn to the rest of the \( \omega \)'s. However, instead of attempting a phase-space optimization of each of them separately, in the next subsection we consider a more general case where the CP-odd observable itself is parameterized with a neural network.

### 2.2 Neural network as a CP-odd observable

Here we consider a case where the output of the neural network is a CP-odd quantity that defines our observable. We build a network with 14 inputs, one per each \( \omega \), and one output \( F(\omega; \alpha) \), which is correctly anti-symmetrized so that \( F(\omega; \alpha) = -F(-\omega; \alpha) \). The loss function is defined in Eq. (2.1) where now \( F(X; \alpha) = F(\omega; \alpha) \).

We again carry out the study of the dependence of the network size with respect to the test sample loss, including non-negligible uncertainties associated with random weight initializations. We scan the neural network architecture parameter space in the same way as in the previous case, starting with a single hidden layer of 1-9 nodes, then adding an additional hidden layer with the number of nodes smaller than or equal to the number of nodes on the first hidden layer. For each configuration we run 50 trainings with different random weight initializations. The results are again shown in Fig. 1, now in the blue box plot. We find a considerable improvement over the phase-space optimization of the single \( \omega_6 \).
Figure 1. A scan in terms of the test loss (sample size 2.5M) over neural network configurations with one (upper plot) or two (lower plot) hidden layers for the phase-space optimized $\omega_6$ (2.5) shown in the purple box plot and the generalized $F(\omega)$ (Sec. 2.2) shown in the blue box plot. The spread in both cases corresponds to 50 different random weight initializations per configuration. For comparison the plain $\omega_6$ (1.2) is shown in gray with the dashed lines showing its 1\textsigma{} statistical uncertainty. The first order approximation of $F(\omega)$, defined in Eq. (2.6), is shown in red for a full set of $\alpha$’s and in black for a smaller set of selected $\alpha$’s, as described in Sec. 2.3.

Again we check the generalizing power of the resulting observables to other $\tilde{\kappa}$ by fixing the model configuration to $\{9, 9\}$ and calculating the significance of the resulting observables with respect to $\tilde{\kappa}$. The results are shown on Fig. 2. We find a consistent improvement over the previous case across all considered $\tilde{\kappa}$. A noticeable improvement in the significance can be seen, however the results have a non-negligible uncertainty associated with random weight initializations. To address this, we next consider this model in the leading order approximation in $\omega$.

2.3 First order approximation of $F(\omega;\alpha)$

To address the arbitrariness of the neural network architecture choice and the associated stability issues connected to different random weight initializations, in this Section we
Figure 2. Comparison of the significances of all the observables considered in this work with respect to $\kappa$. The results correspond to 1M events per $\kappa$ at 14 TeV. Plain $\omega_6 (1.2)$ in gray, phase-space optimized $\omega_6 (2.5)$ in purple, anti-symmetrized neural network $F(\omega; \alpha)$ (Sec. 2.2) in blue, first order approximation of the latter $(\alpha \cdot \omega)_{\text{all}}$ in red and the selected subset of $\alpha \cdot \omega$ parameters in black (Sec. 2.3). See text for details on each observable.

Figure 3. Optimal weights of the linear observable defined in Eq. (2.6). The left plot shows the significances of $\alpha_j$ (defined as their central value divided by their estimated uncertainty) and is used to extract the most significant contributions to the observable. The uncertainties of $\alpha_j$ are estimated using the expected statistical errors of the observable significances, see text for details. The coefficients of the final observable with only the chosen set of parameters are shown on the right plot, where a good agreement between different energies can be seen. Here the uncertainties are defined through the precision of determining the optimal $\alpha_j$ at a given sample size, see text for details.
consider the first order $\omega$ approximation of the form of $F(\omega; \alpha) = \sum_j \alpha_j \omega_j + \mathcal{O}(\omega^3)$ for $j \in \{1, \ldots, 14\}$. The approximation is also justifiable in terms of a Taylor expansion, as most of the events have $|\omega_j| \ll 1$. The observable is then simply

$$\langle \alpha \cdot \omega \rangle_{\text{all}} = \left\langle \sum_j \alpha_j \omega_j \right\rangle,$$

(2.6)

with the subsidiary condition $|\alpha| = 1$. We can optimize over $\alpha_j$ by maximizing the significance

$$\frac{\partial}{\partial \alpha_j} \frac{\langle \alpha \cdot \omega \rangle_{\text{all}}}{\text{std}(\langle \alpha \cdot \omega \rangle_{\text{all}})} = 0.$$

Doing so we obtain a system of 14 quadratic equations

$$\alpha^T M^{(j)} \alpha = 0,$$

(2.8)

where $\alpha = [\alpha_1, \ldots, \alpha_{14}]^T$ and $14 \times 14$ matrices $M^{(j)}$ are given by

$$M^{(j)}_{ik} = \langle \omega_i \omega_j \rangle \langle \omega_k \rangle - \langle \omega_i \omega_k \rangle \langle \omega_j \rangle.$$

(2.9)

We use this approach to extract the optimal weights $\alpha_j$ from $10^7$ events generated with $\tilde{\kappa} = 1$ at 14, 27, and 100 TeV. We estimate the uncertainty associated with the optimal weights in the following way. First we estimate the statistical spread of the significance obtained with optimal $\alpha_j$. Next we allow $\alpha_j$ to float in the intervals $[\alpha_j - \sigma_j, \alpha_j + \sigma_j]$, where $\sigma_j$ are chosen such that the decrease of the significance due to the change in $\alpha_j$ corresponds to the statistical spread of the significance. We perform an efficient scan around the optimal vector $\alpha$ in its 14-dimensional neighborhood using spherical coordinates to trivially fulfill the normalization constraint $\sum_j \alpha_j^2 = 1$. We approximate the significance with a quadratic function around the extremum to find independent, uncorrelated directions in the $\alpha$-space. With this procedure we determine how sharply the optimal $\alpha_j$ are defined. We estimate the statistical error of the significance using $10^6$ events. Clearly the uncertainties $\sigma_j$ are smaller for larger chosen sample size. The results of this approach are shown on Fig. 3, where the left panel shows significances of each $\alpha_j$, gauging their importance at three different energies. In the next step we choose the minimal set of the most important $\alpha_j$ at each energy that results in the optimal significance (2.7) within the expected statistical fluctuations. This minimal set of optimal $\alpha_j$ is shown on the right panel of Fig. 3 where the uncertainties are now defined as the precision of determining each $\alpha_j$ using the optimization procedure (2.7). A good agreement between energies is achieved, leading to one universal observable with 6 well defined parameters: $\alpha_2, \alpha_4, \alpha_6, \alpha_8, \alpha_{10}, \alpha_{13}$. A comparison of the observable $\langle \alpha \cdot \omega \rangle_{\text{all}}$ and of the reduced combination $\alpha \cdot \omega$ to the other approaches in this work is shown on Fig. 2. We reach a similar level of improvement compared to the full $F(\omega; \alpha)$ network with significantly less parameters.

---

4Note that $|\omega| < 1$ by definition, whereas in some cases (also for $\omega_6$) the upper bound is $1/2$. See App. A.

5Notice that the problem is equivalent to a single neuron NN with 14 inputs and one output without the activation function or the bias term.

6This combination clearly depends on our choice of $10^6$ events for the estimation of the statistical error of the significance. If we were to choose a higher number of events, more $\alpha_j$ would become significant and vice versa.
Figure 4. The $2\sigma$ exclusion zones in the $\kappa - \tilde{\kappa}$ plane by assuming a null result at HL-LHC, HE-LHC and FCC-hh for different luminosities. The optimized observable $\alpha \cdot \omega$ is shown in solid black, while the plain $\omega_0$ (1.2) results are shown using dashed lines. At 14TeV order 1 exclusion can be achieved with $350 \text{fb}^{-1}$ which corresponds to the final integrated luminosity of LHC.

Figure 5. The $2\sigma$ exclusion regions at HL-LHC ($3 \text{ ab}^{-1}$), HE-LHC ($15 \text{ ab}^{-1}$) and FCC-hh ($30 \text{ ab}^{-1}$) by assuming a measurement of a $2\sigma$ positive fluctuation in the optimal observable $\alpha \cdot \omega$ with a selected set of $\alpha_j$ (right plot on Fig. 3).

3 Bounds in the $(\kappa, \tilde{\kappa})$ plane

We produce the bounds in the $(\kappa, \tilde{\kappa})$ plane by including showering and hadronization effects using Pythia8 and detector effects using Delphes with the default ATLAS simulation card. As the $t\bar{t}h$ is followed by semileptonic top decays and $h \rightarrow b\bar{b}$ decay, our signal is defined as 4 $b$-jets and two oppositely charged leptons $\ell$. The main irreducible background is $pp \rightarrow t\bar{t}b\bar{b}$ with both tops decaying semileptonically. We use the same event selection
requirements as in Section 3.2 of [23], where the results of using plain $\omega_6$ are shown. We update those bounds for HL- and HE-LHC and produce bounds for FCC-hh for the first time by using the simplified observable (2.6) with the selected subset of weights shown on the right plot of Fig. 3. The results of assuming a null result up to the expected statistical uncertainty for different luminosities at different energies are shown on Fig. 4. A consistent improvement of sensitivity can be achieved by using the optimized combination of $\omega$’s with respect to a single $\omega_6$. Interestingly the significance improvement is consistent between partonic events and after including shower and detector effects even though the optimization was performed at parton level only. This robustness is a welcome benefit of the method, since the computationally costly optimization procedure does not appear to be sensitive to modeling of the hadronic final states and detector effects.

We show the sensitivity of the optimized observable to the sign of $\tilde{\kappa}$ (and $\kappa$) on Fig. 5 by assuming the measurement of a $2\sigma$ positive statistical fluctuation of the SM case, which in our estimate corresponds to the measurement of $\alpha \cdot \omega = (4.2 \pm 2.1) \times 10^{-4}$, $\alpha \cdot \omega = (0.9 \pm 0.45) \times 10^{-4}$ and $\alpha \cdot \omega = (0.2 \pm 0.1) \times 10^{-4}$ for HL-LHC (3 ab$^{-1}$), HE-LHC (15 ab$^{-1}$) and FCC-hh (30 ab$^{-1}$) respectively.

4 Summary and conclusions

Introducing a set of manifestly $CP$-odd observables ($\omega_i$) built from experimentally accessible final state momenta in $pp \rightarrow t\bar{t}h$ production with semileptonically decaying tops, we studied the prospect of their phase-space optimization, parameterizing the optimal weight functions with neural networks. First we considered the phase-space optimization of a single $\omega_6$, improving its performance. Next we studied a general $CP$-odd observable, parameterized directly by an anti-symmetric neural network, which ended with an even higher performance boost. Lastly we studied the first order approximation of this network as a linear combination of the $CP$-odd observables, producing a simpler and more robust observable. We further simplified it by estimating the significance of each term in the linear expansion and keeping only the few most significant terms. The benefit of using the optimized observable, although marginal for realistic numbers of events, especially at the HL-LHC, carries over from parton level final states to the analysis at event reconstruction level, resulting in projections of probing $\tilde{\kappa}$ directly at HL-LHC, HE-LHC and FCC-hh. We found that the LHC at the end of Run 3 will exclude $\kappa \tilde{\kappa} \sim 1.5$ with $2\sigma$ confidence, while FCC-hh would ultimately be sensitive to $\kappa \tilde{\kappa} \sim 0.01$. Finally, our approach to parametrizing CP-odd observables over high-dimensional phase-spaces using manifestly CP-odd NNs could be applied to other high energy particle production and decay processes, as well as to other symmetries. We leave the exploration of these ideas for future work.

Acknowledgments

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P1-0035 and J1-8137). This article is based upon work from COST Action CA16201 PARTICLEFACE supported by COST (European Cooperation in
Finally, we have a triple product:

\[ \omega_1 \sim [(p_{\ell^{-}} \times p_{\ell^{+}}) \cdot p_{h}] \cdot (p_{\ell^{-}} - p_{\ell^{+}}) \cdot p_{h}, \]  
\[ \omega_2 \sim [(p_{\ell^{-}} \times p_{\ell^{+}}) \cdot p_{h}] \cdot ((p_{\ell^{-}} - p_{\ell^{+}}) \cdot (p_{\ell^{-}} + p_{\ell^{+}})) \cdot p_{h}, \]  
\[ \omega_3 \sim [(p_{\ell^{-}} \times p_{\ell^{+}}) \cdot p_{h}] \cdot ((p_{\ell^{-}} - p_{\ell^{+}}) \cdot (p_{h} + p_{b})) \cdot p_{h}, \]  
\[ \omega_4 \sim [(p_{\ell^{-}} \times p_{\ell^{+}}) \cdot (p_{b} + p_{h})] \cdot ((p_{\ell^{-}} - p_{\ell^{+}}) \cdot p_{h}) \cdot p_{h}, \]  
\[ \omega_5 \sim [(p_{\ell^{-}} \times p_{\ell^{+}}) \cdot (p_{b} + p_{h})] \cdot ((p_{\ell^{-}} - p_{\ell^{+}}) \cdot (p_{\ell^{-}} + p_{\ell^{+}})) \cdot p_{h}, \]  
\[ \omega_6 \sim [(p_{\ell^{-}} \times p_{\ell^{+}}) \cdot (p_{b} + p_{h})] \cdot ((p_{\ell^{-}} - p_{\ell^{+}}) \cdot (p_{h} + p_{b})) \cdot p_{h}. \]  

The second class involves \( p_{h} \times p_{b} \) and \( p_{h} - p_{b} \) in the two scalar products:

\[ \omega_7 \sim [(p_{b} \times p_{b}) \cdot p_{h}] \cdot (p_{h} - p_{b}) \cdot p_{h}, \]  
\[ \omega_8 \sim [(p_{b} \times p_{b}) \cdot p_{h}] \cdot ((p_{b} - p_{b}) \cdot (p_{\ell^{-}} + p_{\ell^{+}})) \cdot p_{h}, \]  
\[ \omega_9 \sim [(p_{b} \times p_{b}) \cdot p_{h}] \cdot ((p_{b} - p_{b}) \cdot (p_{h} + p_{b})) \cdot p_{h}, \]  
\[ \omega_{10} \sim [(p_{b} \times p_{b}) \cdot (p_{\ell^{-}} + p_{\ell^{+}})] \cdot (p_{h} - p_{b}) \cdot p_{h}, \]  
\[ \omega_{11} \sim [(p_{b} \times p_{b}) \cdot (p_{\ell^{-}} + p_{\ell^{+}})] \cdot ((p_{b} - p_{b}) \cdot (p_{\ell^{-}} + p_{\ell^{+}})) \cdot p_{h}, \]  
\[ \omega_{12} \sim [(p_{b} \times p_{b}) \cdot (p_{\ell^{-}} + p_{\ell^{+}})] \cdot ((p_{b} - p_{b}) \cdot (p_{b} + p_{b})) \cdot p_{h}, \]  
\[ \omega_{13} \sim [(p_{b} \times p_{b}) \cdot (p_{\ell^{-}} - p_{\ell^{+}})] \cdot ((p_{b} - p_{b}) \cdot (p_{\ell^{-}} - p_{\ell^{+}})) \cdot p_{h}. \]

Finally, we have a triple product:

\[ \omega_{14} \sim [p_{h} \times (p_{\ell^{-}} + p_{\ell^{+}})] \cdot (p_{b} + p_{h}). \]  

All the \( \omega \)s are normalized by the lengths of all the vectors that enter as factors in the scalar products, implying \( |\omega_i| \leq 1 \). In case when \( \omega_i \) is of the form \( A \cdot B A \cdot C \) with \( B \cdot C = 0 \) the upper bound is \( |\omega_i| \leq 1/2 \).

**References**

[1] J. A. Aguilar-Saavedra, *A Minimal set of top-Higgs anomalous couplings*, Nucl. Phys. B821 (2009) 215 [0904.2387].

[2] J. Ellis, D. S. Hwang, K. Sakurai and M. Takeuchi, *Disentangling Higgs-Top Couplings in Associated Production*, JHEP 04 (2014) 004 [1312.5736].

[3] ATLAS, CMS collaboration, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at \( \sqrt{s} = 7 \) and 8 TeV*, JHEP 08 (2016) 045 [1606.02266].
[4] G. Bhattacharyya, D. Das and P. B. Pal, *Modified Higgs couplings and unitarity violation*, *Phys. Rev.* **D87** (2013) 011702 [1212.4651].

[5] B. Grzadkowski and J. F. Gunion, *Using decay angle correlations to detect CP violation in the neutral Higgs sector*, *Phys. Lett.* **B350** (1995) 218 [hep-ph/9501339].

[6] 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.* **C74** (2014) 3065 [1407.5089].

[7] F. Boudjema, R. M. Godbole, D. Guadagnoli and K. A. Mohan, *Lab-frame observables for probing the top-Higgs interaction*, *Phys. Rev.* **D92** (2015) 015019 [1501.03157].

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

[9] 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*, *JHEP* **07** (2016) 056 [1603.03632].

[10] A. V. Gritsan, R. Röntsch, M. Schulze and M. Xiao, *Constraining anomalous Higgs boson couplings to the heavy fermions using matrix element techniques*, *Phys. Rev.* **D94** (2016) 055023 [1606.03107].

[11] J. Li, Z.-g. Si, L. Wu and J. Yue, *Central-edge asymmetry as a probe of Higgs-top coupling in tth production at the LHC*, *Phys. Lett.* **B779** (2018) 72 [1701.00224].

[12] S. Amor Dos Santos et al., *Probing the CP nature of the Higgs coupling in tth events at the LHC*, *Phys. Rev.* **D96** (2017) 013004 [1704.03565].

[13] D. Gonçalves, K. Kong and J. H. Kim, *Probing the top-Higgs Yukawa CP structure in dileptonic tth with M2-assisted reconstruction*, *JHEP* **06** (2018) 079 [1804.05874].

[14] A. Kobakhidze, L. Wu and J. Yue, *Anomalous Top-Higgs Couplings and Top Polarisation in Single Top and Higgs Associated Production at the LHC*, *JHEP* **10** (2014) 100 [1406.1961].

[15] J. Yue, *Enhanced thj signal at the LHC with h → γγ decay and CP-violating top-Higgs coupling*, *Phys. Lett.* **B744** (2015) 131 [1410.2701].

[16] F. Demartin, F. Maltoni, K. Mawatari and M. Zaro, *Higgs production in association with a single top quark at the LHC*, *Eur. Phys. J.* **C75** (2015) 267 [1504.00611].

[17] 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].

[18] M. Kraus, T. Martini, S. Peitzsch and P. Uwer, *Exploring BSM Higgs couplings in single top-quark production*, 1908.09100.

[19] W. Bernreuther and A. Brandenburg, *Tracing CP violation in the production of top quark pairs by multiple TeV proton proton collisions*, *Phys. Rev.* **D49** (1994) 4481 [hep-ph/9312210].

[20] V. Barger, K. Hagiwara and Y.-J. Zheng, *Probing the top Yukawa coupling at the LHC via associated production of single top and Higgs*, 1912.11795.

[21] R. Patrick, A. Scaffidi and P. Sharma, *Top polarisation as a probe of CP-mixing top-Higgs coupling in tth signals*, *Phys. Rev. D* **101** (2020) 093005 [1909.12772].

[22] J. Brod, U. Haisch and J. Zupan, *Constraints on CP-violating Higgs couplings to the third generation*, *JHEP* **11** (2013) 180 [1310.1385].
[23] D. A. Faroughy, J. F. Kamenik, N. Košnik and A. Smolkovič, *Probing the CP nature of the top quark Yukawa at hadron colliders*, *JHEP* **02** (2020) 085 [1909.00007].

[24] G. Durieux and Y. Grossman, *Probing CP violation systematically in differential distributions*, *Phys. Rev. D* **92** (2015) 076013 [1508.03054].

[25] *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1,*.

[26] CMS collaboration, *Projected Performance of an Upgraded CMS Detector at the LHC and HL-LHC: Contribution to the Snowmass Process*, in *Community Summer Study 2013: Snowmass on the Mississippi*, 7, 2013, 1307.7135.

[27] ATLAS collaboration, *Physics at a High-Luminosity LHC with ATLAS*, in *Community Summer Study 2013: Snowmass on the Mississippi*, 7, 2013, 1307.7292.

[28] F. Zimmermann et al., *High-Energy LHC Design*, *J. Phys. Conf. Ser.* **1067** (2018) 022009.

[29] FCC collaboration, *HE-LHC: The High-Energy Large Hadron Collider: Future Circular Collider Conceptual Design Report Volume 4*, *Eur. Phys. J. ST* **228** (2019) 1109.

[30] M. L. Mangano et al., *Physics at a 100 TeV pp Collider: Standard Model Processes*, *CERN Yellow Rep.* (2017) 1 [1607.01831].

[31] R. Contino et al., *Physics at a 100 TeV pp collider: Higgs and EW symmetry breaking studies*, *CERN Yellow Rep.* (2017) 255 [1606.09408].

[32] FCC collaboration, *FCC-hh: The Hadron Collider*, *Eur. Phys. J. ST* **228** (2019) 755.

[33] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro et al., *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015.

[34] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer 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].

[35] J. Bergstra, D. Yamins and D. D. Cox, *Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures*, in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML’13, p. I–115–I–123, JMLR.org, 2013.

[36] J. M. Clavijo, P. Glaysher and J. M. Katzy, *Adversarial domain adaptation to reduce sample bias of a high energy physics classifier*, 2005.00568.

[37] D. Atwood and A. Soni, *Analysis for magnetic moment and electric dipole moment form-factors of the top quark via e^+e^- → têt*, *Phys. Rev.* **D45** (1992) 2405.