Article

Optimized Probes of the *CP* Nature of the Top Quark Yukawa Coupling at Hadron Colliders

Darius A. Faroughy [1], Blaž Bortolato [2], Jernej F. Kamenik [2,3,*] , Nejc Košnik [2,3] and Aleks Smolkovič [2]

[1] Physik-Institut, Universitat Zurich, CH-8057 Zurich, Switzerland; faroughy@physik.uzh.ch
[2] Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia; Blaz.Bortolato@ijs.si (B.B.); nejc.kosnik@ijs.si (N.K.); Aleks.Smolkovic@ijs.si (A.S.)
[3] Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia
[*] Correspondence: jernej.kamenik@ijs.si

Abstract: We summarize our recent proposals for probing the *CP*-odd $i\tilde{\kappa}\bar{t}\gamma^5 th$ interaction at the LHC and its projected upgrades directly using associated on-shell Higgs boson and top quark or top quark pair production. We first recount how to construct a *CP*-odd observable based on top quark polarization in $Wb \to th$ scattering with optimal linear sensitivity to $\tilde{\kappa}$. For the corresponding hadronic process $pp \to thj$ we then present a method of extracting the phase-space dependent weight function that allows to retain close to optimal sensitivity to $\tilde{\kappa}$. For the case of top quark pair production in association with the Higgs boson, $pp \to t\bar{t}h$, with semileptonically decaying tops, we instead show how one can construct manifestly *CP*-odd observables that rely solely on measuring the momenta of the Higgs boson and the leptons and b-jets from the decaying tops without having to distinguish the charge of the b-jets. Finally, we introduce machine learning (ML) and non-ML techniques to study the phase-space optimization of such *CP*-odd observables. We emphasize a simple optimized linear combination $\boldsymbol{\alpha} \cdot \boldsymbol{\omega}$ that gives similar sensitivity as the studied fully fledged ML models. Using $\boldsymbol{\alpha} \cdot \boldsymbol{\omega}$ we review sensitivity projections to $\tilde{\kappa}$ at HL-LHC, HE-LHC, and FCC-hh.

Keywords: top quark; Higgs boson; CP violation; optimized observables; beyond the Standard Model

1. Introduction

The interaction between the heaviest particles of the Standard Model (SM), the top quark t and the Higgs boson h, is an important target for the LHC experiments. It is precisely predicted within the SM. The measured top quark mass m_t and the electroweak condensate value v precisely determine the on-shell scalar (P- and CP-even) coupling $y_t = \sqrt{2}m_t/v$, while P- and CP-odd interactions are absent. Beyond the SM, effective operators of dimension-6 can break this correlation and result in more general (pseudo)scalar $\bar{t}th$ couplings κ ($\tilde{\kappa}$) [1]

$$\mathcal{L}_{ht} = -\frac{y_t}{\sqrt{2}}\bar{t}(\kappa + i\tilde{\kappa}\gamma_5)th \,, \tag{1}$$

which reduce to the SM case at $\kappa = 1$, $\tilde{\kappa} = 0$.

CP-violating h couplings, like $\tilde{\kappa}$, are particularly interesting as any sign of *CP* violation in Higgs processes would constitute an indisputable New Physics (NP) signal. Existing data on Higgs production and decays is already precise enough to constrain any isolated modification of the top Yukawa to $\mathcal{O}(1)$ [2–4]. However, all existing measurements are based on *CP*-even observables with very limited sensitivity to *CP*-odd modifications of the top quark Yukawa. In principle, indirect collider bounds from Higgs decay and production ($gg \to h$, $h \to \gamma\gamma$), and especially the low-energy bounds on electric dipole moments (EDMs) of atoms and nuclei that target specifically *CP*-odd effects [2,5,6], are currently more constraining than direct collider probes. However, these constraints are subject to assumptions about other Higgs interactions, and in particular in the case of EDMs also other contributions unrelated to the Higgs.

At the LHC it is possible to probe these couplings directly with two of the particles in Equation (1) on-shell (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 $pp \to t\bar{t}h$ [2–4,6–23] (The loop induced partonic process $gg \to h \to t\bar{t}$ depends on κ^2, κ, and $\tilde{\kappa}^2$ already on the production side as it is dominated by the top quark loop [5]). The corresponding total cross sections scale as κ^2, $\tilde{\kappa}^2$ (*thj* also as κ), 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.

We have recently addressed this challenge by identifying observables with optimal sensitivity to a single *CP*-odd parameter in both *th* and *t\bar{t}h* associated production at the LHC, which can be realistically measured and exhibit close to optimal sensitivity to *CP*-odd interactions between the Higgs boson and the top quark [24]. The proposed observables in *th* are based on optimization of top-spin correlations previously studied in $t\bar{t}$ production [25]. Unfortunately, the overwhelming irreducible backgrounds make the *thj* channel impractical. On the other hand, in the case of $\bar{t}th$ this procedure becomes intractable in practice and our construction relies instead on *CP*- and *P*-symmetry arguments. In total, we can identify 21 different *CP*-odd observables that can be constructed out of 5 measurable final state momenta and an additional triple-product asymmetry [26]. Namely, assuming $pp \to t\bar{t}h$ production with semileptonically decaying tops, we combine the final state lepton momenta $\boldsymbol{p}_{\ell^+}$, $\boldsymbol{p}_{\ell^-}$, two *b*-jet momenta $\boldsymbol{p}_b$, $\boldsymbol{p}_{\bar{b}}$ from decaying top quarks (although without discriminating their charges (Efficient *b*-jet charge discrimination could allow to construct further *CP*-odd observables, see e.g., [6,20])) and the Higgs momentum $\boldsymbol{p}_h$ in different ways to construct *C*-even, *P*-odd laboratory frame observables ω_i [24].

Due to the high dimensionality and complexity of the phase-space in this process with top quarks decaying semileptonically, using the complete kinematical information accessible experimentally to construct an optimal *CP*-odd observable is challenging. To this end we have employed 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 $\tilde{\kappa}$ [27]. 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 present paper serves as a pedagogical review of the work first reported in Refs. [24,27].

2. Optimal *CP*-Odd Observable in the $bW \to th$ Process

2.1. Parton Level $Wb \to th$ Analysis

We begin by studying the effects of $\tilde{\kappa}$ on top spin observables in the idealized case of single top quark production in partonic process $W(p)b(p') \to t(k)h(k')$. Here the complete polarized scattering amplitudes can be found in a compact analytic form. This process can actually be connected to a more realistic $pp \to thj$ production in the high energy limit, where the *W* and *b* quark mass effects are negligible and the collinear emission of both initial state 'partons' can be described by the corresponding parton distribution functions. (See e.g., Section 3 of Ref. [28] for an extended discussion on the validity of this approximation) Three diagrams contribute to such parton level Higgs-top production in the SM, shown in Figure 1. Neglecting furthermore the mass (and thus the corresponding Yukawa coupling) of the bottom quark, we consider only the first two of the diagrams in Figure 1. The formalism presented here is based on Refs. [25,29,30]. First, we introduce the spin projection operator [31]

$$P(s) = \frac{1}{2}(1 + \gamma_5 \slashed{s}),$$
(2)

where s^μ is a top spin four-vector, defined in a general frame as

$$s^\mu = \left(\frac{\boldsymbol{k} \cdot \hat{\boldsymbol{s}}}{m_t}, \hat{\boldsymbol{s}} + \frac{\boldsymbol{k}(\boldsymbol{k} \cdot \hat{\boldsymbol{s}})}{m_t(E_t + m_t)} \right).$$
(3)

Vector $\boldsymbol{k}$ is the top quark momentum and $\hat{\boldsymbol{s}}$ is an arbitrary unit vector that represents polarization of the top quark. The physical significance of $\hat{\boldsymbol{s}}$ is revealed if we make a rotation-free boost to the top rest frame where we find $s^{*\mu} = (0, \hat{\boldsymbol{s}})$. Note that spatial component of four-vector x^μ in this case transforms as $\boldsymbol{x}^* = \boldsymbol{x} + \left(\frac{\boldsymbol{x} \cdot \boldsymbol{k}}{m_t(E_t + m_t)} - \frac{x^0}{m_t} \right) \boldsymbol{k}$ upon boost to the top rest frame. Therefore $s^2 = -1$, $s \cdot k = 0$, and $\hat{\boldsymbol{s}}$ corresponds to the polarization of the top quark in its rest frame. Projection onto a well defined polarization of the top quark is achieved by placing the projector (2) in front of the bispinors at the amplitude level:

$$u \to P(s)u, \tag{4}$$
$$v \to P(s)v. \tag{5}$$

Sum over top quark polarizations r then becomes

$$\sum_r u(k,r)\bar{u}(k,r) \to P(s)(\slashed{k} + m)P(s) = (\slashed{k} + m)P(s),$$
$$\sum_r v(k,r)\bar{v}(k,r) \to P(s)(\slashed{k} - m)P(s) = (\slashed{k} - m)P(s), \tag{6}$$

and the cross-section is linear in s^μ

$$|\mathcal{M}|^2 = a + b_\mu s^\mu. \tag{7}$$

Here the polarization vector $\hat{\boldsymbol{s}}$ is arbitrary while b_μ contains all the information about the polarization of the top quark in the given process. The parton level cross section can be written as

$$d\sigma = \Phi_{\text{in}} |\mathcal{M}|^2 d\Gamma_{th}, \tag{8}$$

where Φ_{in} is the initial state flux normalization and $d\Gamma_{th}$ is the *th* phase space volume. On the other hand, in the top rest frame it is convenient to introduce the spin density matrix as

$$\rho = A + B_i \sigma_i, \tag{9}$$

such that the unpolarized cross section is proportional to $\overline{|\mathcal{M}|^2} = \text{Tr}[\rho] = 2A$. Here σ are the Pauli matrices. In the density matrix formalism, the expectation value of a generic operator is obtained as

$$\langle \mathcal{O} \rangle = \text{Tr}[\rho \mathcal{O}]. \tag{10}$$

In particular, the polarized cross section along $\hat{\boldsymbol{s}}$ is obtained as the expectation value of the projector:

$$|\mathcal{M}|^2 = \text{Tr}\left[\rho \frac{1}{2}(1 + \hat{\boldsymbol{s}} \cdot \boldsymbol{\sigma}) \right] = A + B_i \hat{s}_i. \tag{11}$$

One can determine the rest-frame coefficients A, B_i from a, b_μ by comparing the expressions for polarized $|\mathcal{M}|^2$, expressed via Equations (7) and (11). The result of this matching are explicit expressions:

$$A = a, \qquad B_i = -b^i + \frac{1}{m_t}\left(b^0 - \frac{\boldsymbol{b} \cdot \boldsymbol{k}}{E_t + m_t} \right) \boldsymbol{k}. \tag{12}$$

The rest-frame polarization of the top quark along a vector $\hat{\boldsymbol{s}}$ is given by the expectation value of $\mathcal{O}_{\hat{s}} = \hat{\boldsymbol{s}} \cdot \frac{\sigma}{2}$,

$$\langle \mathcal{O}_{\hat{s}} \rangle = \boldsymbol{B} \cdot \hat{\boldsymbol{s}}. \tag{13}$$

This observable can be determined for example by measuring the angular distribution of the charged lepton in the semi-leptonic top decay $t \to b(W \to \ell\nu)$ since the charged lepton in top decay is considered to be an almost perfect top spin analyzer, i.e., the angular

decay distribution vanishes when the lepton momentum is opposite to the spin of t [32]. The angular distribution in the top rest frame

$$\frac{1}{\Gamma_t}\frac{d\Gamma_t}{d\cos\theta_\ell} = \frac{1}{2}(1 + \langle\mathcal{O}_{\hat{s}}\rangle\cos\theta_\ell) \tag{14}$$

then allows for experimental extraction of the B^i coefficients [32].
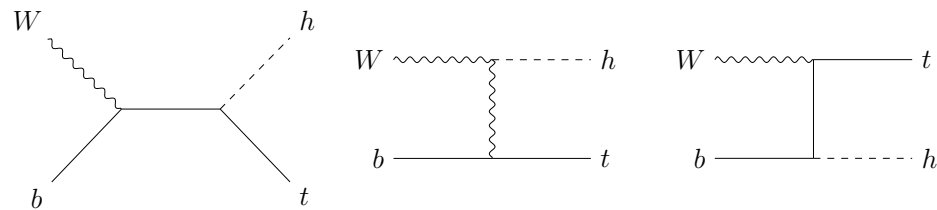
Figure 1. Tree-level diagrams contributing to $Wb \to th$.

Here θ_ℓ is an angle between the lepton and the polarization axis $\hat{s}$ in the top rest frame. The above construction shows that the vector $\hat{s}$ is an arbitrary unit vector defined in the laboratory frame. A particular choice $\hat{s} = k/|k|$ implies that $\langle\mathcal{O}_{\hat{s}}\rangle$ measures the top quark helicity. Another natural choice for $\hat{s}$ is the W momentum $\hat{p}$, also known as the beam basis, which has to be carefully defined in actual pp collisions where the (symmetric) initial state does not allow to define forward and backward directions. Experimentally one has to reconstruct the top quark rest frame in order to be able to trace the angular distribution of the lepton with respect to the chosen $\hat{s}$ and gain access to the coefficients B_i. In the following we will show how to optimize the choice of $\hat{s}$ such that the sensitivity to the CP-violating parameter $\tilde{\kappa}$ is maximized.

In the Wb center-of-mass frame we can define the W and t momenta as

$$\begin{aligned}\hat{p} &= (0,0,1),\\ \hat{k} &= (\sin\theta, 0, \cos\theta),\end{aligned} \tag{15}$$

where θ is the angle between the direction of the top quark and the W boson. We have set the azimuthal angle $\phi = 0$ without loss of generality. The polarization vector components B_i in this case depend on $x = \cos\theta$. Evaluation of the two diagrams in Figure 1 leads to the coefficients $B_{1,2,3}(x)$ of the polarized cross-section. It turns out that in the coordinate system in Equation (15) the analytical expression for $B_2(x)$ is linear in $\tilde{\kappa}$,

$$B_2(x) \equiv \beta(x)\tilde{\kappa}, \tag{16}$$

whereas $B_{1,3}$ do not contain linear $\tilde{\kappa}$ terms. Effectively this means that we should choose the vector $\hat{s}$ to be orthogonal to the plane spanned by the W and t momenta in order to probe $\tilde{\kappa}$ with linear sensitivity. Similar results have been reported in Ref. [18]. In pursuit of maximal sensitivity to $\tilde{\kappa}$ we choose the polarization vector in each event as $\hat{s} = \hat{p} \times \hat{k}/|\hat{p} \times \hat{k}|$. In this case an interesting experimental quantity a two-fold differential cross-section

$$\begin{aligned}\frac{d^2\sigma}{dx\,d\cos\theta_\ell}(Wb \to hb\ell\nu) &= \Sigma(x,\hat{s})\frac{\text{Br}(t \to b\ell\nu)}{2}(1 + \cos\theta_\ell)\\ &+ \Sigma(x,-\hat{s})\frac{\text{Br}(t \to b\ell\nu)}{2}(1 - \cos\theta_\ell)\,,\end{aligned} \tag{17}$$

where we have approximated the intermediate top quark as a narrow resonance and $\Sigma(x,\hat{s}) = d\sigma/dx(Wb \to t^{(\hat{s})}h)$ is the differential production cross section for the top quarks polarized in the $\hat{s}$ direction. Using Equation (11) and inserting $\hat{s}$ we have $\Sigma(x,\pm\hat{s}) =$

$\Phi_{\text{in}}(A(x) \pm \tilde{\kappa}\beta(x))$, where Φ_{in} is the initial flux normalization. Thus we can write Equation (17) as

$$\frac{d^2\sigma}{dx\,d\cos\theta_\ell}(Wb \to hb\ell\nu) = \Phi_{\text{in}}\text{Br}(t \to b\ell\nu)(A(x) + \tilde{\kappa}\beta(x)\cos\theta_\ell). \tag{18}$$

Treating $\tilde{\kappa}$ as a small perturbation we can integrate the distribution in Equation (18) with a phase-space dependent function f that would maximize statistical sensitivity of the integral to $\tilde{\kappa}$. It has been shown in Refs. [33,34] that such an optimal function should be the ratio of the $\tilde{\kappa}$-perturbation to the unperturbed distribution, in our case $f(x, \cos\theta_\ell) = \frac{\beta(x)}{A(x)}\cos\theta_\ell$. The optimal observable is thus

$$\mathcal{O}_{\text{opt.}}^{Wb\to th} \equiv \frac{1}{\sigma}\int dx\,d\cos\theta_\ell\,\frac{d^2\sigma}{dx\,d\cos\theta_\ell}\frac{\beta(x)}{A(x)}\cos\theta_\ell = \frac{1}{N}\sum_{i=1}^{N}\frac{\beta(x_i)}{A(x_i)}\cos\theta_{\ell,i}, \tag{19}$$

where θ_ℓ is the angle between $\hat{s}$ and the lepton momentum in the top center-of mass-frame, as defined in the preceding paragraph. The index $i = 1, \ldots, N$ labels individual events. The prediction scales as $\langle\beta^2\rangle$,

$$\mathcal{O}_{\text{opt.}}^{Wb\to th} = \frac{\tilde{\kappa}}{3}\left[\int dx\,\frac{[\beta(x)]^2}{A(x)}\right]\Big/\left[\int dx\,A(x)\right], \tag{20}$$

where we have integrated over $\cos\theta_\ell$ and left the bounds for $x = \cos\theta$ unspecified. The function $\beta(x)$ is plotted in Figure 2.

To carry over the presented formalism to the realistic case of pp collisions, we have to adapt the beam axis by referring only to experimentally accessible momenta. Using the reconstructed top momentum k as a reference, we define the positive z-direction as the parallel top quark momentum projection $\hat{k}_\parallel$. The top quark is then by definition in the positive hemisphere, $\tilde{x} = \cos\tilde{\theta} \geq 0$, where $\tilde{\theta}$ is the angle between k and $\hat{k}_\parallel$. The optimal polarization direction with linear $\tilde{\kappa}$ sensitivity now becomes $\hat{s} = \hat{k}_\parallel \times \hat{k}_\perp$ upon which an experiment should measure the lepton angle $\tilde{\theta}_\ell$. The cross-section distributions in $\tilde{x}$ and x are related via

$$\frac{d^2\sigma}{d\tilde{x}\,d\cos\tilde{\theta}_\ell} = \frac{d^2\sigma}{dx\,d\cos\theta_\ell}\bigg|_{x=\tilde{x},\cos\theta_\ell=\cos\tilde{\theta}_\ell} + \frac{d^2\sigma}{dx\,d\cos\theta_\ell}\bigg|_{x=-\tilde{x},\cos\theta_\ell=-\cos\tilde{\theta}_\ell}$$
$$= \Phi_{\text{in}}\frac{\text{Br}(t \to b\ell\nu)}{2}\left[\tilde{A}(\tilde{x}) + \tilde{\kappa}\cos\tilde{\theta}_\ell\tilde{\beta}(\tilde{x})\right], \tag{21}$$

where

$$\begin{aligned}\tilde{A}(\tilde{x}) &\equiv A(\tilde{x}) + A(-\tilde{x}),\\ \tilde{\beta}(\tilde{x}) &\equiv \beta(\tilde{x}) - \beta(-\tilde{x}).\end{aligned} \tag{22}$$

The $\cos\theta_\ell$ is flipped in the second term since for $\tilde{x} = -x$ the polarization vector $\hat{s} = \hat{k}_\parallel \times \hat{k}_\perp$ flips the direction compared to the previous definition, $\hat{s} \sim p \times k$. The optimal observable in this case is finally

$$\tilde{\mathcal{O}}_{\text{opt.}}^{Wb\to th} \equiv \frac{1}{\sigma}\int d\tilde{x}\,d\cos\tilde{\theta}_\ell\,\frac{d^2\sigma}{d\tilde{x}\,d\cos\tilde{\theta}_\ell}\cos\tilde{\theta}_\ell\,\frac{\tilde{\beta}(\tilde{x})}{\tilde{A}(\tilde{x})}$$
$$= \frac{\tilde{\kappa}}{3}\left[\int d\tilde{x}\,\frac{[\tilde{\beta}(\tilde{x})]^2}{\tilde{A}(\tilde{x})}\right]\Big/\left[\int d\tilde{x}\,\tilde{A}(\tilde{x})\right], \tag{23}$$

where the optimal weight function is again taken to be the ratio of the $\tilde{\kappa}$-perturbation to the unperturbed distribution (see Equation (21)), in line with Refs. [33,34]. In terms of

experimental N data points with reconstructed $\tilde{\theta}_\ell$ and $\tilde{x}$ the observable is obtained as the following average:

$$\tilde{O}_{\text{opt.}}^{Wb\to th} = \frac{1}{N} \sum_{i=1}^{N} \frac{\tilde{\beta}(\tilde{x}_i)}{\tilde{A}(\tilde{x}_i)} \cos\tilde{\theta}_{\ell,i} \,. \tag{24}$$

In the limit where $\beta(x) = -\beta(-x)$ the observables are equal, $\tilde{O}_{\text{opt.}}^{Wb\to th} = O_{\text{opt.}}^{Wb\to th}$. However in general the $\tilde{O}_{\text{opt.}}^{Wb\to th}$ is expected to result in a weaker statistical significance due to our inability to determine the direction of the top quark with respect to the initial W. In Figure 2 we show that $\beta(x)$ is large at negative x and we have $\tilde{\beta}(\tilde{x}) \approx -\beta(-\tilde{x})$, for representative values of the center-of-mass energy $\sqrt{s}$.
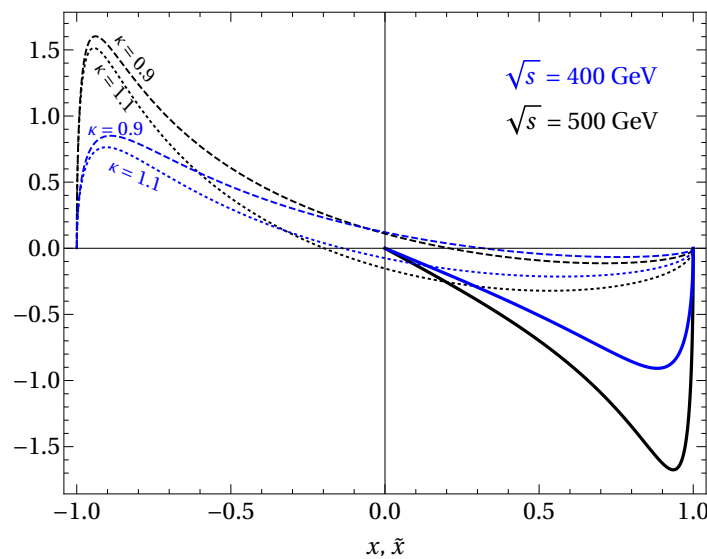
Figure 2. Comparison of the $\beta(x)$ (dashed and dotted) and $\tilde{\beta}(\tilde{x})$ (full line) polarization functions at representative CMS energies $\sqrt{s}$ and two values of κ. We find that $\tilde{\beta}(\tilde{x})$ is independent of κ [24].

2.2. Hadronic Process $pp \to thj$

Here we demonstrate how to carry over the optimal *CP*-violating observable from parton level to the realistic case of pp collisions, but still neglecting reconstruction efficiencies and backgrounds. The parton level observable defined in Equation (23) can be adapted to this case with an additional integration over the parton distribution functions (PDFs). Since the hadronic cross section is a convolution of partonic cross sections it can be split into a $\tilde{\kappa}$-independent piece and the small perturbation proportional to $\tilde{\kappa}$, similar to the partonic cross section in Equation (21). Assuming that the Higgs decays into visible states, the missing p_T is only due to the neutrino originating from the top decay. Thus we can reconstruct the top quark momentum and kinematic quantities of Equation (21). For hadronic collisions one can express the cross section as

$$\frac{d^2\sigma^{pp\to thj}}{d\tilde{x}\, d\cos\tilde{\theta}_\ell} = \mathcal{A}(\tilde{x}) + \tilde{\kappa}\cos\tilde{\theta}_\ell \mathcal{B}(\tilde{x}), \tag{25}$$

and weigh the events with the optimal $f_{\text{opt.}} \propto \cos\tilde{\theta}_\ell \mathcal{B}/\mathcal{A}$. To demonstrate the procedure we have used the Monte Carlo event generator `MadGraph5` [35,36] together with the Higgs Characterisation UFO model [37,38] (for an analysis of next-to-leading order QCD and next-to-next-to leading EW effects see Refs. [8,17,39], respectively) to incorporate the κ and $\tilde{\kappa}$ couplings in the simulation of the $pp \to t(\to b\ell\nu)hj$ signal. The procedure of extracting the optimal weight function $\mathcal{B}/\mathcal{A}$ from MC simulations and using it to produce the optimal observable goes as follows:

1.   Choose the bins for $\tilde{x}$ between $\tilde{x}_{\min} \geq 0$ and $\tilde{x}_{\max} \leq 1$.

2. Fix $\tilde{\kappa}$ and extract from the MC simulation the mean $\langle \cos \tilde{\theta}_\ell \rangle$ in each of the $\tilde{x}$ bins. The obtained value corresponds to the weight $\frac{1}{3} \mathcal{B}/\mathcal{A}$ in this bin, see Equation (25).

3. Use this information to weigh experimental events bin-by-bin with $f_{\text{opt.}} \propto \cos \tilde{\theta}_\ell \mathcal{B}/\mathcal{A}$. The normalization of $f_{\text{opt.}}$ is fixed by the requirement $\int d\tilde{x} \mathcal{B}/\mathcal{A} = 1$.

This optimization procedure is independent of the $\tilde{\kappa}$ value as long as $\tilde{\kappa}$ is sufficiently small. The resulting optimal weight $\mathcal{B}/\mathcal{A}$ is shown in Figure 3, where we compare it for different final states (*thj* or *t̄hj*) and collision energies (14 or 27 TeV). To assess the stability of the proposed method against higher order corrections we have extracted the weight function from simulations at NLO in QCD to estimate the systematic uncertainty associated with QCD effects and found that the difference is within 10% of the LO extraction. Finally, we can compare our optimized approach to the naïve $\tilde{\kappa}$ extraction through the measurement of $\mathcal{O}_{\text{naïve}} = \langle \cos \theta_\ell \rangle$, which in turn corresponds to the case where the weight is independent of $\tilde{x}$, i.e., $\mathcal{B}/\mathcal{A} = 1$. We define the signal significance of an observable $\mathcal{O}$ as

$$\text{Sig}(\tilde{\kappa}) = \frac{\mathcal{O}(\tilde{\kappa})}{\sigma_{\mathcal{O}}(\tilde{\kappa})}. \tag{26}$$

In Figure 4 we show the improvement of the significance when the optimal weight function is applied on simulated signal events without showering or reconstruction effects at 14 TeV.
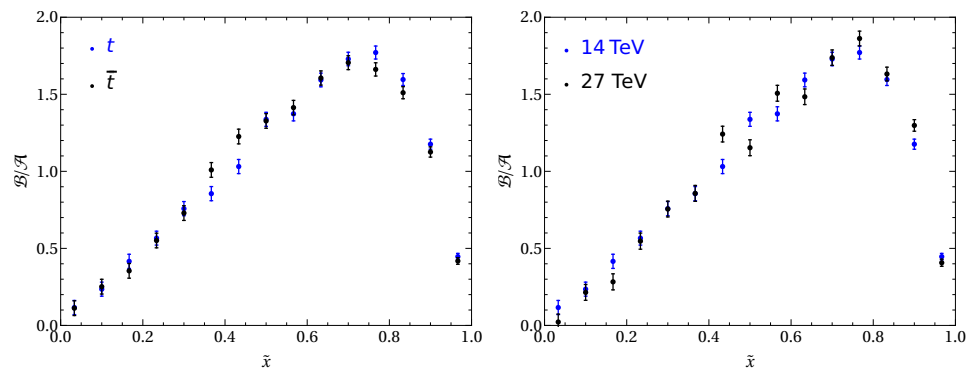
Figure 3. Comparison of the optimal weight $\mathcal{B}/\mathcal{A}$ between the $pp \to thj$ and $pp \to \bar{t}hj$ processes extracted from MC simulations (left). The right panel shows the comparison between 14 and 27 TeV proton collision energies for $pp \to thj$. All plots are obtained using $\tilde{\kappa} = 1$ and with 10^6 MC events [24].
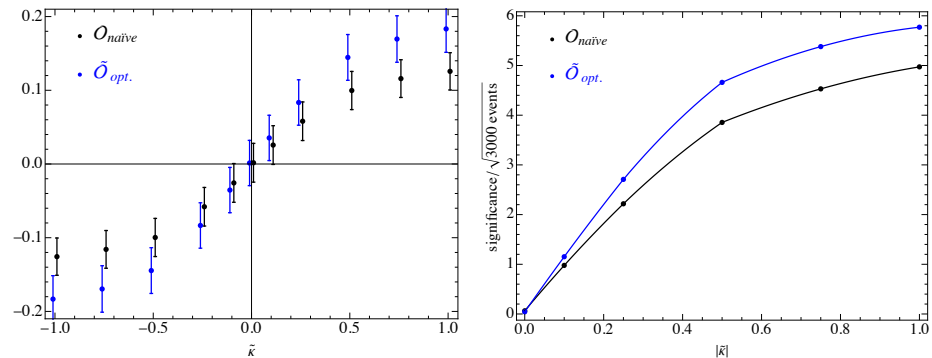
Figure 4. Left: comparison of the optimized spin observable (blue dots) with the naïve observable (black dots) extracted from 3000 $pp \to thj$, $t \to b\ell\nu$ MC events at each choice of $\tilde{\kappa}$. Right: comparison of the significance (defined as the mean value divided by the standard deviation) per $\sqrt{N}$ of the two observables, where N is the number of events [24].

2.3. Limits in the $(\kappa, \tilde{\kappa})$ Plane from $pp \rightarrow thj$ at Event Reconstruction Level

In order to make closer contact with experiments, we can also include the effects of parton showering, detector response and background processes. In our analysis we have used MadGraph5 to generate events at leading order (LO) in QCD for the signal process $pp \rightarrow t(\rightarrow b\ell\nu)h(\rightarrow b\bar{b})j$ plus the conjugate process with $\bar{t}$ at 14 TeV High-Luminosity LHC (HL-LHC) and 27 TeV High-Energy LHC (HE-LHC) center-of-mass energies. (Note that our procedure of obtaining an optimal observable does not depend on the h decay products, therefore this analysis should be taken as a proof of concept with potential for future improvements using e.g., multiple h decay channels) Event generation was performed for multiple values of $(\kappa, \tilde{\kappa})$. The parton level events were subsequently showered and hadronized with Pythia8 [40], and jets are clustered with the anti-k_T algorithm using FastJet [41]. For detector simulation and final state object reconstruction (e.g., lepton isolation and b-tagging) we used Delphes v3.3.3 [42] with the default ATLAS parameters in delphes_card_ATLAS.tcl. The dominant background process in this analysis is $t\bar{t}$ production with additional associated jets. We included this background by generating $pp \rightarrow t\bar{t}$ samples, with one of the tops decayed into the semi-leptonic channel and the other one decayed into the hadronic channel, produced in association with 0, 1, and 2 hard jets. In order to correctly model the hard jets' distributions, we merged the matrix element computations with the MC shower using the MLM [43] prescription. For the event selection we demand the following basic requirements:

- Exactly 3 b-tagged jets with $|\eta(b)| < 5$ and $p_T(b) > 20$ GeV,
- One additional (non-tagged) light jet exclusively in the forward direction with $2 < |\eta(j)| < 5$ and $p_T(j) > 20$ GeV,
- One isolated light lepton $\ell^{\pm} = e^{\pm}, \mu^{\pm}$ with $|\eta(\ell)| < 2.5$ and $p_T(\ell) > 10$ GeV.

In addition, we further select events with one reconstructed Higgs and one reconstructed top quark as follows: first, we calculate the three possible invariant masses from the three reconstructed b-jets (m_{bb}) and only keep the event if at least one bb pair satisfies $|m_H - m_{bb}| < 15$ GeV. For such events, we select as the Higgs decay candidate $h \rightarrow b\bar{b}$ for the pair of b-jets with the invariant mass closest to the Higgs mass. The remaining non-Higgs b-jet is then assumed to come from the top-quark decay. Next, we reconstruct the top-quark by requiring that the combined invariant mass $m_{b\ell\nu}$ of the remaining b-jet, the lepton, and the neutrino (also reconstructed by assuming it to be the unique source of missing energy in the event) to fall inside the mass window of the top-quark defined by $m_t \pm 35$ GeV. In order to further reject the $t\bar{t}$ backgrounds, events with a reconstructed Higgs and top are selected if the combined invariant mass of the b-jets originating from the Higgs and the light jet satisfies the cut $m_{bbj} > 280$ GeV [44]. The final selection efficiency for the thj signal in the SM is 0.32% (0.23%), while for the background it is 0.008% (0.006%) at 14 TeV (27 TeV).

As we fully reconstruct the th system and have access to the lepton momentum from the top decay we have all the necessary information for measuring the optimized spin observable. We use the optimal weight function $\mathcal{B}/\mathcal{A}$ (Figure 3) extracted from the MC simulations to construct a χ^2 with an appropriately weighted signal process. Our results for $pp \rightarrow thj$ generated in the SM are given by the 2σ exclusion limits (shaded blue) shown in Figure 5 for the HE-LHC at a luminosity of 15 ab^{-1}. As can be seen in Equation (23) the observable $\tilde{\mathcal{O}}_{\text{opt.}}$ is normalized to the cross section, which contains terms κ^2, $\tilde{\kappa}^2$, as well as a linear term in κ and a constant term due to second diagram in Figure 1, whereas the numerator is proportional to $\tilde{\kappa}(\kappa + c)$. The behaviour of $\mathcal{O}_{\text{opt}}$ close to the SM point is thus linear in $\tilde{\kappa}$, whereas the cross section has a minimum in κ close to $\kappa = 1$. In the large coupling regime $\tilde{\mathcal{O}}_{\text{opt.}}$ converges to a small value which depends on the direction in which we make the limit $\kappa^2 + \tilde{\kappa}^2 \rightarrow \infty$. The 2σ exclusion has an elliptic shape as shown by the blue contour in Figure 5. We also present the limit (given by the black elliptic contour) assuming a 2σ positive excess above the SM expectation corresponding to a measurement of the optimized spin observable of $\tilde{\mathcal{O}}_{\text{opt.}} = 0.06 \pm 0.03$ whose size and error are statistics-driven. Because of the nature of our observable, the signed fluctuation gives rise to asymmetric

limits in the $\tilde{\kappa}$ direction. In the κ direction the bounds are also not symmetric as $pp \to thj$ production contains linear κ terms. Finally, in order to include background effects, the same statistical analysis would have to be repeated including the $t\bar{t}$ background in the χ^2 fit. However, even with a large background rejection as implemented above, the irreducible background is simply too large and the signal is completely diluted leading to a signal significance of only $S/\sqrt{B} \sim 0.8$ (3.2) at 14 TeV (27 TeV) at a luminosity of 3 ab^{-1} (15 ab^{-1}). This effectively precludes any meaningful extraction of bounds on $\tilde{\kappa}$ from a fit to $\tilde{\mathcal{O}}_{\text{opt}}$. We leave the possibility of further optimizing the cuts in order to reduce the backgrounds or including other Higgs decay channels as a future challenge. In the remainder of this review we instead focus on the related but more abundant process of associated top quark pair and Higgs boson production.
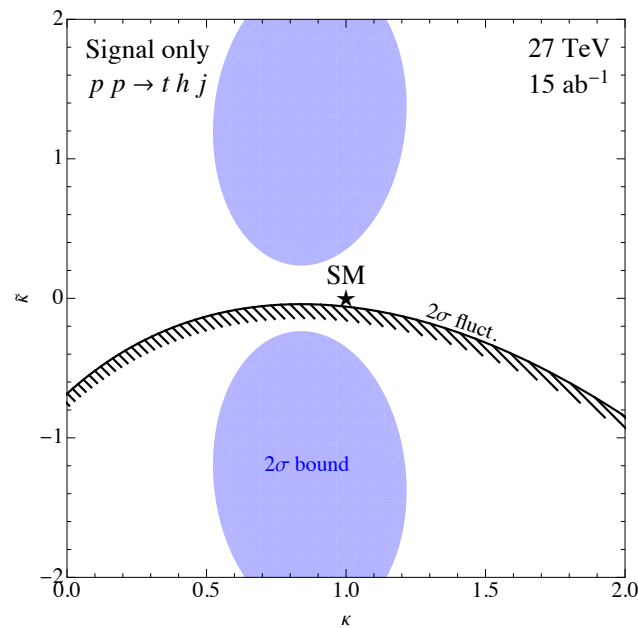
Figure 5. Bounds in the $(\kappa, \tilde{\kappa})$ plane using the optimized observable $\tilde{\mathcal{O}}_{\text{opt}}$ for the single-top associated production with a Higgs boson. The blue shaded region corresponds to the 2σ ($\chi^2 > 6.18$) exclusion zone assuming the measurement of the SM at the HE-LHC (15 ab^{-1}). The black line and stripes shows the 2σ excluded region for a 2σ positive fluctuation at the HE-LHC (see text for details) [24].

3. CP-Odd Observables in the $pp \to t\bar{t}h$ Process

In this Section we introduce laboratory-frame accessible and phase-space optimized *CP*-odd observables in the process $pp \to t\bar{t}h$ in which both top quarks decays semi-leptonically: $t \to b\, l^+\nu$ and $\bar{t} \to \bar{b}\, l^-\bar{\nu}$. This process has a higher S/B ratio compared to $pp \to thj$ and has indeed been measured by the LHC collaborations [45,46] (For the state of the art predictions of the differential distributions see e.g., Ref. [39]) The top quarks in $pp \to t\bar{t}h$ are known to be unpolarized, independent of the value of $\tilde{\kappa}$ [2]. The information on the underlying κ and $\tilde{\kappa}$ parameters is in principle contained in the correlations among the top quark spins, however due to the experimental difficulties of extracting the top quark polarizations in this process, this approach is unfeasible. This is the reason we focus in this Section on manifestly *CP*-odd observables, directly sensitive to $\tilde{\kappa}$ and easily accessible by measurements of final-state momenta in the lab frame.

3.1. Laboratory Frame CP-Odd Observables in $t\bar{t}h$ Production

The accessible final-state momenta [6] of this decay in the lab frame are 3-momenta of b-jet $\boldsymbol{p}_b$, $\bar{b}$-jet $\boldsymbol{p}_{\bar{b}}$, 3-momenta of leptons $\boldsymbol{p}_{\ell^+}$ and $\boldsymbol{p}_{\ell^-}$, and 3-momenta of the Higgs $\boldsymbol{p}_h$. In practice differentiating between b-jet and $\bar{b}$-jet is difficult (For recent attempts in extracting the charge of the b-jet see Refs. [47–50]), therefore we should consider only

observables which are invariant under $p_b \leftrightarrow p_{\bar{b}}$ transformation. We construct these observables from 3-vector quantities with well defined C and P eigenvalues, that are given in Table 1.

Table 1. Vector quantities with well-defined C and P eigenvalues in 3 dimensional Euclidean space. More complicated objects with well-defined C and P eigenvalues can be constructed using variables in this table.

	p_h	$p_{\ell-} + p_{\ell+}$	$p_{\ell-} - p_{\ell+}$	$p_b + p_{\bar{b}}$	$p_b - p_{\bar{b}}$
C	$+$	$+$	$-$	$+$	$-$
P	$-$	$-$	$-$	$-$	$-$
CP	$-$	$-$	$+$	$-$	$+$

We use quantities in Table 1 to construct CP-odd variables ω of mass-dimension 5, that are C-even and invariant under $p_b \leftrightarrow p_{\bar{b}}$ transformation. In order to systematically obtain all distinct ω's we proceed as follows. First, we construct variables of the form $\omega \sim V_1 \times V_2 \cdot V_3 \, V_4 \cdot V_5$ (Notice that the possibility of a nested cross product $(((V_1 \times V_2) \times V_3) \times V_4) \cdot V_5$ can also be reduced to this form) using $V_j \in \{p_h, p_{\ell-} + p_{\ell+}, p_{\ell-} - p_{\ell+}, p_b + p_{\bar{b}}, p_b - p_{\bar{b}}\}$ for $j \in \{1, ..., 5\}$. Doing so, we find 150 potential quintuple products. We symmetrize them with respect to C-conjugation and $p_b \leftrightarrow p_{\bar{b}}$ transformation. The non-zero quintuple products are ω variables, however they may be linearly dependent. Indeed some of the obtained ω's are connected via the following Euclidean identity

$$\delta_{ab}\epsilon_{cde} - \delta_{ac}\epsilon_{deb} + \delta_{ad}\epsilon_{ebc} - \delta_{ae}\epsilon_{bcd} = 0, \tag{27}$$

which can be written as:

$$a\,(b \times c \cdot d) - b\,(c \times d \cdot a) + c\,(d \times a \cdot b) - d\,(a \times b \cdot c) = 0, \tag{28}$$

where a, b, c and d are four arbitrary vectors in 3 dimensional Euclidean space. The sign of individual terms in the last expression corresponds to the sign of the cyclic permutation of the four vectors.

The first class of ω's involves $p_{\ell+}$ and $p_{\ell-}$ in the mixed product, $p_{\ell-} - p_{\ell+}$ in the scalar product. Both products are invariant under $p_b \leftrightarrow p_{\bar{b}}$:

$$\omega_1 \sim [(p_{\ell-} \times p_{\ell+}) \cdot p_h][(p_{\ell-} - p_{\ell+}) \cdot p_h], \tag{29}$$

$$\omega_2 \sim [(p_{\ell-} \times p_{\ell+}) \cdot p_h][(p_{\ell-} - p_{\ell+}) \cdot (p_{\ell-} + p_{\ell+})], \tag{30}$$

$$\omega_3 \sim [(p_{\ell-} \times p_{\ell+}) \cdot p_h][(p_{\ell-} - p_{\ell+}) \cdot (p_b + p_{\bar{b}})], \tag{31}$$

$$\omega_4 \sim [(p_{\ell-} \times p_{\ell+}) \cdot (p_b + p_{\bar{b}})][(p_{\ell-} - p_{\ell+}) \cdot p_h], \tag{32}$$

$$\omega_5 \sim [(p_{\ell-} \times p_{\ell+}) \cdot (p_b + p_{\bar{b}})][(p_{\ell-} - p_{\ell+}) \cdot (p_{\ell-} + p_{\ell+})], \tag{33}$$

$$\omega_6 \sim [(p_{\ell-} \times p_{\ell+}) \cdot (p_b + p_{\bar{b}})][(p_{\ell-} - p_{\ell+}) \cdot (p_b + p_{\bar{b}})]. \tag{34}$$

The second class involves $p_b \times p_{\bar{b}}$ and/or $p_b - p_{\bar{b}}$ in both mixed and scalar products:

$$\omega_7 \sim [(p_b \times p_{\bar{b}}) \cdot p_h][(p_b - p_{\bar{b}}) \cdot p_h], \tag{35}$$

$$\omega_8 \sim [(p_b \times p_{\bar{b}}) \cdot p_h][(p_b - p_{\bar{b}}) \cdot (p_{\ell-} + p_{\ell+})], \tag{36}$$

$$\omega_9 \sim [(p_b \times p_{\bar{b}}) \cdot p_h][(p_b - p_{\bar{b}}) \cdot (p_b + p_{\bar{b}})], \tag{37}$$

$$\omega_{10} \sim [(p_b \times p_{\bar{b}}) \cdot (p_{\ell-} + p_{\ell+})][(p_b - p_{\bar{b}}) \cdot p_h], \tag{38}$$

$$\omega_{11} \sim [(p_b \times p_{\bar{b}}) \cdot (p_{\ell-} + p_{\ell+})][(p_b - p_{\bar{b}}) \cdot (p_{\ell-} + p_{\ell+})], \tag{39}$$

$$\omega_{12} \sim [(p_b \times p_{\bar{b}}) \cdot (p_{\ell-} + p_{\ell+})][(p_b - p_{\bar{b}}) \cdot (p_b + p_{\bar{b}})], \tag{40}$$

$$\omega_{13} \sim [(p_b \times p_{\bar{b}}) \cdot (p_{\ell-} - p_{\ell+})][(p_b - p_{\bar{b}}) \cdot (p_{\ell-} - p_{\ell+})], \tag{41}$$

$$\omega_{14} \sim [(p_{\ell-} \times p_{\ell+}) \cdot (p_b - p_{\bar{b}})][(p_b - p_{\bar{b}}) \cdot (p_{\ell-} - p_{\ell+})]. \tag{42}$$

The third class involves mixed product of $\boldsymbol{p}_h$, $\boldsymbol{p}_{\ell^-} \pm \boldsymbol{p}_{\ell^+}$, and $\boldsymbol{p}_b \pm \boldsymbol{p}_{\bar{b}}$:

$$\omega_{15} \sim [\boldsymbol{p}_h \times (\boldsymbol{p}_{\ell^-} + \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b - \boldsymbol{p}_{\bar{b}})][(\boldsymbol{p}_b - \boldsymbol{p}_{\bar{b}}) \cdot \boldsymbol{p}_h], \tag{43}$$

$$\omega_{16} \sim [\boldsymbol{p}_h \times (\boldsymbol{p}_{\ell^-} + \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b - \boldsymbol{p}_{\bar{b}})][(\boldsymbol{p}_b - \boldsymbol{p}_{\bar{b}}) \cdot (\boldsymbol{p}_{\ell^-} + \boldsymbol{p}_{\ell^+})], \tag{44}$$

$$\omega_{17} \sim [\boldsymbol{p}_h \times (\boldsymbol{p}_{\ell^-} + \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b - \boldsymbol{p}_{\bar{b}})][(\boldsymbol{p}_b - \boldsymbol{p}_{\bar{b}}) \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}})], \tag{45}$$

$$\omega_{18} \sim [\boldsymbol{p}_h \times (\boldsymbol{p}_{\ell^-} - \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}})][(\boldsymbol{p}_{\ell^-} - \boldsymbol{p}_{\ell^+}) \cdot \boldsymbol{p}_h], \tag{46}$$

$$\omega_{19} \sim [\boldsymbol{p}_h \times (\boldsymbol{p}_{\ell^-} - \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}})][(\boldsymbol{p}_{\ell^-} - \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_{\ell^-} + \boldsymbol{p}_{\ell^+})], \tag{47}$$

$$\omega_{20} \sim [\boldsymbol{p}_h \times (\boldsymbol{p}_{\ell^-} - \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}})][(\boldsymbol{p}_{\ell^-} - \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}})], \tag{48}$$

$$\omega_{21} \sim [\boldsymbol{p}_h \times (\boldsymbol{p}_{\ell^-} - \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b - \boldsymbol{p}_{\bar{b}})][(\boldsymbol{p}_{\ell^-} - \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b - \boldsymbol{p}_{\bar{b}})]. \tag{49}$$

There are further possibilities with a mixed product $\boldsymbol{p}_h \times (\boldsymbol{p}_{\ell^-} + \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}})$ multiplied by one of 8 C-even, $b \leftrightarrow \bar{b}$ even scalar products $\{\boldsymbol{p}_h \cdot \boldsymbol{p}_h, \boldsymbol{p}_h \cdot (\boldsymbol{p}_{\ell^-} + \boldsymbol{p}_{\ell^+}), \boldsymbol{p}_h \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}}), (\boldsymbol{p}_{\ell^-} + \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}}), (\boldsymbol{p}_{\ell^-} \pm \boldsymbol{p}_{\ell^+})^2, (\boldsymbol{p}_b \pm \boldsymbol{p}_{\bar{b}})^2\}$. Since the mixed product itself already has the desired symmetry properties, those 8 quintuple products do not bring additional new information, with respect to a mixed (triple) product that is our final observable:

$$\omega_{22} \sim \boldsymbol{p}_h \times (\boldsymbol{p}_{\ell^-} + \boldsymbol{p}_{\ell^+}) \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}}). \tag{50}$$

Note that there are nonlinear relations between ω's, such as $\omega_1 \omega_6 = \omega_3 \omega_4$, that we do not exploit to further reduce the set. Namely, ratios of ω's can contain singularities in the available phase-space and as such would be difficult to reconstruct by a neural network optimizer.

All the ω's are normalized by the lengths of the vectors that enter as factors in the scalar products,

$$\omega_i = \frac{[\boldsymbol{V}_1 \times \boldsymbol{V}_2 \cdot \boldsymbol{V}_3][\boldsymbol{V}_4 \cdot \boldsymbol{V}_5]}{|\boldsymbol{V}_1 \times \boldsymbol{V}_2||\boldsymbol{V}_3||\boldsymbol{V}_4||\boldsymbol{V}_5|}, \tag{51}$$

and the upper bound $|\omega_i| \leq 1$ is generally valid. For cases when ω_i has a vector $\boldsymbol{a}$ present both in the mixed and scalar products, e.g., $(\boldsymbol{V}_1 \times \boldsymbol{V}_2 \cdot \boldsymbol{a})(\boldsymbol{V}_3 \cdot \boldsymbol{a})$, and furthermore with $\boldsymbol{V}_1 \times \boldsymbol{V}_2 \cdot \boldsymbol{V}_3 = 0$, a stricter upper bound $|\omega_i| \leq 1/2$ applies (for $\omega_{1,6,7,11,13,14,15,16,20,21}$).

Having constructed manifestly CP-odd variables ω, we now show how they can be used to extract information on $\tilde{\kappa}$ in an optimal way. The $t\bar{t}h$ production cross section can be written as

$$\frac{d\sigma}{d\boldsymbol{x} d\omega_i} = A(\boldsymbol{x}, |\omega_i|) + \tilde{\kappa}\kappa B(\boldsymbol{x}, \omega_i). \tag{52}$$

where $\boldsymbol{x}$ are CP-even phase space variables, A is a manifestly CP-even and B is a manifestly CP-odd function of ω_i: $B(\boldsymbol{x}, \omega_i) = -B(\boldsymbol{x}, -\omega_i)$. The $\kappa\tilde{\kappa}$ dependence is due to the interference of scalar and pseudoscalar amplitudes. A simple CP-odd observable is an average of a single variable ω_i

$$\mathcal{O}_i = \frac{1}{\sigma} \int d\boldsymbol{x} \, d\omega_i \frac{d\sigma}{d\boldsymbol{x} d\omega_i} \omega_i = \frac{1}{N} \sum_{j=1}^{N} \omega_i^{(j)}, \tag{53}$$

where N is the number of experimental events. The standard deviation σ_i of such an observable is given by

$$\sigma_i^2 = \frac{1}{N}\left[\frac{1}{N}\sum_j (\omega_i^{(j)})^2 - \mathcal{O}_i^2\right]. \tag{54}$$

For large enough N the distribution of $\mathcal{O}_i$ is approximately Gaussian and the corresponding significance of this observable is:

$$\text{Sig}_i \equiv \frac{\mathcal{O}_i}{\sigma_i} = \sqrt{N}\frac{\mathcal{O}_i}{\sqrt{\frac{1}{N}\sum_j (\omega_i^{(j)})^2 - \mathcal{O}_i^2}}. \tag{55}$$

Equation (55) holds for $\omega_1, ..., \omega_{22}$ and their linear combinations. By studying the behavior of all 22 such observables $\mathcal{O}_i$ on MC simulations of $pp \to t\bar{t}h$ with semi-leptonically

decaying top quarks we have found that $\mathcal{O}_6$ and $\mathcal{O}_{14}$ are the most promising in terms of their significance. Our next aim is to show how one can achieve better sensitivity to $\tilde{\kappa}$ by introducing optimized observables in two directions: the phase-space optimization of a single ω_i and a construction of an optimal *CP*-odd observable as a combination of all available ωs.

3.2. NN-Based Optimized CP-Odd Observables

Due to the high dimensionality and complexity of the phase-space in the $t\bar{t}h$ process with top quarks decaying semileptonically, 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 note that existing general purpose ML inference tools are already able to optimize sensitivity to a given parameter (see e.g., Ref. [51]). The purpose of our work was to do so in an economic way that manifestly respects the symmetries of the problem.

We implemented the training and evaluation of neural networks using the `TensorFlow` framework [52]. In all cases, we used a sample of 10^7 $pp \to ht(\to b\ell^+\nu)\bar{t}(\to \bar{b}\ell^-\bar{\nu})$ events generated at LO using `Madgraph5` [35] together with the Higgs Characterisation UFO model [37,38] with $\kappa, \tilde{\kappa} = 1$. We split the sample into separate training (7.5 M) and test (2.5 M) samples. After training at fixed $\tilde{\kappa} = 1$ we also tested the observables at other values of $\tilde{\kappa}$ and κ, both ranging from -1 to 1. In these tests 1 M events have been used.Unless stated otherwise the results are shown for events in pp collisions at 14 TeV. We randomly initialized the neural network weights using the default Glorot uniform initializer and used the Adam optimizer with a custom varying learning rate $l(e) = l(e-1)/(1+0.8^e)$ where e is the current epoch and the initial learning rate is set to 0.1. We use `relu` for the activation function. We trained all networks using the novel loss function

$$\text{loss}(\boldsymbol{\alpha}) = \left(\frac{\text{mean}(\mathcal{F}(\boldsymbol{X};\boldsymbol{\alpha}))}{\text{std}(\mathcal{F}(\boldsymbol{X};\boldsymbol{\alpha}))/\sqrt{N}} \right)^{-2}, \tag{56}$$

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 mean($\mathcal{F}(\boldsymbol{X};\boldsymbol{\alpha})$) that should be minimized in order to achieve optimal statistical sensitivity. Here N is the size of the event sample, $\boldsymbol{\alpha}$ are the free neural network weights and biases and $\boldsymbol{X}$ stands for the values of *CP*-even and/or *CP*-odd phase-space variables in the given event. We emphasize that such a choice of the loss function is unique to the problem at hand - we are using the optimization procedure to directly maximize the significance of each considered observable. We can 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 found 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 performed a scan over a set of possible NN configurations with up to 2 hidden layers and up to 9 nodes per NN layer. (We have also considered an automated algorithm to determine the optimal NN architecture (i.e., Hyperopt [53], see also Ref. [54] for one of its recent uses.). Here instead we present results of manual scans over a set of possible NN configurations in order to have better control over the NN parameters. We found the results of both approaches comparable. We choose this cutoff for representational purposes, however we have checked that our results do not change significantly when using larger networks, namely up to three hidden layers of 30 nodes each.)

3.2.1. Phase-Space Optimization of a Single ω

First we present the optimization of the ω_6 and ω_{14} variables based on phase-space averaging. We do not follow the optimization procedure based on separating A and B in Equation (52) since this would require cumbersome multidimensional binning [33]. We use a vector of easily accessible *CP*-even Mandelstam variables x:

$$x = \begin{pmatrix} (\boldsymbol{p}_{\ell^+} + \boldsymbol{p}_{\ell^-}) \cdot \boldsymbol{p}_h \\ (\boldsymbol{p}_{\ell^+} + \boldsymbol{p}_{\ell^-}) \cdot (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}}) \\ (\boldsymbol{p}_b + \boldsymbol{p}_{\bar{b}}) \cdot \boldsymbol{p}_h \\ \boldsymbol{p}_{\ell^+} \cdot \boldsymbol{p}_{\ell^-} \\ \boldsymbol{p}_b \cdot \boldsymbol{p}_{\bar{b}} \end{pmatrix}. \tag{57}$$

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

$$f(x; \alpha) = C(x; \alpha) + \tilde{\kappa}^2 D(x; \alpha) + \mathcal{O}(\tilde{\kappa}^4). \tag{58}$$

Using (52) we can now express the observable

$$\begin{aligned} \langle f(x; \alpha)\omega_i \rangle &= \frac{1}{\sigma} \int \frac{d\sigma}{dxd\omega_i} f(x; \alpha)\omega_i \, dxd\omega_i \\ &= \frac{\tilde{\kappa}\kappa}{\sigma} \int B(x, \omega_i)C(x; \alpha)\omega_i \, dxd\omega_i + \frac{\tilde{\kappa}^3\kappa}{\sigma} \int B(x, \omega_i)D(x; \alpha)\omega_i \, dxd\omega_i + \mathcal{O}(\tilde{\kappa}^5), \end{aligned} \tag{59}$$

with the usual definition of the average (The phase space average of a function is defined as $\langle \# \rangle \equiv \frac{1}{\sigma} \int \frac{d\sigma}{dxd\omega} \# \, dxd\omega$). The integration region in $x - \omega_i$ space is symmetric with respect to the transformation $\omega_i \to -\omega_i$, therefore integrals of the arguments which are antisymmetric in ω_i vanish. Due to this reason all contributions to the expected value of the observable $\langle f(x; \alpha)\omega_i \rangle$ are proportional to odd powers of $\tilde{\kappa}$. 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 (56) we have $\mathcal{F}(x, \omega_i; \alpha) = f(x; \alpha) \omega_i$.

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 studied 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 Figure 6 in the purple box plot for the case of ω_6 and the orange box plot for the case of ω_{14}. The plain ω_6-based observable is shown in gray, with the dashed lines denoting its 1σ statistical uncertainty, while the same holds true for plain ω_{14} in black. We found that the phase-space optimization of ω_6 gives a noticeable improvement over plain ω_6 when using a large enough network. Interestingly, ω_{14} seems to be close to optimal on its own, as the phase space optimization does not introduce noticeable improvement. Moreover, the optimized ω_6 gives a similar performance to ω_{14}, hinting that we have reached maximal performance achievable with a single ω.

To test how well the resulting networks generalize to other values of $\tilde{\kappa}$ we used the 50 converged $\{9, 9\}$ models to calculate the dependence of the resulting observable significances with respect to $\tilde{\kappa}$ on the aforementioned fresh sets of 1 M events per $\tilde{\kappa}$. This is shown on Figure 7 where a consistent improvement over simple $\langle \omega_6 \rangle$ can be seen at all considered $\tilde{\kappa}$, while again a marginal improvement is confirmed for ω_{14}, with the optimized ω_6 hovering around ω_{14}.
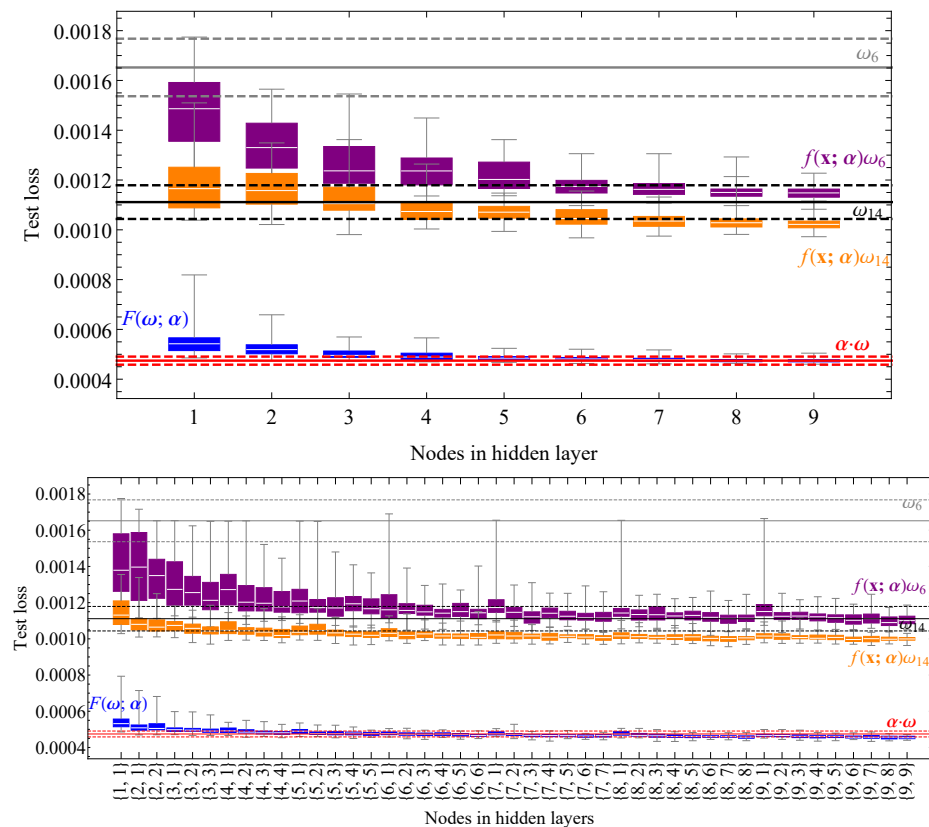
Figure 6. A scan in terms of the test loss (sample size 2.5M, $\kappa, \tilde{\kappa} = 1$) over neural network configurations with one (upper plot) or two (lower plot) hidden layers for the phase-space optimized ω_6 and ω_{14} (59) shown in the purple and orange box plot and the generalized $F(\boldsymbol{\omega}; \boldsymbol{\alpha})$ (Section 3.2.2) shown in the blue box plot. The spread in all cases corresponds to 50 different random weight initializations per configuration. For comparison the plain ω_6 and ω_{14} are shown in gray and black with the dashed lines showing their 1σ statistical uncertainty. The first order approximation of $F(\boldsymbol{\omega}; \boldsymbol{\alpha})$, defined in Equation (60), is shown in red as described in Section 3.2.3 [27].

As the results of optimizing single ω_6 and ω_{14} point to a maximal performance possible using a single ω and the chosen set of phase space variables, we now turn to the rest of the ω's. In the next subsection we consider a more general case where the *CP*-odd observable itself is parameterized with a neural network.

3.2.2. Neural Network as a *CP*-Odd Observable

We consider the case where the output of the neural network is a *CP*-odd quantity that defines our observable. To this end, we build a network with 22 inputs, one per each ω_i, and one output $F(\boldsymbol{\omega}; \boldsymbol{\alpha})$, which is correctly anti-symmetrized so that $F(\boldsymbol{\omega}; \boldsymbol{\alpha}) = -F(-\boldsymbol{\omega}; \boldsymbol{\alpha})$. In terms of the loss function (56) we have $\mathcal{F}(\boldsymbol{X}; \boldsymbol{\alpha}) = F(\boldsymbol{\omega}; \boldsymbol{\alpha})$. Note that since we include the complete irreducible set $\boldsymbol{\omega}$ in this non-linear construction, it effectively also covers the case of a simple phase-space optimization of any (linear combination of) ω_i, since all relevant *CP*-even phase-space variables can be recovered by taking suitable products of $\omega_i \omega_{i'}$.

We again carried 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 scanned 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 ran 50 trainings with different random weight initializations. The results are shown in Figure 6 in the blue box

plot. We find a considerable improvement over the phase-space optimizations of ω_6 or ω_{14}. The improvement is consistent in the entire range of $\tilde{\kappa} \in [0.1, 1.0]$ and is most striking at large $\tilde{\kappa}$.

Again we can 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 Figure 7. We find a consistent improvement over the previous case across all considered $\tilde{\kappa}$. A noticeable improvement in the significance can be seen. In order to better understand the physics underlying the optimization, we next consider this model in the leading order approximation in ω.

3.2.3. First Order Approximation of $F(\omega; \alpha)$

To address the arbitrariness of the neural network architecture choice and to better understand the underlying physics, we finally consider the first order approximation of the function $F(\omega; \alpha)$, which can be expanded in a Taylor series $F(\omega; \alpha) = \sum_j \alpha_j \omega_j + \mathcal{O}(\omega^3)$ for $j \in \{1, \ldots, 22\}$, since for most events the values of the *CP*-odd variables are small: $|\omega_i| \ll 1$, $i \in \{1, ..., 22\}$. In the first order approximation the optimal *CP*-odd observable can be written as

$$\mathcal{O}_{\alpha \cdot \omega} = \left\langle \sum_j \alpha_j \omega_j \right\rangle, \tag{60}$$

where α_j are parameters that satisfy a subsidiary condition $|\alpha| \equiv \sum_{j=1}^{22} \alpha_j^2 = 1$. Their values can be found by maximizing the significance:

$$\frac{\partial}{\partial \alpha_k} \frac{\mathcal{O}_{\alpha \cdot \omega}}{\text{std}(\alpha \cdot \omega)} = 0, \tag{61}$$

The left side of the expression above can be expanded in the following way:

$$\frac{\partial}{\partial \alpha_k} \frac{\mathcal{O}_{\alpha \cdot \omega}}{\text{std}(\alpha \cdot \omega)} = \frac{2 \langle \alpha \cdot \omega \rangle}{\text{std}(\alpha \cdot \omega)} \left\{ \langle \omega_k \rangle \langle (\alpha \cdot \omega)^2 \rangle - \langle (\alpha \cdot \omega) \omega_k \rangle \right\}$$

$$= \frac{2 \langle \alpha \cdot \omega \rangle}{\text{std}(\alpha \cdot \omega)} \left\{ \sum_{i,j=1}^{22} \alpha_i \alpha_j \left[\langle \omega_i \omega_k \rangle \langle \omega_j \rangle - \langle \omega_i \omega_j \rangle \langle \omega_k \rangle \right] \right\}. \tag{62}$$

where in the last step we take α outside the $\langle - \rangle$, for example $\langle \alpha \cdot \omega \rangle = \sum_j \alpha_j \langle \omega_j \rangle$. By assuming $\tilde{\kappa} \neq 0$ we have $\mathcal{O}_{\alpha \cdot \omega} \neq 0$. From this, together with the condition in Equation (61), it follows that the expression in the curly brackets of Equation (62) must be equal to 0. Hence, we obtain a system of 22 quadratic equations (The problem is equivalent to a single neuron NN with 22 inputs and one output without the activation function and the bias term)

$$\alpha^T M^{(k)} \alpha = 0, \tag{63}$$

where $\alpha = [\alpha_1, ..., \alpha_{22}]^T$ and the 22×22 matrices $M^{(k)}$ are given by (Notice that by definition $M_{ik}^{(k)} = 0$ for all i and k, therefore $\det(M^{(k)}) = 0$ for all k)

$$M_{ij}^{(k)} = \langle \omega_i \omega_k \rangle \langle \omega_j \rangle - \langle \omega_i \omega_j \rangle \langle \omega_k \rangle. \tag{64}$$

The system of equations in Equation (63) can be solved numerically. The solution vector α is undetermined up to a real non-zero constant. We normalize it such that $\alpha \cdot \alpha = 1$ and choose the solution with mostly positive values.
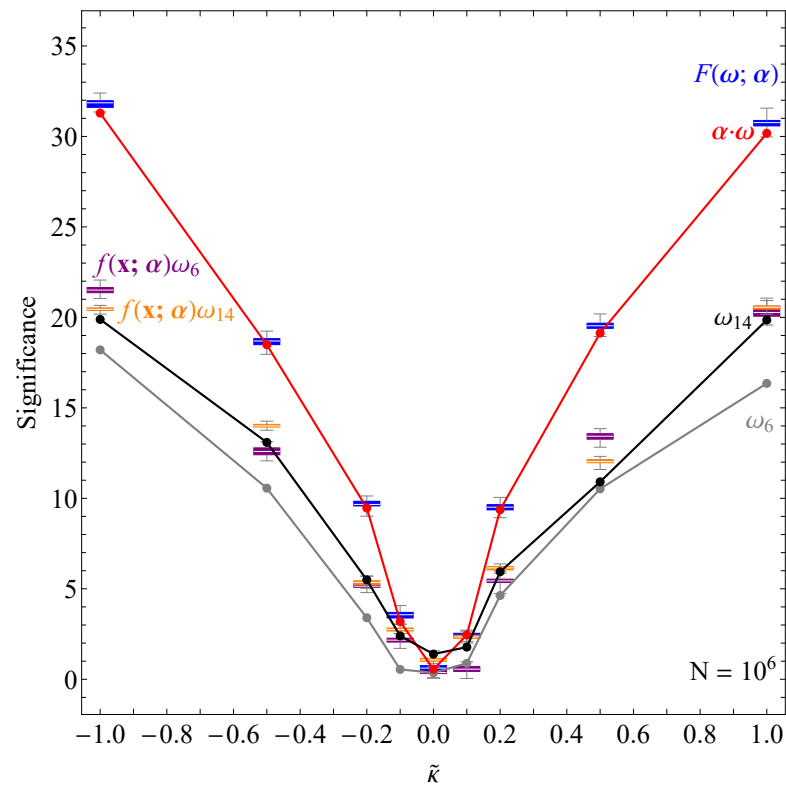
Figure 7. Comparison of the significances (defined as the mean value divided by the standard deviation) of all the observables considered in this work with respect to $\tilde{\kappa}$ and at fixed $\kappa = 1$. The results correspond to 1M events per $\tilde{\kappa}$ at 14 TeV. Plain ω_6 in gray, ω_{14} in black, phase-space optimized ω_6 and ω_{14} (59) in purple and orange, anti-symmetrized neural network $F(\omega; \alpha)$ (Section 3.2.2) in blue and the first order approximation of the latter $\alpha \cdot \omega$ in red (Section 3.2.3). See text for details on each observable [27].

We used this approach to extract the optimal weights α_j from 10^7 events generated with $\tilde{\kappa} = \kappa = 1$ at 14, 27, and 100 TeV. We also estimated the uncertainty associated with the optimal weights using the following procedure: First we estimate the statistical spread of the significance obtained with optimal α. Next we allow a single α_j to float in the intervals $[\alpha_j - \sigma_j, \alpha_j + \sigma_j]$, where σ_j is chosen such that the decrease of the significance due to the change in α_j corresponds to the statistical spread of the significance. We perform an efficient scan around the optimal vector α in its 22-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 α-space. With this procedure we determine how sharply the optimal α_j are defined. In practice, we estimated the statistical error of the significance using 10^7 events. Clearly the uncertainties σ_j are larger for smaller chosen sample size. The results of this approach are shown in Figure 8, where the upper (lower) panel shows the estimated error (significance) for each α_j at 14, 27, and 100 TeV. A comparison of the observable $\alpha \cdot \omega$ to other approaches discussed previously is shown in Figure 7. We reach a similar level of improvement compared to the full $F(\omega; \alpha)$ network with significantly fewer parameters.
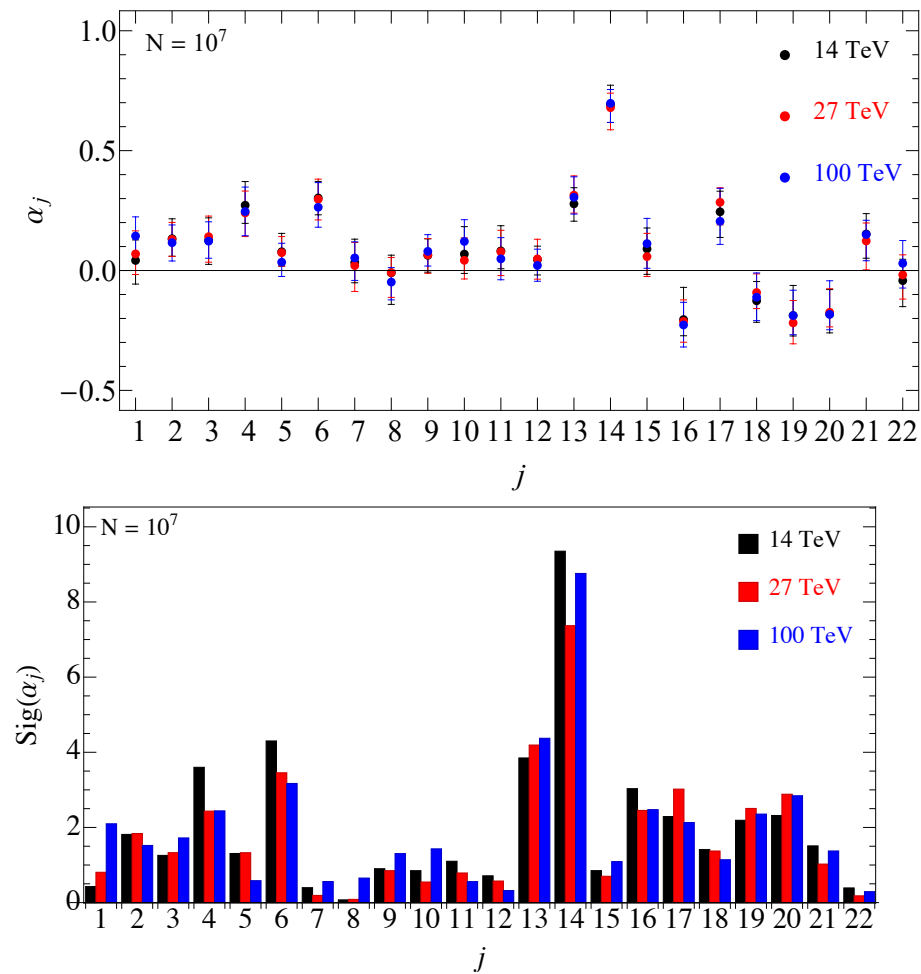
Figure 8. Optimal weights of the linear observable defined in Equation (60). The upper plot shows uncertainties of α_j, estimated using the expected statistical errors of the observable significances, see text for details. The lower plot shows the significances of α_j (defined as their central value divided by their estimated uncertainty) [27].

### 3.3. Bounds in the $(\kappa, \tilde{\kappa})$ Plane

We could now produce the bounds in the $(\kappa, \tilde{\kappa})$ plane by varying both κ and $\tilde{\kappa}$ at generator level and 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 $h \to b\bar{b}$ decay, our signal is defined as 4 b-jets and two oppositely charged leptons ℓ. We included the main irreducible background $pp \to t\bar{t}b\bar{b}$ with both tops decaying semileptonically and used the event selection requirements:

- ≥ 4 jets of any flavor with $|\eta(j)| < 5$ and $p_T(j) > 20$ GeV.
- ≥ 3 of the above jets are b-tagged.
- 2 oppositely charged light leptons with $|\eta(\ell)| < 2.5$ and $p_T(\ell) > 10$ GeV.

Furthermore, in order to identify the b-jets from t, $\bar{t}$ decays we counted the number N_b of tagged b-jets and performed the following selections: if $N_b \geq 4$, we compute the invariant masses m_{bb} of all possible b-jet pairs and select the pair with invariant mass closest to the Higgs mass $m_h = 125$ GeV. If the selected pair falls inside the Higgs mass window ($|m_{bb} - m_h| < 15$ GeV) we removed the pair from the list of b-jets and selected from this list the highest p_T b-jets as our candidate top quark decay b-jets. However, if $N_b = 3$ we computed all possible invariant masses m_{bj} where j are non-b jets in the event. We selected as the $h \to b\bar{b}$ candidate the bj pair that minimizes $|m_h - m_{bj}|$ and falls inside the Higgs

mass window $m_h \pm 15$ GeV. The remaining two *b*-jets are taken as the candidate top quark decay *b*-jets.

We present the bounds for HL- and HE-LHC, FCC-hh by using the optimal observable (60) with the weights shown in the upper plot of Figure 8. The bounds coming from a null result up to the expected statistical uncertainty for different luminosities at different energies are shown on Figure 9, where we also show the expected sensitivity to $(\kappa, \tilde{\kappa})$ from the $t\bar{t}h$ production cross-section measurements using the projected uncertainties of $\delta\kappa/\kappa \sim (0.04, 0.02, 0.009)$ at HL-LHC, HE-LHC [55] and FCC-hh [56], respectively. For direct comparison with [24] we also show the expected bounds from using a single ω_6 and have checked that using ω_{14} does not change the single ω bounds significantly.
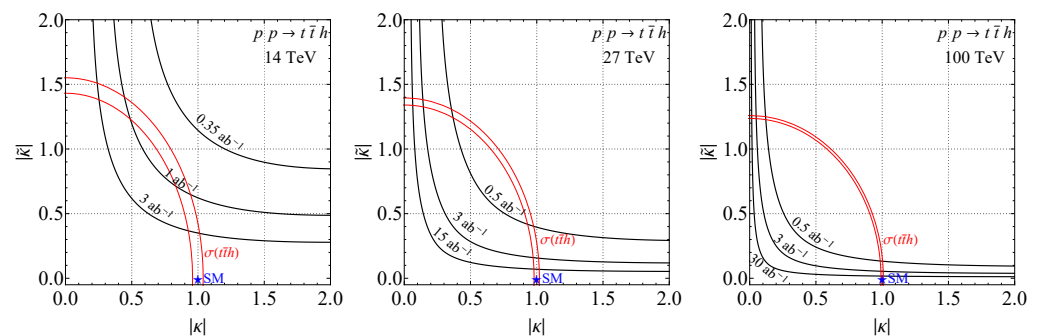
Figure 9. The 2σ exclusion zones in the $\kappa - \tilde{\kappa}$ plane using the optimized observable $\boldsymbol{\alpha} \cdot \boldsymbol{\omega}$ by assuming a null result at HL-LHC, HE-LHC and FCC-hh for different luminosities (left: 14 TeV; middel: 27 TeV; right: 100 TeV). The projected sensitivity of $t\bar{t}h$ production cross-section measurements is shown in red, see text for details. At 14 TeV $\mathcal{O}(1)$ exclusion of $\tilde{\kappa}$ can be achieved using $\boldsymbol{\alpha} \cdot \boldsymbol{\omega}$ with 350 fb^{-1} which corresponds to the final integrated luminosity of the LHC [27].

A consistent improvement of sensitivity compared to using single ω_i can be achieved by using the optimized combination of ω's, constraining the parameter space significantly in orthogonal directions compared to the cross-section measurements. Interestingly the significance improvement is consistent between partonic events and after including shower, detector effects, as well as the dominant background and realistic object reconstruction, 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. It is also reassuring that the optimization does not significantly rely on specific phase-space regions particularly affected by the background. We expect the results to be also robust against higher order QCD corrections as in the case of $pp \rightarrow thj$, where we have checked this explicitly.

We show the sensitivity of the optimized observable to the sign of $\tilde{\kappa}$ (and κ) in Figure 10 by assuming the measurement of a 2σ positive statistical fluctuation of the SM case, which in our estimate corresponds to the measurement of $\boldsymbol{\alpha} \cdot \boldsymbol{\omega} = (4.6 \pm 2.3) \times 10^{-4}$, $\boldsymbol{\alpha} \cdot \boldsymbol{\omega} = (0.9 \pm 0.45) \times 10^{-4}$ and $\boldsymbol{\alpha} \cdot \boldsymbol{\omega} = (0.2 \pm 0.1) \times 10^{-4}$ for HL-LHC (3 ab^{-1}), HE-LHC (15 ab^{-1}) and FCC-hh (30 ab^{-1}) respectively.
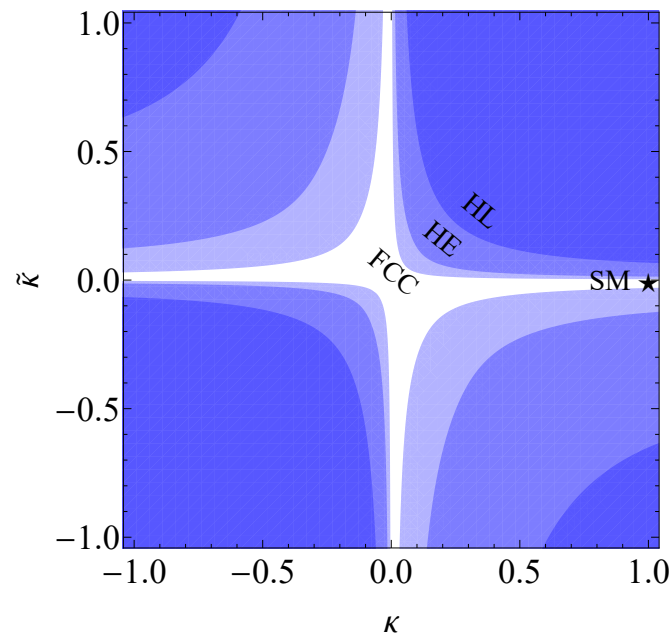
Figure 10. The 2σ exclusion regions at HL-LHC ($3\ \text{ab}^{-1}$), HE-LHC ($15\ \text{ab}^{-1}$) and FCC-hh ($30\ \text{ab}^{-1}$) by assuming a measurement of a 2σ positive fluctuation in the optimal observable $\boldsymbol{\alpha} \cdot \boldsymbol{\omega}$ (60) [27].

4. Conclusions

In order to establish, directly and with minimal additional assumptions, the presence of a *CP*-odd component of the top quark Yukawa ($\tilde{\kappa}$), we have studied manifestly *CP*-odd observables in th and $t\bar{t}h$ production at the LHC and its prospective upgrades.

For the thj final states we have relied on the possibility of reconstructing the t quark momentum and accessing the t polarization. We have identified a particular polarization direction which is perpendicular to the th plane, where the top polarization along this direction would undoubtedly point to the presence of the *CP*-odd coupling $\tilde{\kappa}$. We have presented a method for optimizing the phase space dependent weight and shown its sensitivity at the HL- and HE-LHC for the semileptonic top and $h \to b\bar{b}$ mode. The handful of signal events offer discriminating power, sensitive to the sign of $\tilde{\kappa}$, however the irreducible background due to $t\bar{t}$+jets severely dilutes the sensitivity of the proposed observable.

On the other hand, $t\bar{t}h$ production has a considerably larger cross section at LHC energies compared to thj, while suffering more moderately from irreducible backgrounds. Due to the complexity of the final state kinematics with multiple undetected particles we have in this case proposed variables ω_i that only depend on the lab-frame accessible momenta and are manifestly *P*- and *CP*-odd. Furthermore, we have studied the prospect of their phase-space optimization, parameterizing the optimal weight functions with neural networks. In particular, we have studied a general *CP*-odd observable, parameterized directly by a *CP* anti-symmetric neural network, which results in better performance compared to any individual ω_i. Finally, we have studied the first order approximation of this network as a linear combination of the *CP*-odd observables, producing a simpler and more robust observable, $\boldsymbol{\alpha} \cdot \boldsymbol{\omega}$. One benefit of using $\boldsymbol{\alpha} \cdot \boldsymbol{\omega}$ optimized at parton level is that it retains close to optimal sensitivity to the *CP*-odd coupling even after detector simulation, event selection and reconstruction, allowing to probe $\tilde{\kappa}$ directly at HL-LHC, HE-LHC and FCC-hh. We have found that, at the end of Run 3, the LHC will exclude $\kappa\tilde{\kappa} \sim 1$ with 2σ confidence, while FCC-hh will be sensitive to $\kappa\tilde{\kappa} \sim 0.01$. Note that these observables represent highly complementary probes of the top Yukawa sector compared to $pp \to t\bar{t}h$ cross-section measurements. In particular in their optimized form they would allow to break the degeneracy in the $(\kappa, \tilde{\kappa})$ plane and significantly reduce the allowed parameter space even at modest sensitivities accessible at the (HL)LHC.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aguilar-Saavedra, J.A. A Minimal set of top-Higgs anomalous couplings. *Nucl. Phys.* **2009**, *B821*, 215–227. doi:10.1016/j.nuclphysb.2009.06.022. [CrossRef]
2. Ellis, J.; Hwang, D.S.; Sakurai, K.; Takeuchi, M. Disentangling Higgs-Top Couplings in Associated Production. *J. High Energy Phys.* **2014**, *2014* , 4. [CrossRef]
3. Aad, G.; Abbott, B.; Abdallah, J.; Abdinov, O.; Abeloos, B.; Aben, R.; AbouZeid, O.S.; Abraham, N.L.; Abramowicz, H.; Abreu, H.; et al. 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. *J. High Energy Phys.* **2016**, *8*, 45. [CrossRef]
4. Bhattacharyya, G.; Das, D.; Pal, P.B. Modified Higgs couplings and unitarity violation. *Phys. Rev. D* **2013**, *87*, 011702. [CrossRef]
5. Brod, J.; Haisch, U.; Zupan, J. Constraints on CP-violating Higgs couplings to the third generation. *J. High Energy Phys.* **2013**, *11*, 180. [CrossRef]
6. Boudjema, F.; Godbole, R.M.; Guadagnoli, D.; Mohan, K.A. Lab-frame observables for probing the top-Higgs interaction. *Phys. Rev. D* **2015**, *92*, 015019. [CrossRef]
7. Grzadkowski, B.; Gunion, J.F. Using decay angle correlations to detect CP violation in the neutral Higgs sector. *Phys. Lett. B* **1995**, *350*, 218–224. [CrossRef]
8. Demartin, F.; Maltoni, F.; Mawatari, K.; Page, B.; Zaro, M. Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction. *Eur. Phys. J. C* **2014**, *74*, 3065. [CrossRef]
9. Buckley, M.R.; Goncalves, D. Boosting the Direct CP Measurement of the Higgs-Top Coupling. *Phys. Rev. Lett.* **2016**, *116*, 091801. [CrossRef]
10. Mileo, N.; Kiers, K.; Szynkman, A.; Crane, D.; Gegner, E. Pseudoscalar top-Higgs coupling: Exploration of CP-odd observables to resolve the sign ambiguity. *J. High Energy Phys.* **2016**, *7*, 056. [CrossRef]
11. Gritsan, A.V.; Röntsch, R.; Schulze, M.; Xiao, M. Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques. *Phys. Rev. D* **2016**, *94*, 055023. [CrossRef]
12. Li, J.; Si, Z.g.; Wu, L.; Yue, J. Central-edge asymmetry as a probe of Higgs-top coupling in $t\bar{t}h$ production at the LHC. *Phys. Lett. B* **2018**, *779*, 72–76. [CrossRef]
13. dos Santos, S.A.; Fiolhais, M.C.N.; Frederix, R.; Gonçalo, R.; Gouveia, E.; Martins, R.; Onofre, A.; Pease, C.; Peixoto, H.; Reigoto, A.; et al. Probing the CP nature of the Higgs coupling in $t\bar{t}h$ events at the LHC. *Phys. Rev. D* **2017**, *96*, 013004. [CrossRef]
14. Gonçalves, D.; Kong, K.; Kim, J.H. Probing the top-Higgs Yukawa CP structure in dileptonic $t\bar{t}h$ with M_2-assisted reconstruction. *J. High Energy Phys.* **2018**, *6*, 079. [CrossRef]
15. Kobakhidze, A.; Wu, L.; Yue, J. Anomalous Top-Higgs Couplings and Top Polarisation in Single Top and Higgs Associated Production at the LHC. *J. High Energy Phys.* **2014**, *10*, 100. [CrossRef]
16. Yue, J. Enhanced *thj* signal at the LHC with $h \to \gamma\gamma$ decay and $\mathcal{CP}$-violating top-Higgs coupling. *Phys. Lett. B* **2015**, *744*, 131–136. [CrossRef]
17. Demartin, F.; Maltoni, F.; Mawatari, K.; Zaro, M. Higgs production in association with a single top quark at the LHC. *Eur. Phys. J. C* **2015**, *75*, 267. [CrossRef]
18. Barger, V.; Hagiwara, K.; Zheng, Y.J. Probing the Higgs Yukawa coupling to the top quark at the LHC via single top+Higgs production. *Phys. Rev. D* **2019**, *99*, 031701. [CrossRef]
19. Kraus, M.; Martini, T.; Peitzsch, S.; Uwer, P. Exploring BSM Higgs couplings in single top-quark production. *arXiv* **2019**, arXiv:1908.09100.
20. Bernreuther, W.; Brandenburg, A. Tracing CP violation in the production of top quark pairs by multiple TeV proton proton collisions. *Phys. Rev. D* **1994**, *49*, 4481–4492. [CrossRef]
21. Barger, V.; Hagiwara, K.; Zheng, Y.J. Probing the top Yukawa coupling at the LHC via associated production of single top and Higgs. *arXiv* **2019**, arXiv:1912.11795.

22. Patrick, R.; Scaffidi, A.; Sharma, P. Top polarisation as a probe of CP-mixing top-Higgs coupling in *tjh* signals. *Phys. Rev. D* **2020**, *101*, 093005. [CrossRef]
23. Ferroglia, A.; Fiolhais, M.C.; Gouveia, E.; Onofre, A. Role of the *t̄th* rest frame in direct top-quark Yukawa coupling measurements. *Phys. Rev. D* **2019**, *100*, 075034. [CrossRef]
24. Faroughy, D.A.; Kamenik, J.F.; Košnik, N.; Smolkovič, A. Probing the *CP* nature of the top quark Yukawa at hadron colliders. *J. High Energy Phys.* **2020**, *2*, 085. [CrossRef]
25. Fajfer, S.; Kamenik, J.F.; Melic, B. Discerning New Physics in Top-Antitop Production using Top Spin Observables at Hadron Colliders. *J. High Energy Phys.* **2012**, *8*, 114. [CrossRef]
26. Durieux, G.; Grossman, Y. Probing CP violation systematically in differential distributions. *Phys. Rev. D* **2015**, *92*, 076013. [CrossRef]
27. Bortolato, B.; Kamenik, J.F.; Košnik, N.; Smolkovič, A. Optimized probes of *CP* -odd effects in the *t̄th* process at hadron colliders. *Nucl. Phys. B* **2021**, *964*, 115328. [CrossRef]
28. Mangano, M.L.; Zanderighi, G.; Saavedra, J.A.A.; Alekhin, S.; Badger, S.; Bauer, C.W.; Becher, T.; Bertone, V.; Bonvini, M.; Boselli, S.; et al. Physics at a 100 TeV pp Collider: Standard Model Processes. *CERN Yellow Rep.* **2017**, 1–254. [CrossRef]
29. Dicus, D.A.; Sudarshan, E.C.G.; Tata, X. Factorization Theorem for Decaying Spinning Particles. *Phys. Lett.* **1985**, *154B*, 79–85. [CrossRef]
30. Bernreuther, W.; Brandenburg, A.; Si, Z.G.; Uwer, P. Top quark pair production and decay at hadron colliders. *Nucl. Phys.* **2004**, *B690*, 81–137. [CrossRef]
31. Bjorken, J.D.; Drell, S.D. *Relativistic Quantum Mechanics*; International Series in Pure and Applied Physics; McGraw-Hill: New York, NY, USA, 1965.
32. Atwood, D.; Bar-Shalom, S.; Eilam, G.; Soni, A. CP violation in top physics. *Phys. Rept.* **2001**, *347*, 1–222. [CrossRef]
33. Atwood, D.; Soni, A. Analysis for magnetic moment and electric dipole moment form-factors of the top quark via $e^+e^- \rightarrow t\bar{t}$. *Phys. Rev. D* **1992**, *45*, 2405–2413. [CrossRef]
34. Gunion, J.F.; Grzadkowski, B.; He, X.G. Determining the top - anti-top and Z Z couplings of a neutral Higgs boson of arbitrary CP nature at the NLC. *Phys. Rev. Lett.* **1996**, *77*, 5172–5175. [CrossRef]
35. Alwall, J.; Frederix, R.; Frixione, S.; Hirschi, V.; Maltoni, F.; Mattelaer, O.; Shao, H.S.; Stelzer, T.; Torrielli, P.; Zaro, M. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *J. High Energy Phys.* **2014**, *7*, 79. [CrossRef]
36. Artoisenet, P.; Frederix, R.; Mattelaer, O.; Rietkerk, R. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *J. High Energy Phys.* **2013**, *3*, 15. [CrossRef]
37. Degrande, C.; Duhr, C.; Fuks, B.; Grellscheid, D.; Mattelaer, O.; Reiter, T. UFO - The Universal FeynRules Output. *Comput. Phys. Commun.* **2012**, *183*, 1201–1214. [CrossRef]
38. Artoisenet, P.; de Aquino, P.; Demartin, F.; Frederix, R.; Frixione, S.; Maltoni, F.; Mandal, M.K.; Mathews, P.; Mawatari, K.; Ravindran, V.; et al. A framework for Higgs characterisation. *J. High Energy Phys.* **2013**, *11*, 43. [CrossRef]
39. Broggio, A.; Ferroglia, A.; Frederix, R.; Pagani, D.; Pecjak, B.D.; Tsinikos, I. Top-quark pair hadroproduction in association with a heavy boson at NLO+NNLL including EW corrections. *J. High Energy Phys.* **2019**, *8*, 39. [CrossRef]
40. Sjostrand, T.; Mrenna, S.; Skands, P.Z. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.* **2008**, *178*, 852–867. [CrossRef]
41. Cacciari, M.; Salam, G.P.; Soyez, G. FastJet User Manual. *Eur. Phys. J.* **2012**, *C72*, 1896. [CrossRef]
42. de Favereau, J.; Delaere, C.; Demin, P.; Giammanco, A.; Lemaître, V.; Mertens, A.; Selvaggi, M. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *J. High Energy Phys.* **2014**, *2*, 57. [CrossRef]
43. Mangano, M.L.; Moretti, M.; Piccinini, F.; Treccani, M. Matching matrix elements and shower evolution for top-quark production in hadronic collisions. *J. High Energy Phys.* **2007**, *1*, 13. [CrossRef]
44. Farina, M.; Grojean, C.; Maltoni, F.; Salvioni, E.; Thamm, A. Lifting degeneracies in Higgs couplings using single top production in association with a Higgs boson. *J. High Energy Phys.* **2013**, *5*, 22. [CrossRef]
45. Aaboud, M.; Aad, G.; Abbott, B.; Abdinov, O.; Abeloos, B.; Abidi, S.; AbouZeid, O.; Abraham, N.; Abramowicz, H.; Abreu, H.; et al. Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in *pp* collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *Phys. Rev.* **2018**, *D97*, 072016. [CrossRef]
46. ATLAS Collaboration . Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Phys. Lett.* **2018**, *B784*, 173–191. [CrossRef]
47. Krohn, D.; Schwartz, M.D.; Lin, T.; Waalewijn, W.J. Jet Charge at the LHC. *Phys. Rev. Lett.* **2013**, *110*, 212001. [CrossRef]
48. Fraser, K.; Schwartz, M.D. Jet Charge and Machine Learning. *J. High Energy Phys.* **2018**, *10*, 93. [CrossRef]
49. *A New Tagger for the Charge Identification of b-Jets*; Technical Report ATL-PHYS-PUB-2015-040; CERN: Geneva, Switzerland, 2015.
50. *Measurement of the Jet Vertex Charge Algorithm Performance for Identified b-Jets in t̄t Events in pp Collisions with the ATLAS Detector*; Technical Report ATLAS-CONF-2018-022; CERN: Geneva, Switzerland, 2018.
51. Brehmer, J.; Kling, F.; Espejo, I.; Cranmer, K. MadMiner: Machine learning-based inference for particle physics. *Comput. Softw. Big Sci.* **2020**, *4*, 3. [CrossRef]
52. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: tensorflow.org (accessed on 21 May 2021).

53. Bergstra, J.; Yamins, D.; Cox, D.D. 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 (ICML'13)—Volume 28, Atlanta, GA, USA, 16–21 June 2013; pp. 115–123.

54. Clavijo, J.M.; Glaysher, P.; Katzy, J.M. Adversarial domain adaptation to reduce sample bias of a high energy physics classifier. *arXiv* **2020**, arXiv:2005.00568.

55. Cepeda, M.; Gori, S.; Ilten, P.; Kado, M.; Riva, F.; Abdul Khalek, R.; Aboubrahim, A.; Alimena, J.; Alioli, S.; Alves, A.; 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*; Dainese, A., Mangano, M., Meyer, A.B., Nisati, A., Salam, G., Vesterinen, M.A., Eds.; 2019; Volume 7, pp. 221–584. Available online: https://e-publishing.cern.ch/index.php/CYRM/article/view/952doi:10.23731/CYRM-2019-007.221 (accessed on 21 May 2021). [CrossRef]

56. Abada, A.; Abbrescia, M.; AbdusSalam, S.S.; Abdyukhanov, I.; Fernandez, J.A.; Abramov, A.; Aburaia, M.; Acar, A.O.; Adzic, P.R.; Agrawal, P. et al. FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1. *Eur. Phys. J. C* **2019**, *79*, 474. [CrossRef]