Scalar mass dependence of angular variables in $t\bar{t}\phi$ production

Duarte Azevedo,[a] **Rodrigo Capucha,**[a] **António Onofre**[b] **and Rui Santos**[a,c]

[a]*Centro de Física Teórica e Computacional, Faculdade de Ciências, Universidade de Lisboa,*
Campo Grande, Edifício C8 1749-016 Lisboa, Portugal

[b]*Departamento de Física, Universidade do Minho,*
4710-057 Braga, Portugal

[c]*ISEL — Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa,*
1959-007 Lisboa, Portugal

E-mail: drpazevedo@fc.ul.pt, rodrigocapucha@hotmail.com,
antonio.onofre@cern.ch, rasantos@fc.ul.pt

ABSTRACT: In this paper we explore CP discrimination in the associated production of top-quark pairs $(t\bar{t})$ with a generic scalar boson (ϕ) at the LHC. We probe the CP-sensitivity of several observables for a varying scalar boson mass and CP-number, either CP-even $(\phi = H)$ or CP-odd $(\phi = A)$, using dileptonic final states of the $t\bar{t}\phi$ system, with $\phi \to b\bar{b}$. We show that CP-searches are virtually impossible for ϕ boson masses above a few hundred GeV in this channel. A full phenomenological analysis was performed, using Standard Model background and signal events generated with `MadGraph5_aMC@NLO` and reconstructed using a kinematic fit. The most sensitive CP-observables are used to compute Confidence Levels (CLs), as a function of luminosity, for the exclusion of different signal hypotheses with scalar and pseudoscalar boson masses that range from $m_\phi = 40\,\text{GeV}$ up to $200\,\text{GeV}$. We finalize by analysing the impact of a measurement (or limit) of the CP-violating angle in the parameter space of a complex two-Higgs doublet model known as the C2HDM.

Contents

1 Introduction

The CP-nature of the discovered Higgs boson is still an open issue. Although the ATLAS and CMS collaborations established that the discovered 125 GeV Higgs [1, 2] cannot be a pure pseudoscalar with more than 99% confidence level (CL), a mixed state with a significant CP-odd component is still possible. The need for further sources of CP-violation was first discussed by Sakharov as one of the three conditions for baryogenesis to occur [3]. This is an important motivation to look for physics Beyond the Standard Model (BSM) and, in particular, to models with extra sources of CP-violation. One of the simplest extensions of the Standard Model (SM) we can build with a CP-violating scalar sector is to add an extra scalar doublet to the SM field content while keeping the same gauge symmetries. The CP-conserving version of the model is commonly referred to as two-Higgs doublet model (2HDM), while the simplest CP-violating version of the model is known as complex 2HDM (C2HDM) and has been the subject of many studies [4–17]. Due to its simplicity, the C2HDM is seen as an ideal benchmark model to test the CP quantum numbers of scalars at the LHC. In the C2HDM, all three neutral scalars may have a mixture of CP-even and CP-odd components and there is no restriction to the mass of these scalars. Although one of the scalars has to be the discovered 125 GeV Higgs boson, it can be any of the three neutral scalars predicted by the theory, from the lightest to the heaviest.

The search for BSM physics is a major goal of the LHC experiments. The measurement of the Yukawa couplings has become a primary target, since it is decisive to establish the CP-nature of the scalars in case a new scalar is discovered. The relation between the scalar and the pseudoscalar components in the Yukawa couplings can be directly probed either in the production or in the decays of the scalars, depending on the final state fermions.

Examples of the use of asymmetries to probe the CP-nature of the Higgs boson in the top quark Yukawa coupling were discussed in [18–24], while the decays of the tau leptons were used to probe the τ-lepton Yukawa coupling [25–29]. These studies for the top-quark and for the τ-lepton are now being discussed in detail by the ATLAS and CMS collaborations. In the case of the top quark we will be probing the Yukawa coupling directly in the production process. For the τ-leptons, the Higgs decay is used. It is important to point out that there is still room for very large pseudoscalar components in the couplings to b-quarks and τ-leptons, for any of the scalars in some Yukawa versions of the C2HDM, considering the recently announced ACME collaboration's constraint on the electron EDM,

$$|d_e| < 1.1 \times 10^{-29} \text{ e cm}, \tag{1.1}$$

from measurements of the ThO molecule[30], as recently reported in [31, 32].

In this work we will examine in detail how the asymmetries and the angular variables distributions previously proposed change when the mass of the scalar differs from the measured Higgs mass. Many studies have examined several angular variables in $t\bar{t}\phi$ production (with $m_\phi = 125\,\text{GeV}$). These have considered a next-to-leading order (NLO) computation, including resummation of soft emission corrections to next-to-leading-logarithmic accuracy (NLL) [33–38]. Nontheless, a detailed study for a scalar with a mass either below or above this value is still not available in the literature.

We will build on a series of papers where the issue of the CP-nature of the discovered Higgs boson was thoroughly studied in associated production with top quark pairs [20, 21, 39, 40]. We will discuss the same set of angular variables with several goals in mind. The first one is to answer the question: if a new scalar or pseudoscalar boson exists, what is the confidence level to exclude a signal hypothesis (either CP-even or CP-odd) assuming the SM holds, as a function of the LHC luminosity and ϕ boson mass? The second one focuses on determining the confidence level for exclusion of a pure CP-odd signal in case a new massive scalar boson is found, as a function of the ϕ boson mass and the LHC luminosity. The third one relates to setting the confidence level for the exclusion of the SM (once again as a function of mass and luminosity) assuming a new CP-even scalar particle signal is found.

The outline of the paper is as follows. In section 2, we describe the ϕ boson mass dependence of the several angular distributions to be studied. In section 3, we present and discuss our main results. In section 4, we consider the impact of the discovery of a new Higgs boson on the parameter space of the main benchmark model for CP-violation studies, the complex version of the two-Higgs doublet model (C2HDM). Our conclusions are presented in section 5.

2 Theoretical limitations on asymmetries measurements

The most general Yukawa interaction of a boson (ϕ), with no definite value of CP, to a top quark pair can be written as

$$\mathcal{L} = \kappa_t y_t \bar{t}(\cos\alpha + i\gamma_5 \sin\alpha)t\phi, \tag{2.1}$$

where y_t is the SM Yukawa coupling, κ_t parametrises the total coupling strength relative to the SM and the angle α parametrises the CP-phase, which is related to the parameters in the Higgs potential. We will refer to $\phi = H$ for the pure CP-even scenario and $\phi = A$ for the pure CP-odd case. The pure CP-even case is recovered by setting $\cos\alpha = \pm 1$ while the pure CP-odd case is obtained by fixing $\cos\alpha = 0$.

In previous works [18–21] several angular variables were proposed, not only to increase the sensitivity in discriminating signals from irreducible backgrounds at the LHC in $t\bar{t}\phi$ final states, but also as a means to probe the CP nature of the Yukawa coupling in $t\bar{t}\phi$ production at the LHC. The results in [20, 21] showed that we can define a minimal set of variables to obtain the best possible sensitivity, to achieve both goals in a very effective way. While these studies assumed a mass of $125\,\text{GeV}$ for the ϕ boson, in this paper we extend their use to a wider mass range, from $40\,\text{GeV}$ to $500\,\text{GeV}$. This is discussed in the following sections.

2.1 $t\bar{t}H$ and $t\bar{t}A$ angular distributions

A first set of variables is introduced [20] using θ_Y^X, defined as the angle between the direction of the Y system 3-momentum (in the rest frame of X) with respect to the momentum direction of the X system (in the rest frame of its parent system). When reconstructing the signal angular distributions, we consider successive two body decays of the $t\bar{t}\phi$ system down to the final state particles i.e., the quarks (or jets), the charged and the neutral leptons, which originated from the decays of the t, $\bar{t}$ quarks and ϕ boson. If the decay chain of the $t\bar{t}\phi$ system is labelled (123), the successive decays considered include all possible combinations of the type $(123) \to 1 + (23)$, $(23) \to 2 + (3)$ and $(3) \to 4 + 5$ (see figure 1). We then build three families of observables: $f(\theta_1^{123})g(\theta_4^3)$, $f(\theta_1^{123})g(\theta_3^{23})$ and $f(\theta_3^{23})g(\theta_4^3)$, with $f, g = \{\sin, \cos\}$. The momentum direction of the (123) system is measured with respect to the laboratory (LAB) frame, where the net 3-momentum of the protons colliding is zero. Particles 1 to 3 are either the t or the $\bar{t}$ quark, or the Higgs boson, while particle 4 can be any of the products of the decay of the top quarks and the Higgs boson, including the intermediate W bosons. We use two ways of computing particle 4 Lorentz vector in the centre-of-mass of particle 3. One is by using the laboratory four-momentum of both particles 3 and 4, and boost particle 4 directly to the centre-of-mass frame of particle 3 (*direct* boost). The other, is to boost particles 3 and 4 sequentially through all intermediate centre-of-mass systems until particle 4 is evaluated in the centre-of-mass frame of particle 3 (*sequential* boost or *seq.* boost).

We will also use the variables b_2 and b_4 as defined in [18, 40] in the LAB and $t\bar{t}\phi$ centre-of-mass systems ($b_2^{t\bar{t}\phi}$ and $b_4^{t\bar{t}\phi}$, respectively),

$$b_2 = (\vec{p}_t \times \hat{k}_z).(\vec{p}_{\bar{t}} \times \hat{k}_z)/(|\vec{p}_t|.|\vec{p}_{\bar{t}}|), \quad b_4 = (p_t^z.p_{\bar{t}}^z)/(|\vec{p}_t|.|\vec{p}_{\bar{t}}|), \tag{2.2}$$

where the z-direction corresponds to the beam line. It is worth noting that b_2 and b_4 have a natural physics interpretation. They depend on the t and $\bar{t}$ polar angles, θ_t and $\theta_{\bar{t}}$ respectively, with respect to the z-direction, and can be expressed as $b_2 = \sin\theta_t \times \sin\theta_{\bar{t}}$ and $b_4 = \cos\theta_t \times \cos\theta_{\bar{t}}$.
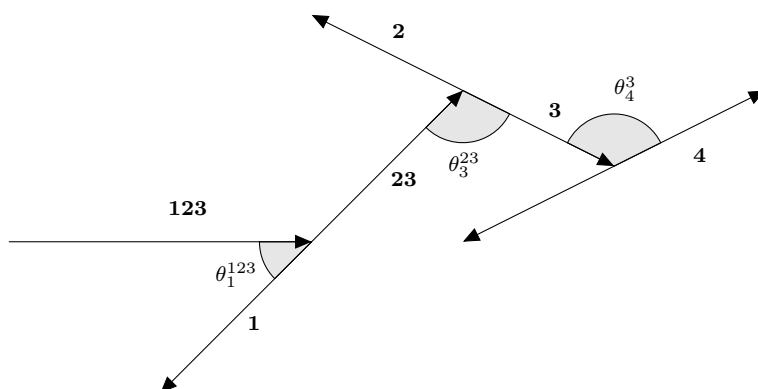
Figure 1. Schematic representation of the $t\bar{t}\phi$ decay chain and angles between the different systems and decay products.
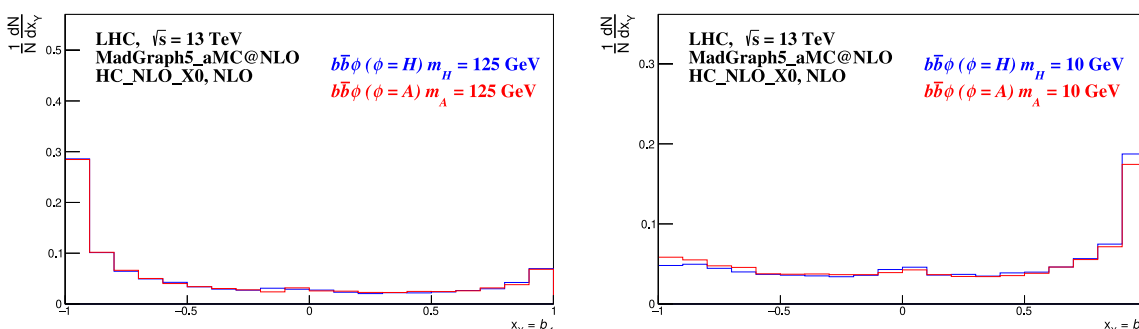

Figure 2. Parton level b_4 distributions at Next-to-leading order (NLO), normalized to unity, for $m_\phi = 125\,\text{GeV}$ (left) and $m_\phi = 10\,\text{GeV}$ (right).

Forward-backward asymmetries associated to each of the observables under study were defined according to [20]

$$A^Y_{FB} = \frac{\sigma(x_Y > x'_Y) - \sigma(x_Y < x'_Y)}{\sigma(x_Y > x'_Y) + \sigma(x_Y < x'_Y)}, \tag{2.3}$$

where $\sigma(x_Y > x'_Y)$ and $\sigma(x_Y < x'_Y)$ correspond to the total cross section for x_Y above and below x'_Y. The latter is the central value of the x_Y domain.

The reason why these distributions allow us to probe the CP-nature of a scalar in the $t\bar{t}\phi$ coupling lies ultimately in the behaviour of the cross section as a function of the particle's CP value. In fact, as discussed in [18], the amplitude for the process $pp \to t\bar{t}\phi$ has two terms: one that does not depend on the mixing angle, α, and another that is proportional to $\cos 2\alpha$. Hence, only the latter is sensitive to a CP-odd component of the Yukawa coupling. This term is proportional to the top quark mass and therefore its contribution is important, as long as the Higgs boson mass is of the same order of magnitude. One could ask if the process $pp \to b\bar{b}\phi$ could be used to probe the Yukawa structure of the $b\bar{b}\phi$ vertex. The answer is clearly negative because the CP-asymmetric term is now proportional to m_b^2, that is, at least three orders of magnitude smaller. In the left panel of figure 2, we present the b_4 distribution, at parton level, for the process
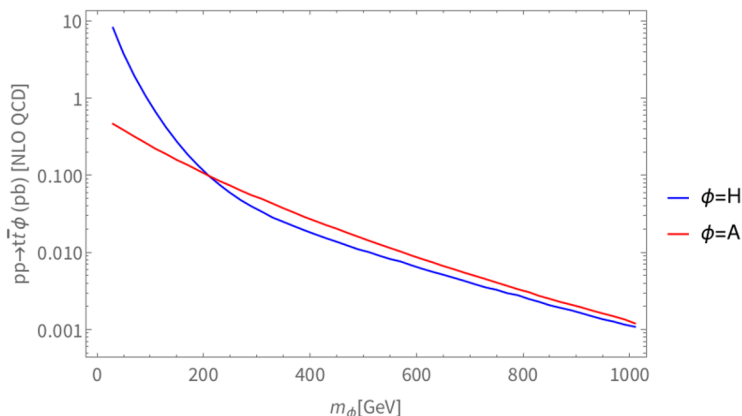
Figure 3. Total cross section for the process $pp \to t\bar{t}H$ (blue) and $pp \to t\bar{t}A$ (red) as a function of the ϕ boson mass, at NLO, for a centre-of-mass energy of $13\,\mathrm{TeV}$, at the LHC. Details about the generation of the events represented through figures 3, 4, 5 and 6, can be found in section 3.1.

$pp \to b\bar{b}\phi$ for $m_\phi = 125\,\mathrm{GeV}$. In blue, we present the pure scalar case while in red we show the pure pseudoscalar one. As expected no difference is found in the distributions. We have checked that the distributions of all other angular variables follow the same trend and again no difference was seen. Finally we repeated the procedure for a very light scalar, with a mass of $m_\phi = 10\,\mathrm{GeV}$, with similar null results as we show on the right side of the same figure. The case of the $b\bar{b}\phi$ final state has been previously discussed in [41, 42].

In figure 3, we present the total cross section, at NLO, for a centre-of-mass energy of $13\,\mathrm{TeV}$, at the LHC, for the process $pp \to t\bar{t}H$ (blue) and $pp \to t\bar{t}A$ (red) as a function of the ϕ boson mass. The fact that the CP-asymmetric term is much larger compared to the $b\bar{b}\phi$ case means that CP-discrimination between the different CP-components of the Higgs is now possible. Figures 4 and 5 show the b_2 and b_4 distributions for $t\bar{t}H$ and $t\bar{t}A$ events with different ϕ boson masses, computed in the LAB and in the centre-of-mass frame of the $t\bar{t}\phi$ system, respectively. They are shown at parton level without any cuts. Next-to-leading order corrections and shower effects (NLO+Shower) are also included. Clear differences are now visible between the scalar and pseudoscalar signals, and also between the distributions computed in the LAB and in the centre-of-mass frames.

In order to study the CP-sensitivity as a function of the ϕ boson mass, forward-backward asymmetries of some variables were computed for each CP-component of the top quark Yukawa coupling, i.e., CP-even and CP-odd. The variables are,

- $X = b_2$,

- $X = b_4$,

- $X = \sin\theta_\phi^{t\bar{t}\phi} \sin\theta_{\bar{t}}^{t\bar{t}}$,

- $X = \sin\theta_t^{t\bar{t}\phi} \sin\theta_{W^+}^{\phi}$ (with *sequential* boost).

The full normalized distributions, at parton level, are shown in figure 6. As hinted by the behaviour of the cross sections, for large enough Higgs masses, the difference between
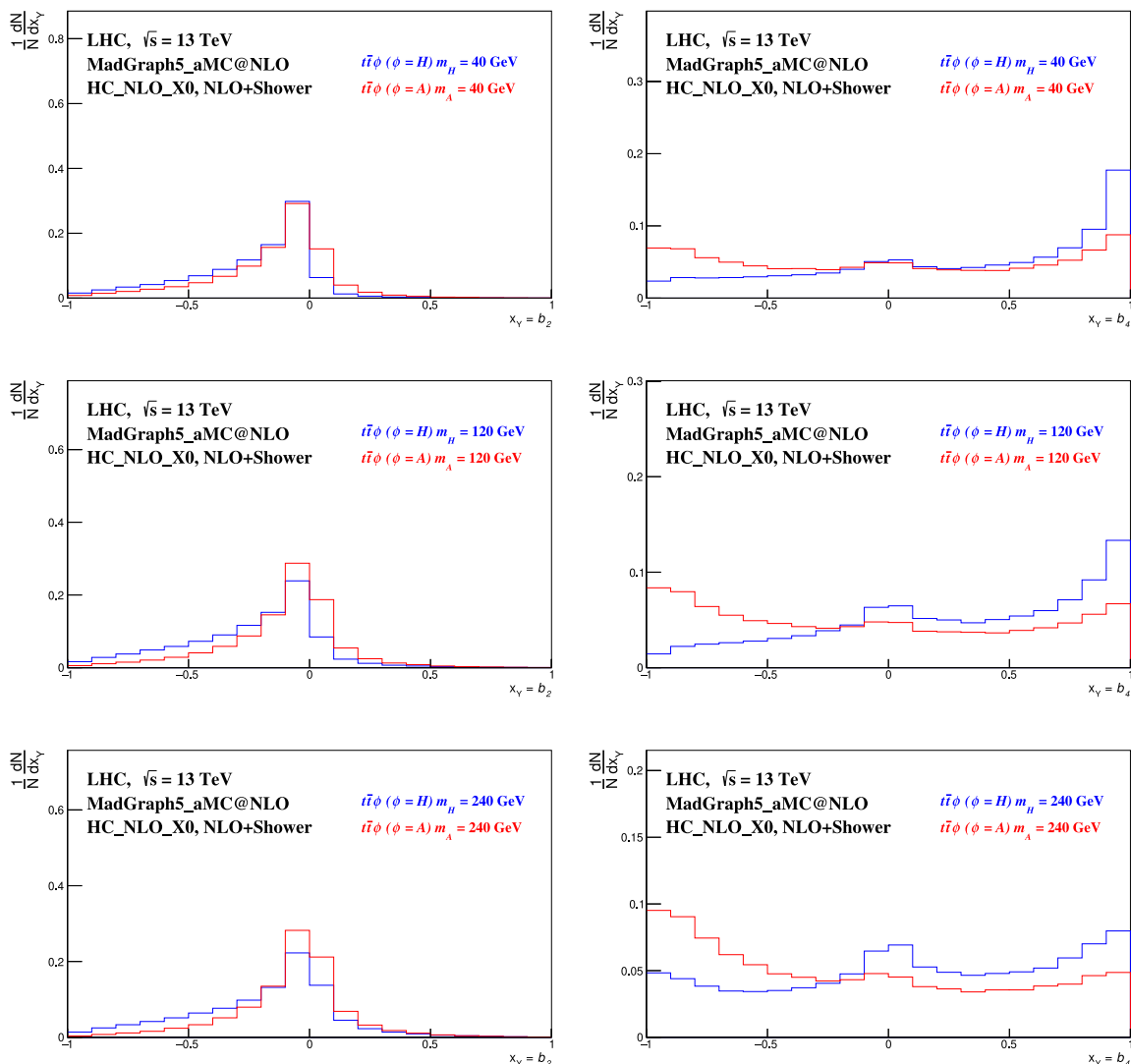
– 5 –

Figure 4. Full normalized distributions at NLO+Shower for the variables b_2 (left) and b_4 (right) in the LAB frame, without any selection cuts nor reconstruction, for both the pure scalar (blue line) and pure pseudoscalar (red line) signals with $m_\phi = 40$, 120 and 240 GeV.

CP-even and CP-odd distributions disappears. Although this behaviour was confirmed for all variables, the exact mass value for which the difference becomes negligible depends on the choice of variables. The maximum value of the ϕ boson mass for which a meaningful difference between distributions exists is 400 GeV.
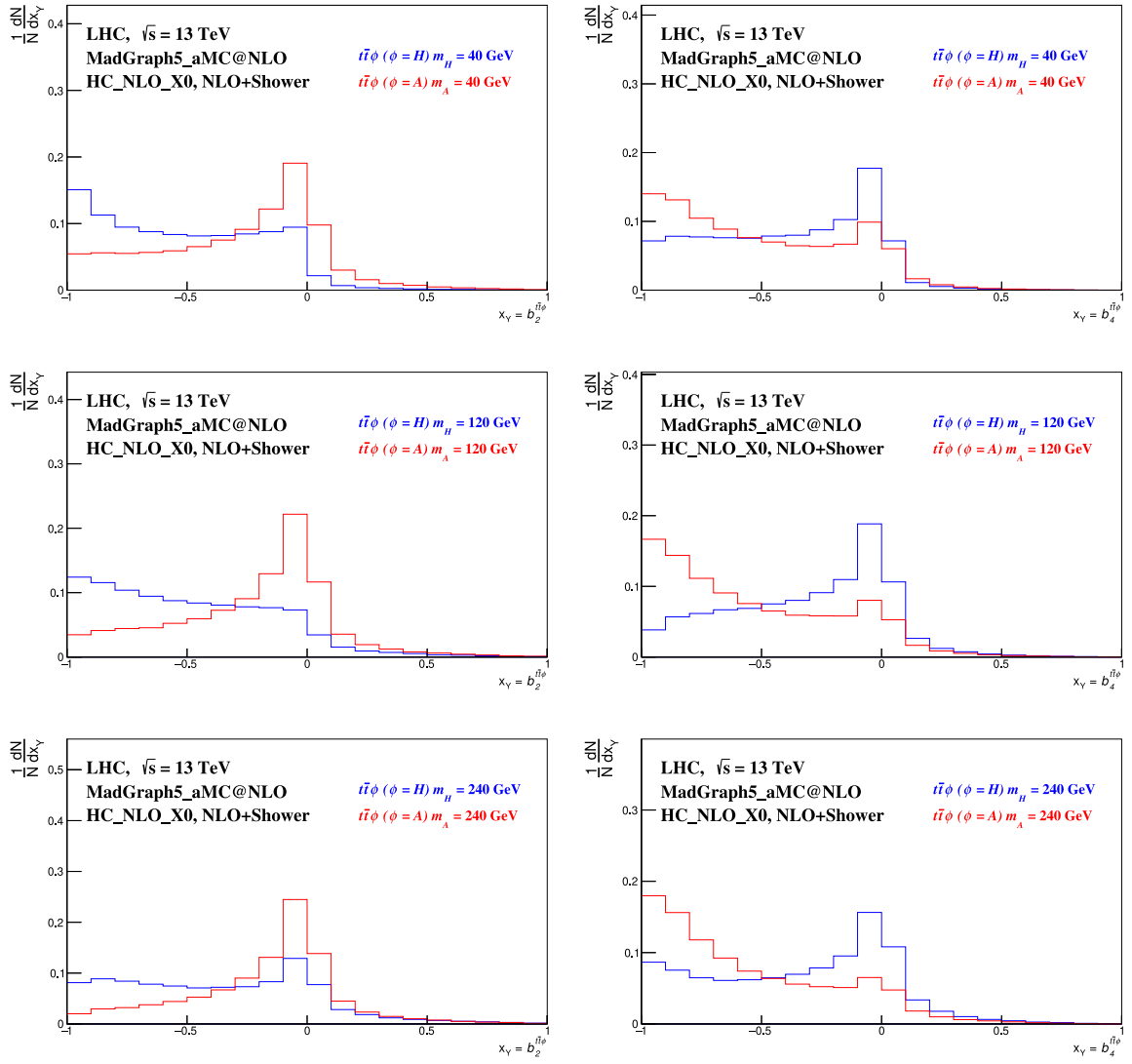
Figure 5. Same as in figure 4, but now the b_2 and b_4 distributions are computed in the centre-of-mass of the $t\bar{t}\phi$ system.
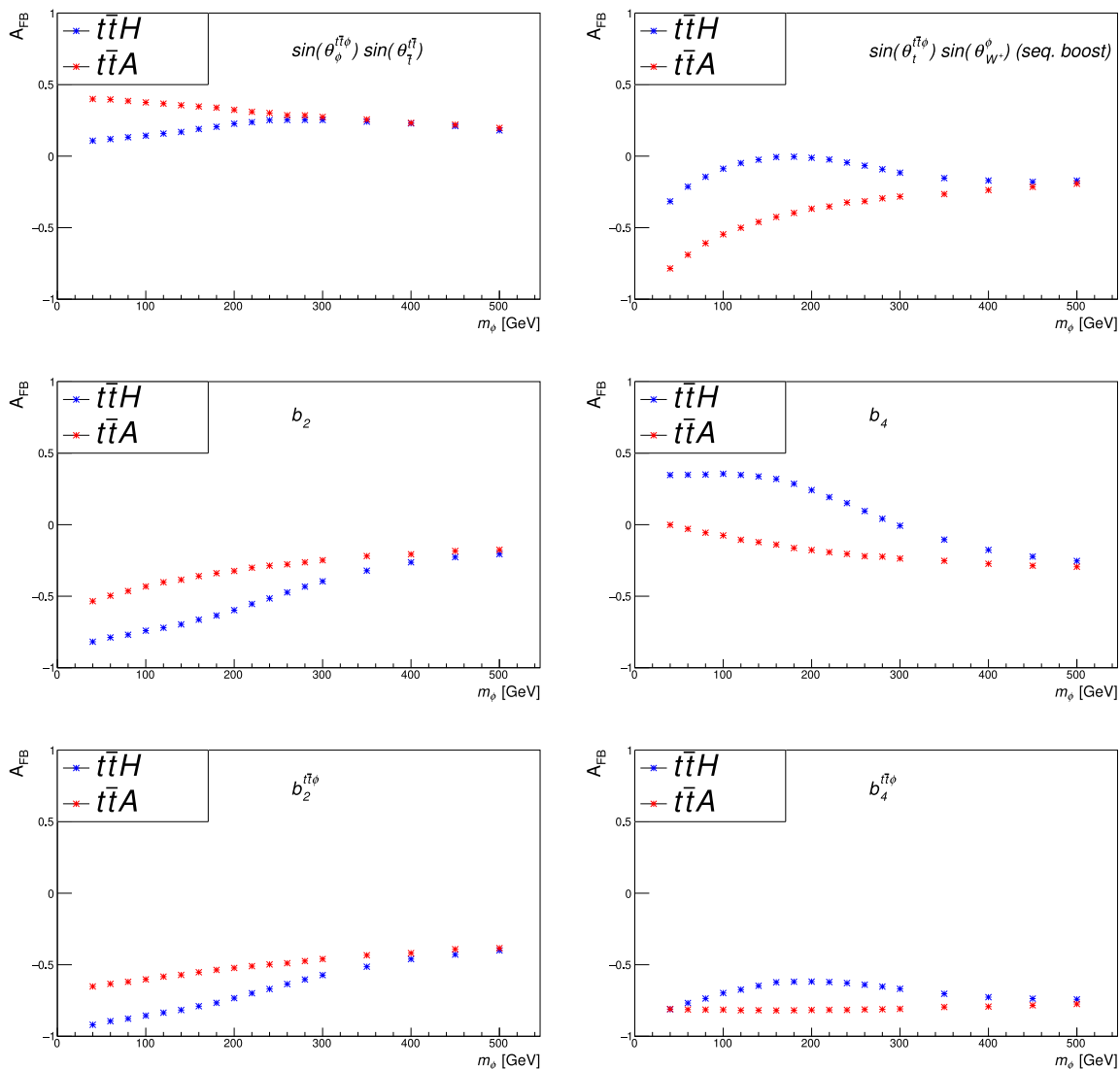
Figure 6. Forward-backward asymmetries as a function of the ϕ boson mass, without cuts, at NLO+Shower. On top, we have two angular distributions, $\sin\theta_\phi^{t\bar{t}\phi} \sin\theta_{\bar{t}}^{t\bar{t}}$ (left panel) and $\sin\theta_t^{t\bar{t}\phi} \sin\theta_{W^+}^\phi$ with a *sequential* boost (right panel). On the middle and last rows, we show the variables b_2 (left) and b_4 (right), in the LAB and centre-of-mass frames, respectively.

3 Results and discussion

3.1 Generation of events

Signal events from $pp \rightarrow t\bar{t}\phi$ associated production at the LHC (with $\phi = \{H, A\}$), were generated at NLO with the Higgs Characterization model HC_NLO_X0 [43], using MadGraph5_aMC@NLO[44]. The pure CP-even and the pure CP-odd odd samples were generated by setting the CP-phase to $\cos \alpha = 1$ or 0, respectively, following equation (2.1), with $\kappa_t = 1$. Several samples, for both scalar and pseudoscalar signals, were generated with masses m_ϕ between 40 and 300 GeV, in steps of 20 GeV, and also the four masses m_ϕ = 350, 400, 450 and 500 GeV. While the CP-even and CP-odd bosons were only allowed to decay to a pair of b-quarks ($\phi \rightarrow b\bar{b}$), the $t\bar{t}$ system was assumed to decay to a pair of b-quarks and two intermediate $W^\pm$ gauge bosons which, in turn, decay to two charged leptons and two neutrinos $t(\bar{t}) \rightarrow bW^+(\bar{b}W^-) \rightarrow b\ell^+\nu_\ell(\bar{b}\ell^-\bar{\nu}_\ell)$. Following the decay of all intermediate massive particles, the signal final state is characterized by the presence of two oppositely charged leptons, two neutrinos and two $b\bar{b}$ quark pairs, at parton level. Only W boson decays to electrons (e) and muons (μ) were considered as signal. This configuration defines the dileptonic channel.

In addition to the signal samples, backgrounds from SM processes were also generated using MadGraph5_aMC@NLO. The dominant background, a pair of top- and b-quarks ($t\bar{t}b\bar{b}$), as well as the associated production of top-quarks with the SM Higgs boson ($t\bar{t}H_{\mathrm{SM}}$), were generated at NLO. For the latter, a SM Higgs boson mass of $m_{H_{\mathrm{SM}}} = 125$ GeV was assumed. These two backgrounds lead to the same partonic final state as the signal.

The remaining backgrounds considered, which were all generated at tree-level (LO), are:

- $t\bar{t}$+3 jets i.e., top-quark pair production with up to three light jets.

- $t\bar{t}V$+ jets i.e., top-quark pair production with one gauge boson ($V = Z, W^\pm$), plus up to one light jet.

- Single top quark production through the s-, t-channel (with up to one additional jet) and Wt associated production.

- W+4 jets, i.e., $W^\pm$ boson production with up to four light jets.

- $Wb\bar{b}$+2 jets, i.e., $W^\pm$ boson production with two jets from the hadronization of b-quarks (b-jets), and up to two additional light jets.

- Z+4 jets i.e., Z boson production with up to four light jets.

- $Zb\bar{b}$+2 jets i.e., Z boson production with a pair of b-jets plus up to two light jets.

- WW, WZ, ZZ+3 jets i.e., diboson production with up to three jets.

All events were generated assuming proton collisions at the LHC with a centre-of-mass energy of 13 TeV. The masses of the top quarks (m_t) and the W bosons (m_W), were set to 173 GeV and 80.4 GeV, respectively, while their widths were set to the default

MadGraph5_aMC@NLO values of 1.4915 GeV and 2.0476 GeV, respectively. For all samples, the NNPDF2.3 [45] parton distribution functions (PDFs), were used. The renormalization and factorisation scales were fixed to the sum of the transverse masses of all final state particles and partons. The decay of particles was performed by MadSpin [46] for signal and background events in order to preserve spin correlations among the decay products and with the respective heavy parent resonances. Parton shower and hadronization was performed by Pythia6 [47]. The matching between the generator and the parton shower used the MLM scheme [48] for the LO samples and the MC@NLO matching [49] for the NLO events. For a fast, parametrised detector simulation of a LHC-like experiment, we used Delphes [50] with the default ATLAS parameter card. For jet reconstruction of the signal and background events, FastJet [51] is employed with the anti-k_t algorithm [52] with a cone size of $\Delta R = 0.7$.[1] Transverse momentum (p_T) cuts are applied to jets such that, in any events, these objects are kept if the following condition is met

$$p_T^{\text{jet}} \geq 10 \text{ GeV}. \tag{3.1}$$

No additional cuts were applied to the transverse momentum of leptons nor to the pseudo-rapidity (η) of jets and leptons (at generation level).

3.2 Kinematic reconstruction

After event generation, hadronization and detector simulation, we use a kinematic reconstruction to assign detector level jets to partons from the hard-scattering process and, using the detected charged leptons, reconstruct the massive intermediate particles i.e., the top quarks, the W and ϕ bosons. This, unavoidably, requires the reconstruction of the undetected neutrinos, which is performed on an event by event basis, using the MadAnalysis5 [53] framework.

Only events with at least two charged leptons of opposite charge and four or more jets are selected and reconstructed. Both leptons and jets were required to have $p_T \geq 20$ GeV and $|\eta| \leq 2.5$, which leads to signal selection efficiencies that vary from 9% (12%) to 18% (19%) for masses of the scalar (pseudoscalar) from 40 GeV to 200 GeV, respectively. The uncertainties on these numbers are smaller than 0.2%. It should be stressed at this point, that no attempt to optimize the selection was applied by looking for instance, to boosted jets, which is outside the scope of this paper.

One of the main challenges of the kinematic reconstruction, is the assignment of jets to the reconstructed parton level objects, that match correctly the decay particles of the top quarks, the W and the ϕ bosons. In order to check the performance of the kinematic reconstruction a *truth-match* approach was used for the assignment, by finding the four jets with smallest ΔR distance to the parton level b-quarks. As we expect a one-to-one correspondence, a wrong assignment leads, necessarily, to combinatorial background. As when dealing with real data no possible *truth-match* information is available, an association criteria needs to be applied to the events. This relies on a multivariate analysis method

---

[1] $\Delta R \equiv \sqrt{\Delta\phi^2 + \Delta\eta^2}$, where $\Delta\phi\,(\Delta\eta)$ correspond to the difference in the azimuthal angle (pseudo-rapidity) of two objects.
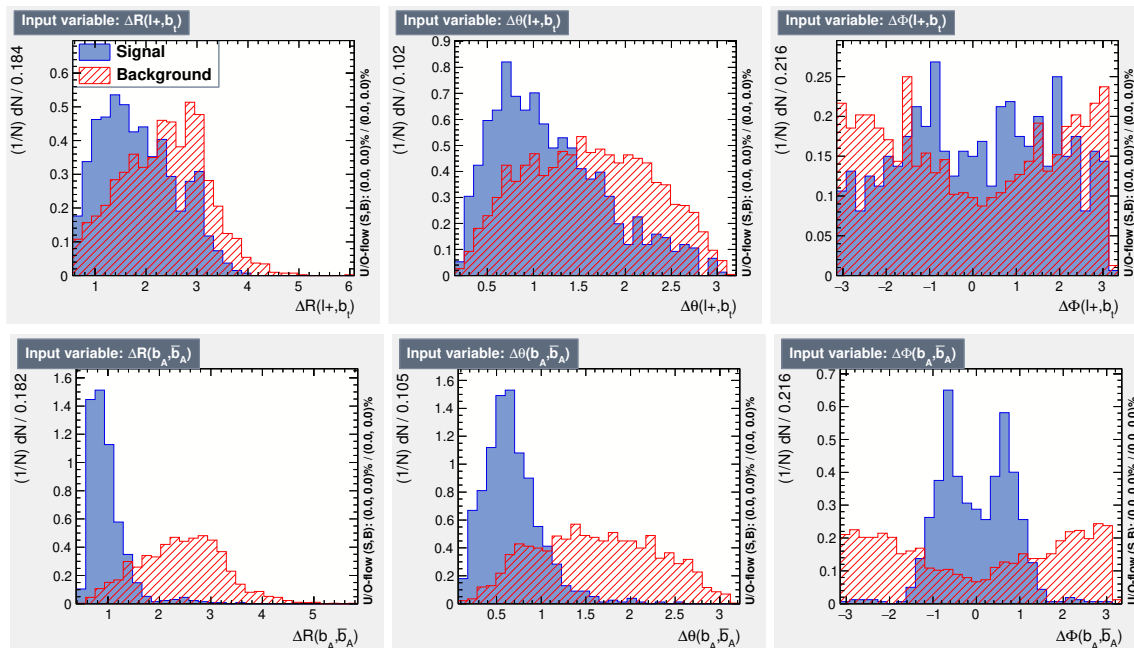
Figure 7. Distributions of the TMVA input variables for the signal (blue) and background (red) samples for $t\bar{t}A$ events with $m_A = 40\,\text{GeV}$. The angular variables ΔR, $\Delta\theta$ and $\Delta\Phi$ for the pairs (l^+, b_t) (top) and $(b_A, \bar{b}_A)$ (bottom) are computed at parton level.

tailored for each case, CP-even and CP-odd, using TMVA [54]. In both cases, two samples labelled as signal and background were created from simulated $t\bar{t}\phi$ signal events and used for training and testing. While signal samples contain kinematic distributions only from the correct association, background samples contain equivalent kinematic distributions from wrong associations. The following variables were used for training the methods: ΔR, $\Delta\Phi$, $\Delta\theta$ for the pairs (b_t, l^+), $(\bar{b}_{\bar{t}}, l^-)$ and $(b_\phi, \bar{b}_\phi)$, where b_t ($\bar{b}_{\bar{t}}$) represents the bottom (anti-bottom) quark from the top (anti-top) decay and b_ϕ ($\bar{b}_\phi$) represents the bottom (anti-bottom) quark from the Higgs decay. The invariant mass of the first two pairs, at parton level, and the invariant mass of the system $(b_\phi, \bar{b}_\phi)$ at the detector level, were also considered. These variables and their correlations are shown in figures 7 and 8 for the CP-odd case with $m_A = 40\,\text{GeV}$. We have found that, for all the mass values of the CP-even and CP-odd signals, the methods with best performance are the Boosted Decision Tree (BDT) and the Gradient Boosted Decision Tree (BDTG). The latter is the method used in the kinematic reconstruction. During the testing phase, the jet combination chosen is the one returning the highest value of the BDTG discriminant. The *Receiver Operating Characteristic* (ROC) curve and the BDT and BDTG discriminant distributions are shown in figures 9 and 10, respectively, for $t\bar{t}A$ with $m_A = 40\,\text{GeV}$.

In events with jet multiplicity above six, only the six highest p_T jets are considered. The reason for this choice relates to the fact that, in about 95% of all signal events, the jets corresponding to the hadronization of parton level b-quarks are among the six with highest p_T. Jet combinations also need to verify $m_{l^+ b_t} (m_{l^- \bar{b}_{\bar{t}}}) < 150\,\text{GeV}$ and $20\,\text{GeV} < m_{b_\phi \bar{b}_\phi} < 300\,\text{GeV}$.
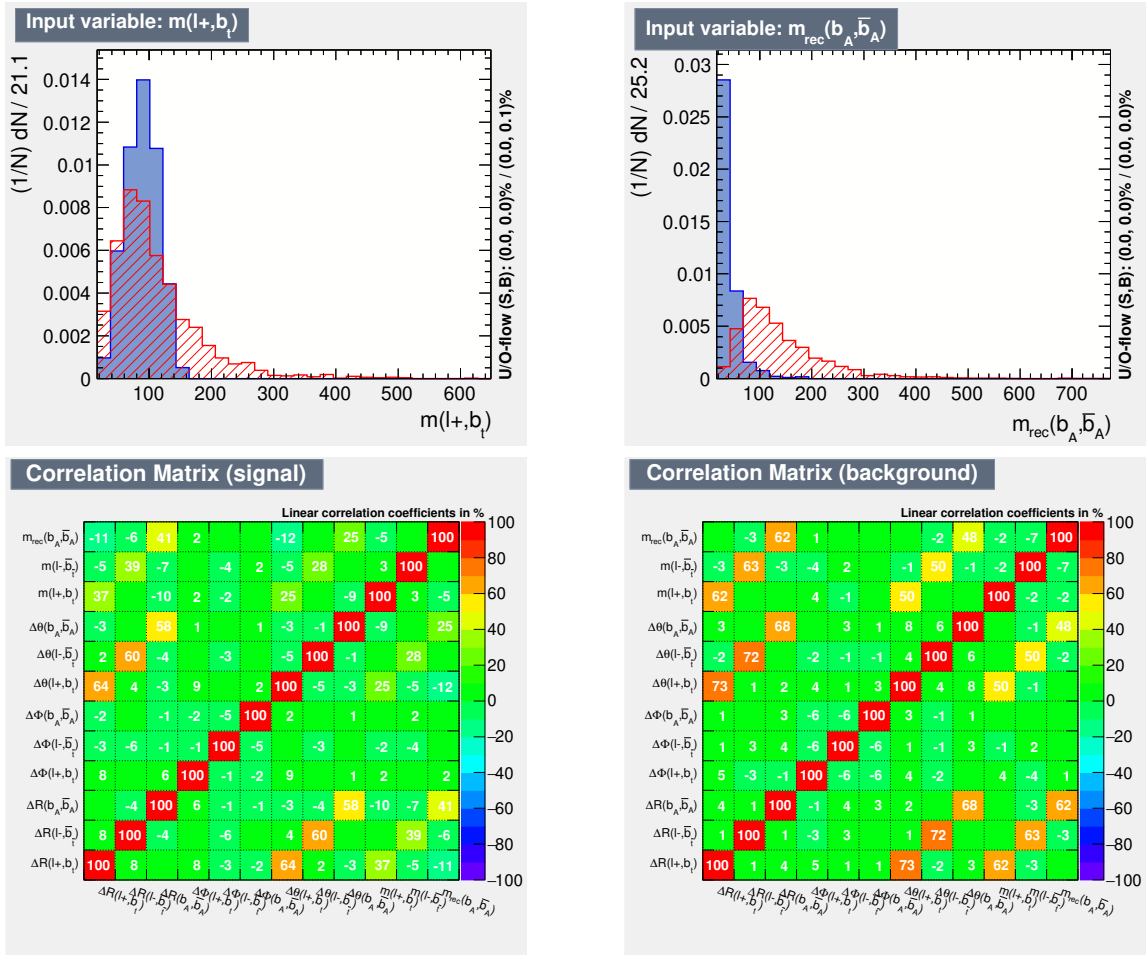
Figure 8. Invariant masses for the systems (l^+, b_t) at parton level (top left) and $(b_A, \bar{b}_A)$ obtained with truth-matching (top right), for $t\bar{t}A$ events with $m_A = 40\,\mathrm{GeV}$. Below, we have the matrix correlations between the TMVA input variables for the signal (left) and background (right).
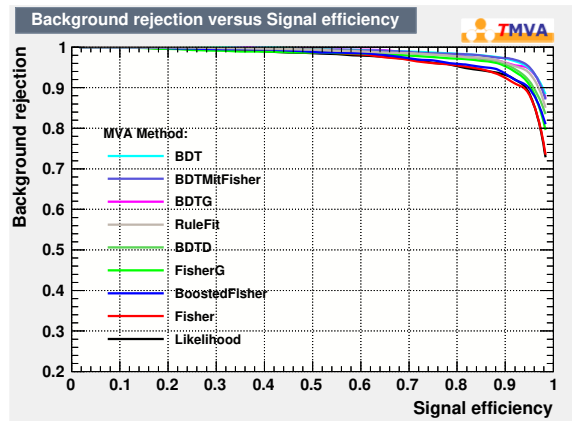

Figure 9. Most performant methods in terms of correct jet assignment for $t\bar{t}A$ events with $m_A = 40\,\mathrm{GeV}$. We found that typically BDT or BDTG are the best for all ϕ boson masses.
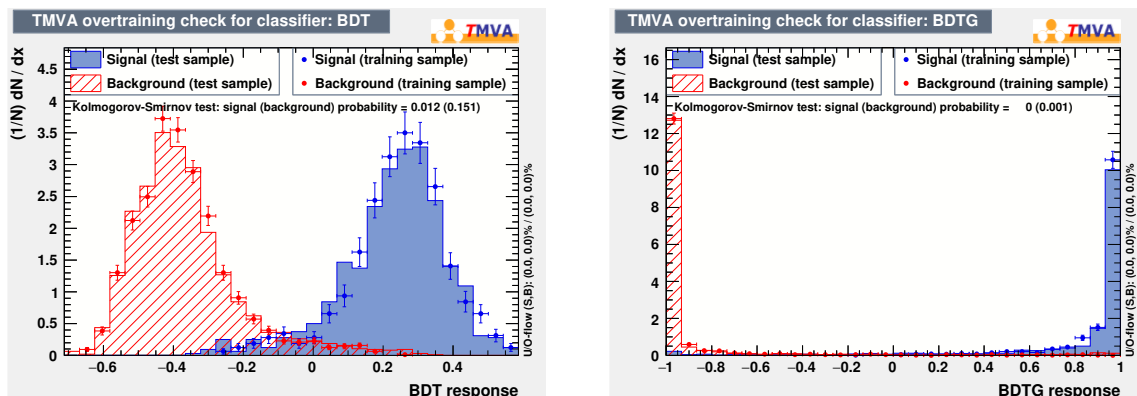
Figure 10. Distributions of the BDT (left) and BDTG (right) discriminants for the signal and background in training and test samples, for $t\bar{t}A$ events with $m_A = 40\,\text{GeV}$.

Following the pairing of jets and leptons, the reconstruction of the undetected neutrinos 3-momentum is performed by solving the following set of equations,

$$
\begin{aligned}
(p_{l^+} + p_\nu)^2 &= m_W^2, \\
(p_{l^-} + p_{\bar{\nu}})^2 &= m_W^2, \\
(p_{W^+} + p_b)^2 &= m_t^2, \\
(p_{W^-} + p_{\bar{b}})^2 &= m_t^2, \\
p_\nu^x + p_{\bar{\nu}}^x &= \not{E}^x, \\
p_\nu^y + p_{\bar{\nu}}^y &= \not{E}^y.
\end{aligned}
\tag{3.2}
$$

In the first four, relativistic mass constraints are imposed to signal and background events, by assuming the four-momentum of the W bosons ($p_{W^\pm}$), with masses set to m_W, are reconstructed using the charged leptons and neutrinos four-momentum $p_{\ell^\pm}$ and $p_{\nu(\bar{\nu})}$, respectively. The top quarks, t and $\bar{t}$, with masses set to m_t, are reconstructed with the b-quarks four-momentum, p_b and $p_{\bar{b}}$, correctly paired to the respective W^+ and W^-. In the last two equations, the 3-momentum x and y components of the undetected neutrinos (anti-neutrinos), p_ν^x ($p_{\bar{\nu}}^x$) and p_ν^y ($p_{\bar{\nu}}^y$), respectively, fully account for the x and y components of the missing transverse energy ($\not{E}$).

Since top quarks and W bosons have non-zero widths, their mass distributions follow Breit-Wigner probability distribution functions (*p.d.f.*s), with pole masses fixed to m_t and m_W, respectively. In order to reconstruct the neutrino and anti-neutrino four-momenta, we generate random top and anti-top quark masses from 1-dimensional parton level *p.d.f.*s, and generate, consistently, random $W^\pm$ masses, following 2-dimensional mass *p.d.f.*s of (m_{W^+}, m_t) and $(m_{W^-}, m_{\bar{t}})$. This ensures kinematic correlations are preserved when generating the top quark and W boson masses. We then solve the equations for all momentum components of the neutrinos. If no solution is found, the mass generation is repeated up-to a maximum of 500 trials. If there is still no solution, the event is discarded. Additionally, as the mass equations are of quadratic form, several solutions may exist for a single event. In order to choose the best one, a likelihood function is constructed using *p.d.f.*s from the

transverse momenta of the neutrinos, the top quarks and the $t\bar{t}$ system, respectively $P(p_{T_\nu})$, $P(p_{T_{\bar\nu}})$, $P(p_{T_t})$, $P(p_{T_{\bar t}})$, $P(p_{T_{t\bar t}})$, all obtained from parton level distributions. Furthermore, we consider the two dimensional mass $p.d.f.$ of the $t\bar{t}$ pair, $P(m_t, m_{\bar t})$, and the mass of the reconstructed Higgs, $P(m_\phi)$, obtained with truth-matching. The likelihood is defined according to

$$L_{t\bar{t}\phi} \propto \frac{1}{p_{T_\nu} p_{T_{\bar\nu}}} P(p_{T_\nu}) P(p_{T_{\bar\nu}}) P(p_{T_t}) P(p_{T_{\bar t}}) P(p_{T_{t\bar t}}) P(m_t, m_{\bar t}) P(m_\phi). \qquad (3.3)$$

A normalization factor $1/(p_{T_\nu} p_{T_{\bar\nu}})$ is applied in the likelihood because energy losses due to radiation emission and effects from detector resolutions will tend to increase the reconstructed neutrino four-momentum. This factor compensates for too extreme values of the neutrinos p_T, giving less weight to solutions of that type. We have checked, after event selection and considering only truth-matched signal events, that 66% to 73% of the total number of events are correctly reconstructed, corresponding to ϕ masses in the range 40 GeV to 300 GeV (for both scalar and pseudoscalar $t\bar{t}\phi$ signals). If truth-match is not applied, the reconstruction efficiency varies from 49% (51%) to 63% (62%), for scalars (pseudoscalars), in the same mass range. In this case, the number of times the reconstruction results in the same jet configuration as the one found with truth-match varies from 29% (31%) to 49% (55%) for the same mass range of scalar (pseudoscalar) signals. It is worth mentioning here that the current kinematic reconstruction nicely extends the one discussed in [21] to a wider mass range of scalar and pseudoscalar bosons with very similar performance numbers, if not better.

Figure 11 shows two-dimensional p_T distributions of the W^+ (top-left), the top quark (top-right), the $t\bar{t}$ system (bottom-left) and the Higgs boson (bottom-right) after kinematic reconstruction of $t\bar{t}H$ events, for $m_H = 40$ GeV. The correlation between the parton level (x-axis) and reconstructed (y-axis) p_T distributions, is clearly visible. The same behaviour is observed for the $t\bar{t}A$ signals, as well as for the other scalar boson masses considered. The main difference for these distributions as we increase the Higgs mass is that the density of points in higher p_T regions will also increase. The choice of the 40 GeV case was made for representation purposes only.

In figure 12, we show the neutrino reconstructed p_T versus the parton level value (left) and the distribution of the Higgs boson reconstructed masses, obtained with truth-matching, for several masses of the scalar boson (right). In spite of the wider spread of values in the neutrino p_T distribution, a clear correlation between the parton level and reconstructed p_T is observed.

Additional selection criteria were applied to events following the kinematical reconstruction (final selection cuts), to further increase the signal to background ratio. The depletion of $Z + 4$ jets and $Zb\bar{b} + 2$ jets backgrounds is accomplished by selecting events with a dilepton invariant mass ($m_{\ell^+\ell^-}$) outside a window around the Z boson mass ($m_Z = 91$ GeV). That is defined by $|m_{\ell^+\ell^-} - m_Z| > 10$ GeV. Most backgrounds, notably the $t\bar{t} + 3$ jets, are mitigated by selecting events with at least 3 b-jets. In figure 13, the expected number of events that survive the full selection criteria, for the different SM backgrounds is shown at the LHC and for an integrated luminosity of 100 fb^{-1}. The
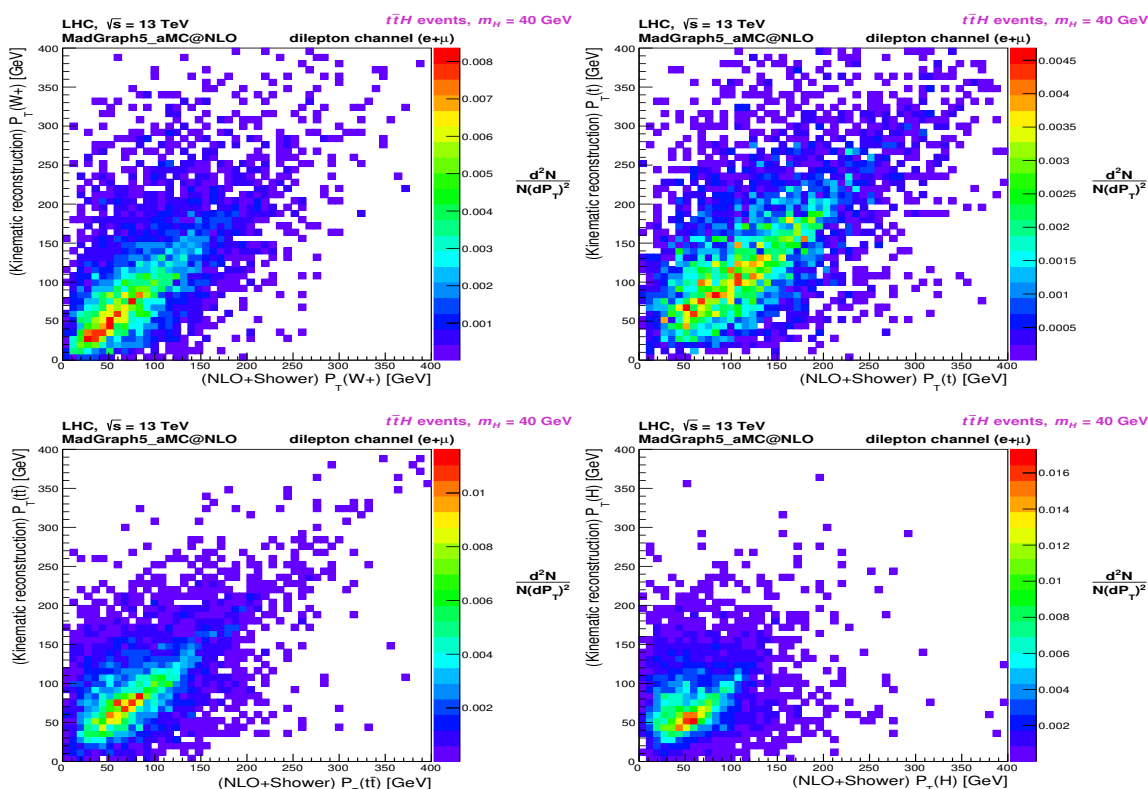
Figure 11. Two-dimensional distributions of p_T in $t\bar{t}H$ events (similar distributions can be found for the $t\bar{t}A$). The horizontal axes represent variables recorded at NLO+Shower, and the vertical axes represent the corresponding variables after kinematic reconstruction. Upper-left: distribution for W^+. A similar distribution is obtained for W^-, but is not shown here. Upper-right: distribution for t. A similar distribution is obtained for $\bar{t}$, but is not shown here. Lower-left: distribution for $t\bar{t}$. Lower-right: distribution for H. All distributions are shown for a Higgs mass of 40 GeV.
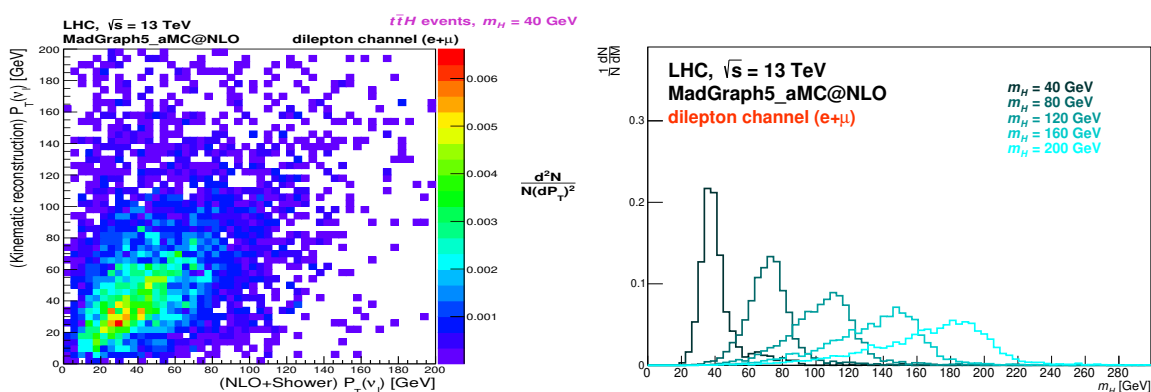

Figure 12. Two-dimensional distribution of the neutrino p_T in $t\bar{t}H$ events (left): the NLO+Shower p_T (x-axis) against the reconstructed p_T (y-axis) is shown. Distribution of the reconstructed Higgs boson mass with truth-matched jets in $t\bar{t}H$ events (right).
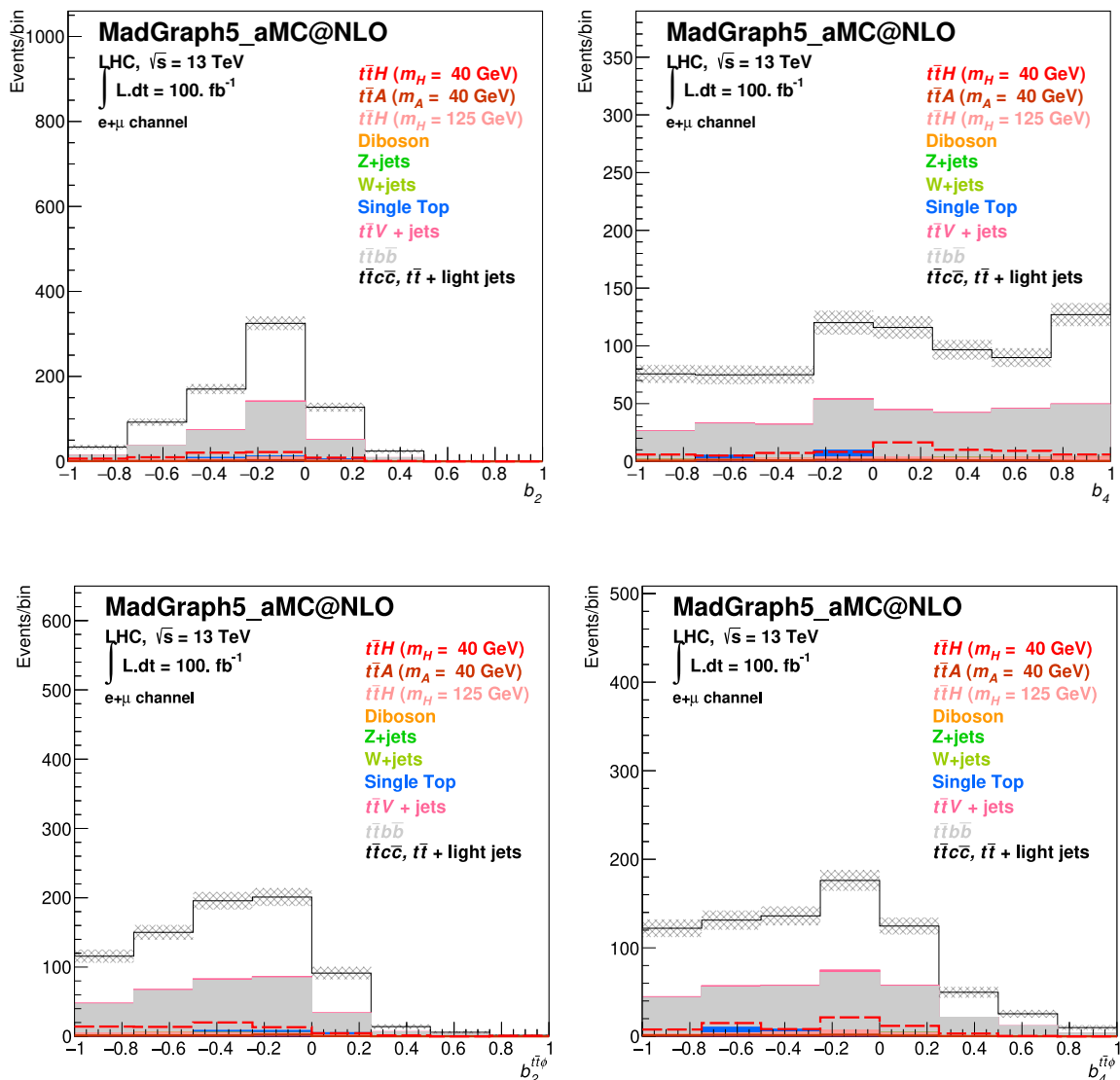
Figure 13. Expected number of background and signal events for the distributions b_2 (top left), b_4 (top right), $b_2^{t\bar{t}\phi}$ (bottom left) and $b_4^{t\bar{t}\phi}$ (bottom right), for a luminosity of $100\,\mathrm{fb}^{-1}$. Kinematic reconstruction and final selection cuts are considered.

distributions are compared to the CP-even and CP-odd signals, with $m_\phi = 40\,\mathrm{GeV}$, for different observables. The $Z+$jets includes the $Z + 4$ jets and the $Zb\bar{b}+2$ jets contributions. The $W+$jets, includes the contributions from $W + 4$ jets and $Wb\bar{b}+2$ jets. Diboson events are composed of the $WW + 3$ jets, $WZ + 3$ jets and $ZZ + 3$ jets backgrounds, $t\bar{t}c\bar{c}$, $t\bar{t} +$ light jets is the $t\bar{t} + 3$ jets process and $t\bar{t}H$ ($m_H = 125\,\mathrm{GeV}$) is the $t\bar{t}H_{\mathrm{SM}}$ process.

3.3 CLs results for different exclusion scenarios

In this section, CLs on the exclusion of scalar and pseudoscalar signals $t\bar{t}\phi$ ($\phi = H, A$) evaluated for different scenarios, are computed as a function of the LHC luminosity, up to the High-Luminosity Phase (HL-LHC). Several mass values of the ϕ boson are considered,

in the range 40–200 GeV. For larger scalar boson masses, the sensitivity with the current analysis is lost, because the total production cross sections for $t\bar{t}\phi$ is too small, for both the scalar and pseudoscalar bosons. For Higgs masses lower than 40 GeV, we saw a significant degradation of the reconstruction efficiency, for the resolved $t\bar{t}\phi$ analysis considered.

The b_2 and b_4 distributions are used to set the CLs evaluated in both the LAB and $t\bar{t}\phi$ centre-of-mass systems, for comparison. The contribution of all SM backgrounds is taken into account, normalized to the LHC luminosity, as well as the different signal hypotheses. For each scenario under study, one million pseudo experiments are generated, using bin-by-bin Poisson fluctuations around a mean value, which is set to the number of events in each individual bin of the distributions. The probability that a H_0 and an alternative H_1 hypothesis can describe the pseudo experiment is evaluated for each of them. The likelihood ratio of the H_1 and H_0 probabilities is used as test statistics, to compute the CLs with which hypothesis H_1 can be excluded assuming H_0 is true. The expected CLs for exclusion were calculated as a function of the integrated luminosity, from 100 to 3000 fb^{-1}, using the b_2 and b_4 observables. The calculation of the CLs follows the prescription set by [55, 56]. The different scenarios under consideration are:

- Scenario 1: Exclusion of the SM plus a new CP-even scalar particle, assuming the SM. In this case, H_0 is the SM only hypothesis,[2] while H_1 is the SM plus a new CP-even signal;

- Scenario 2: Exclusion of the SM plus a new CP-odd scalar particle, assuming the SM. In this case, H_0 is the SM only hypothesis, while H_1 is the SM plus a new CP-odd signal;

- Scenario 3: Exclusion of the SM plus a new CP-odd scalar particle, assuming the SM plus a new CP-even scalar particle of the same mass. In this case, H_0 is the SM plus a new CP-even signal hypothesis, while H_1 is the SM plus a new CP-odd signal;

- Scenario 4: SM exclusion, assuming the SM plus a new CP-even scalar particle. In this case, H_0 is the SM plus a new CP-even signal hypothesis, while H_1 is the SM only.

Other scenarios of interest could also be considered like, for instance, the exclusion of the SM plus a new CP-even scalar particle assuming the SM plus a new CP-odd scalar particle of the same mass, or the SM exclusion assuming the SM plus a new CP-odd scalar particle. As these scenarios were judged similar to some of the ones considered already, they were not shown in this paper.

In figure 14, we show the luminosity required for exclusion, at a given CL (2σ for the first three scenarios and 5σ for scenario 4), for each of the scenarios considered, as a function of the Higgs mass. If for a given scalar boson mass no points are shown, the exclusion is not possible, even at the end of HL-LHC ($\mathcal{L} = 3000$ fb^{-1}), for that mass. In each plot, the luminosity for exclusion is shown for the b_2 and b_4 variables.

---

[2]Consisting of diboson, Z + jets, W + jets, single top, $t\bar{t}$V + jets, $t\bar{t}b\bar{b}$, $t\bar{t}c\bar{c}$, $t\bar{t}$ + light jets and $t\bar{t}H$ ($m_H = 125$ GeV) events.

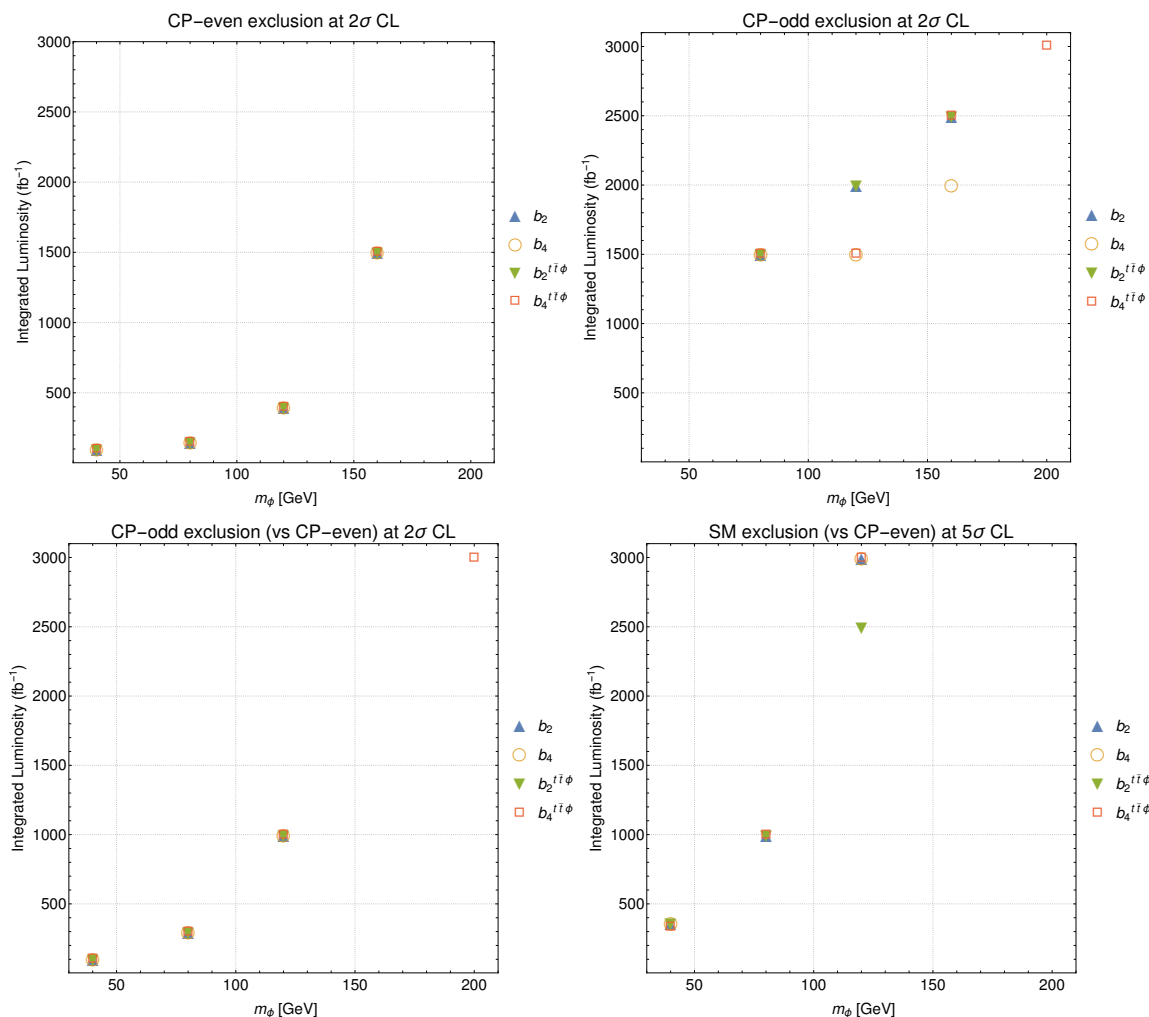**Figure 14.** Luminosity needed to exclude scenarios 1 (top left), 2 (top right) and 3 (bottom left) at the 2σ level, and scenario 4 (bottom right) at the 5σ level, as a function of the ϕ boson mass.

The top left of figure 14 (Scenario 1) tells us that, with the current LHC luminosity, we can exclude a CP-even scalar boson with CLs that exceed 2σ if its mass is $m_\phi \lesssim 80\,\text{GeV}$. For masses around the SM Higgs boson mass we require roughly $300\,\text{fb}^{-1}$ to achieve the 2σ CLs exclusion. This will be obtained during the incoming RUN's of the LHC. Masses of the CP-even scalar boson above $200\,\text{GeV}$ cannot be excluded even at the end of HL-LHC, with the dileptonic channel alone.

The top right of figure 14 (Scenario 2) shows that the exclusion CLs are quite different for the CP-odd case, greatly due to the reduced cross section when compared with the CP-even case. To exclude pseudoscalars at 2σ with respect to the SM, and with masses in the range $m_\phi = 80 - 200\,\text{GeV}$, a luminosity of at least $\sim 1500\,\text{fb}^{-1}$ is required at the LHC. For the $m_\phi = 40\,\text{GeV}$ case, the reconstruction efficiency is the lowest, thus the CLs are worse relative to the other CP-odd scalar boson masses.

If a new CP-even scalar is discovered (Scenario 3), a CP-odd exclusion is possible for the mass range considered, with diminishing CLs for an increasing ϕ boson mass, as shown

in the bottom left of figure 14. The exception for $m_\phi = 160\,\text{GeV}$, with much lower CLs, is due to both hypotheses presenting almost the same number of events distributed similarly in the variables considered, hence both hypotheses have similar results which degrades the sensitivity for exclusion. Without considering the uncertainties, the best variable is b_4 in the laboratory frame (see figure 19 of appendix A).

In the bottom right of figure 14 (Scenario 4), we show that the SM only hypothesis will be excluded at the 5σ CL at some stage of the LHC lifetime, for masses of the discovered new CP-even scalar below $m_\phi \lesssim 120\,\text{GeV}$. For the case of $m_\phi = 160\,\text{GeV}$, it can be excluded with almost 3σ at the end of HL-LHC (see figure 20 of appendix A).

The expected confidence levels, as a function of the integrated LHC luminosity, are shown in appendix A.

4 Interpretation in the framework of the C2HDM

We will now interpret the results in the framework of the C2HDM. We will briefly review the relevant aspects of the C2HDM to be used in the discussion (for a detailed description of the model see [17]). In the C2HDM the scalar potential is explicitly CP-violating and is invariant under a Z_2 symmetry $\Phi_1 \to \Phi_1, \Phi_2 \to -\Phi_2$, softly broken by the m_{12}^2 term,

$$
V = m_{11}^2|\Phi_1|^2 + m_{22}^2|\Phi_2|^2 - \left(m_{12}^2\,\Phi_1^\dagger\Phi_2 + \text{h.c.}\right) + \frac{\lambda_1}{2}(\Phi_1^\dagger\Phi_1)^2 + \frac{\lambda_2}{2}(\Phi_2^\dagger\Phi_2)^2
$$
$$
+\lambda_3(\Phi_1^\dagger\Phi_1)(\Phi_2^\dagger\Phi_2) + \lambda_4(\Phi_1^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1) + \left[\frac{\lambda_5}{2}(\Phi_1^\dagger\Phi_2)^2 + \text{h.c.}\right], \tag{4.1}
$$

where the doublets Φ_i $(i = 1, 2)$ develop real vacuum expectation values (VEVs) v_1 and v_2. All parameters are real except for m_{12}^2 and λ_5. We define $\tan\beta \equiv \frac{v_2}{v_1}$ and the rotation matrix that takes us from the gauge to the mass eigenstates is

$$
\begin{pmatrix} H_1 \\ H_2 \\ H_3 \end{pmatrix} = R \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix}, \tag{4.2}
$$

with

$$
R = \begin{pmatrix} c_1 c_2 & s_1 c_2 & s_2 \\ -(c_1 s_2 s_3 + s_1 c_3) & c_1 c_3 - s_1 s_2 s_3 & c_2 s_3 \\ -c_1 s_2 c_3 + s_1 s_3 & -(c_1 s_3 + s_1 s_2 c_3) & c_2 c_3 \end{pmatrix}, \tag{4.3}
$$

where $s_i = \sin\alpha_i$, $c_i = \cos\alpha_i$ $(i = 1, 2, 3)$, and

$$
-\pi/2 < \alpha_1 \le \pi/2, \qquad -\pi/2 < \alpha_2 \le \pi/2, \qquad -\pi/2 < \alpha_3 \le \pi/2. \tag{4.4}
$$

The Higgs boson masses are ordered such that $m_{H_1} \le m_{H_2} \le m_{H_3}$. In the C2HDM, there are four types of Yukawa models. However, the top Yukawa couplings are the same in all four types and therefore this discussion is valid for all of them. The Yukawa Lagrangian for the up quarks in all four types has the form

$$
\mathcal{L}_{Yi} = -\frac{m_f}{v}\bar{\psi}_f \left[\frac{R_{i2}}{s_\beta} - i\frac{R_{i3}}{t_\beta}\gamma_5\right]\psi_f H_i, \tag{4.5}
$$

where ψ_f denotes the fermion fields with mass m_f, i is the scalar index, $v^2 = v_1^2 + v_2^2$ (fixed by the W boson mass) and $t_\beta = \frac{v_2}{v_1}$.

What we want to understand now is what can be concluded for the parameter space of this specific model once we have either a measurement or an exclusion for a given ϕ boson mass (and luminosity). We will just analyse a simple situation where the 125 GeV Higgs is H_2 and the lightest Higgs is H_1, with a mass below 125 GeV. We start by mapping equation (4.5) into equation (2.1).,

$$\begin{cases} \kappa_t \cos\alpha = \dfrac{s_1 c_2}{s_\beta} \\ \kappa_t \sin\alpha = -\dfrac{s_2}{t_\beta} \end{cases} \qquad s_\beta^2 \kappa_t^2 = s_1^2 c_2^2 + s_2^2 c_\beta^2. \tag{4.6}$$

The values of κ_t and α are free to vary in their allowed range (taking into account available theoretical and experimental constraints) because no scalar was found below 125 GeV. Let us start by noting that $\sin\alpha = 0$ and $\sin\alpha_2 = 0$ are equivalent. This means that the CP-even limit is obtained unambiguously. The $H_1 VV$ coupling, where V is a vector boson, is proportional to $\cos\alpha_2$ which vanishes for $\alpha_2 = \pi/2$.

What will be measured or constrained in the experiment is α and κ_t. Also, the limits in this work were set for the pure scalar and pure pseudoscalar scenarios. For these scenarios we get, respectively,

$$\begin{cases} \sin\alpha = 0 \implies \kappa_t = \pm\dfrac{s_1}{s_\beta}, \\ \cos\alpha = 0 \implies \kappa_t = \pm\dfrac{s_2}{t_\beta} \ (\text{if } s_1 = 0) \quad \text{or} \quad \kappa_t = \pm\dfrac{1}{t_\beta} \ (\text{if } c_2 = 0), \end{cases} \tag{4.7}$$

and a measurement or limit on κ_t will set a limit on the parameters of the model. For the particular scenario where $c_2 = 0$ we obtain a limit on $\tan\beta$. Because $\tan\beta$ is already constrained to be above one by low energy physics measurements (see [17]) information can only be added if we increase the limit. This is in fact the case, if the limit is for instance $\kappa_t \leq 1/10$ we get $\tan\beta \geq 10$ ($c_2 = 0$). With the same limit for κ_t, in the remaining two scenarios the bound on the parameters is $s_1 \leq 1/10$ ($s_2 = 0$) and $s_2 \leq t_\beta/10$ ($s_1 = 0$). In figure 15 we present the luminosity needed to exclude κ_t at the 2σ level for the pure CP-even case (scenario 1), for a CP-even scalar boson mass of 40 GeV. Note that this is the most favourable scenario for discovery (and for exclusion). As can be seen the value of κ_t attainable is close to 0.3 by the end of the LHC run. However, because this is a study using the dileptonic final state, we can expect to reach values of κ_t of the order of 1/10 for an analysis which includes all other decay channels.

The next question to ask is what are the constraints on the parameter space in scenarios where one is either close to CP-even or to the CP-odd scenario. In figure 16 (left), we present the allowed points in the C2HDM parameter space (c_1 vs. s_2) if a measurement of κ_t and $\sin\alpha$ is in the ranges $0.1 \leq \kappa_t \leq 1.2$ and $0.1 \leq \sin\alpha \leq 0.2$. We also force $1 \leq \tan\beta \leq 10$. In the top plot we see the variation with κ_t, in the middle with $\sin\alpha$ and on the bottom with $\tan\beta$. This is the case where we are close to the CP-even limit.
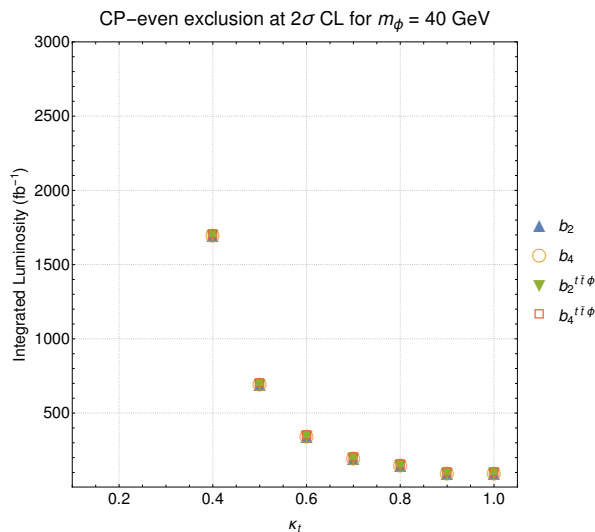
Figure 15. Luminosity needed to exclude κ_t at the 2σ level for the pure CP-even case (scenario 1), for a CP-even scalar boson mass of $40\,\text{GeV}$.

In figure 16 (right), we present the scenario when we are close to the CP-odd limit, that is $0.8 \leq \sin\alpha \leq 0.9$. The most striking point is that although in each case we are closer to one of the limits, CP-even or CP-odd, the allowed parameter space is quite large and we clearly need some other sources of measurement to constraint the parameter space.

5 Conclusions

In this paper we examine the possibility of determining the CP nature of the heavier quarks (b and t quarks) Yukawa interactions with a generic scalar boson ϕ, in $q\bar{q}\phi$ production at the LHC. We found that strategies to achieve this goal suggested in the literature for the case of the top quark do not work for the bottom quark, even at parton level. This was also confirmed for very light Higgs bosons with masses of the order of $10\,\text{GeV}$. The underlying reason is the CP-asymmetric term, responsible for the CP-asymmetries, which is proportional to m_f^2. Hence, this term is only meaningful when the fermion mass is of the order of m_ϕ.

Previous works established that several kinematic distributions for $t\bar{t}\phi$ are sensitive to the CP-components of the top quark Yukawa coupling. These studies assumed $m_\phi = 125\,\text{GeV}$ and an extension of the study to other masses of scalar and pseudoscalar Higgs bosons was still missing in the literature. In this paper, we investigate the dilepton final states of $t\bar{t}\phi$ (with $\phi = H, A$) for several masses of the CP-even or CP-odd boson (ϕ). We found that for the masses considered, there is still a good level of discrimination between scalar and pseudoscalar Yukawa interactions, at parton level. However, the differences between those cases become smaller as the Higgs mass increases, and vanish around $m_\phi = 450\,\text{GeV}$.
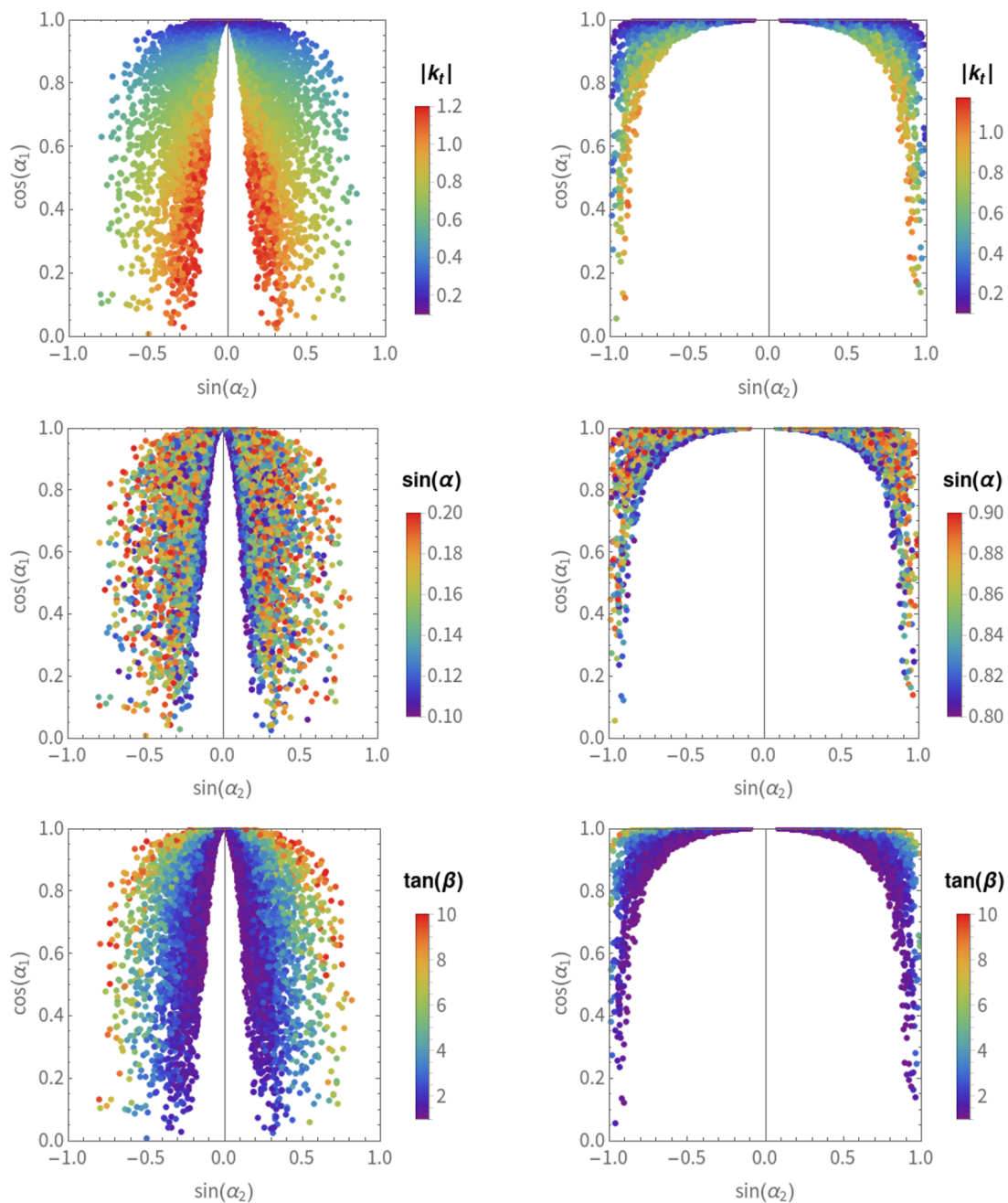
Figure 16. Points allowed in the plane c_1 vs. s_2 for $0.1 \leq \kappa_t \leq 1.2$ and $1 \leq \tan\beta \leq 10$. In the left we impose $0.1 \leq \sin\alpha \leq 0.2$ (CP-even like) and in the right we impose $0.8 \leq \sin\alpha \leq 0.9$ (CP-odd like). On the top, we color superimpose κ_t, in the middle $\sin\alpha$ and on the bottom $\tan\beta$.

A full kinematic reconstruction was applied to signal and background events, to reconstruct the four momenta of the undetected neutrinos, allowing to estimate the experimental sensitivity of the CP-search. CLs are presented for the exclusion of several scenarios as a function of the luminosity, for different Higgs boson masses. Generally, it is shown that the required luminosity for exclusion at a given CL increases with the ϕ boson mass. Given the current LHC luminosity, of $150\,\mathrm{fb}^{-1}$, exclusion of the SM plus a pure CP-even Higgs with masses of 40 and $80\,\mathrm{GeV}$ and SM couplings, assuming the SM only, is already possible. For $m_H > 200\,\mathrm{GeV}$, CP-searches will require the inclusion of additional channels. We also found that the SM plus a CP-odd scalar exclusion, assuming the SM only, is harder than the CP-even exclusion for CP-odd Higgs masses up to $160\,\mathrm{GeV}$. For higher masses, the opposite is true. If a new Higgs is found, we have enough sensitivity to exclude the possibility of the scalar being purely CP-odd in the explored mass range, again assuming SM-like couplings. In this work, only the dileptonic final states of the $t\bar{t}\phi$ system is considered in the CLs evaluation, at the LHC. A natural follow up would be to combine several $t\bar{t}\phi$ decay channels, to further improve the results obtained in this paper.

Finally, the impact of a new discovery was discussed for the C2HDM. If a new particle is found to be an exact CP-eigenstate, this will impose further constrains on typical 2HDM parameters such as $\tan\beta$. In case the new particle is just close to either the CP-even or the CP-odd scenarios, the allowed parameter space would still be very large and will require other measurements to further constrain it.

Acknowledgments

A   CLs in the dileptonic channel

In this section, figures 17 to 20 show the expected exclusion CLs, for the different scenarios and Higgs masses considered in section 3.3, using only the dileptonic channel. In each plot, the exclusion CLs using the b_2 and b_4 variables, as a function of the integrated luminosity, are shown for each given mass of the ϕ boson.

In figure 17, the exclusion of the SM plus a new CP-even scalar particle, assuming the SM, is represented (see section 3.3 for details).

In figure 18, the exclusion of the SM plus a new CP-odd scalar particle, assuming the SM, is represented (see section 3.3 for details).

In figure 19, the exclusion of the SM plus a new CP-odd scalar particle, assuming the SM plus a new CP-even scalar particle of the same mass, is represented (see section 3.3 for details).

In figure 20, the SM exclusion, assuming the SM plus a new CP-even scalar particle, is represented (see section 3.3 for details).
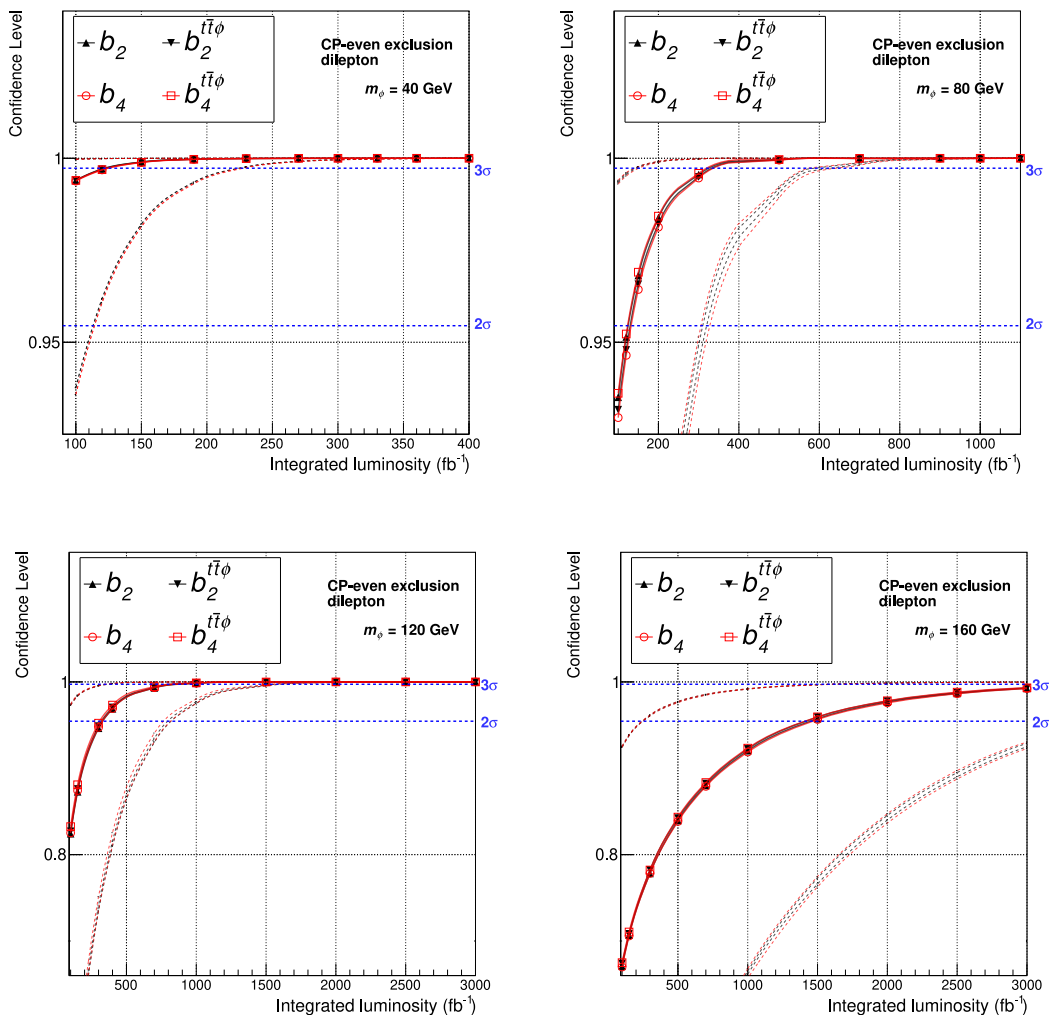
Figure 17. Expected CLs for CP-even exclusion assuming the SM (scenario 1), as a function of the integrated luminosity. The luminosity is increased until CL= 1 or $\mathcal{L} = 3000\,\mathrm{fb}^{-1}$. Each plot represents a different ϕ boson mass. The vertical axis scale also changes, for visualization purposes. The red and black dashed lines represent the $\pm 1\sigma$ bands for the variables considered. The blue dashed lines represent the 2σ or 3σ CL lines.
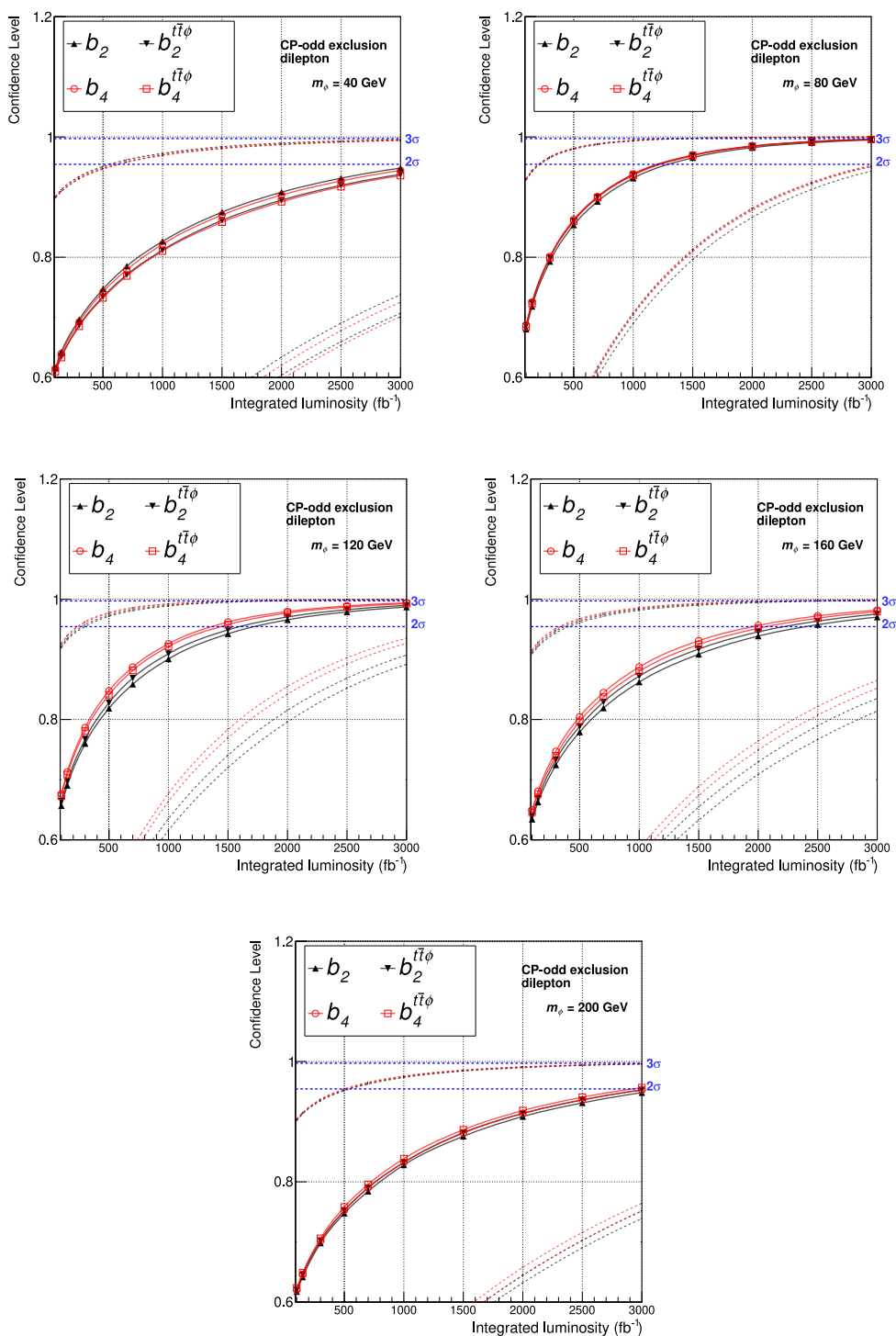
Figure 18. Expected CLs for CP-odd exclusion assuming the SM (scenario 2), as a function of the integrated luminosity.
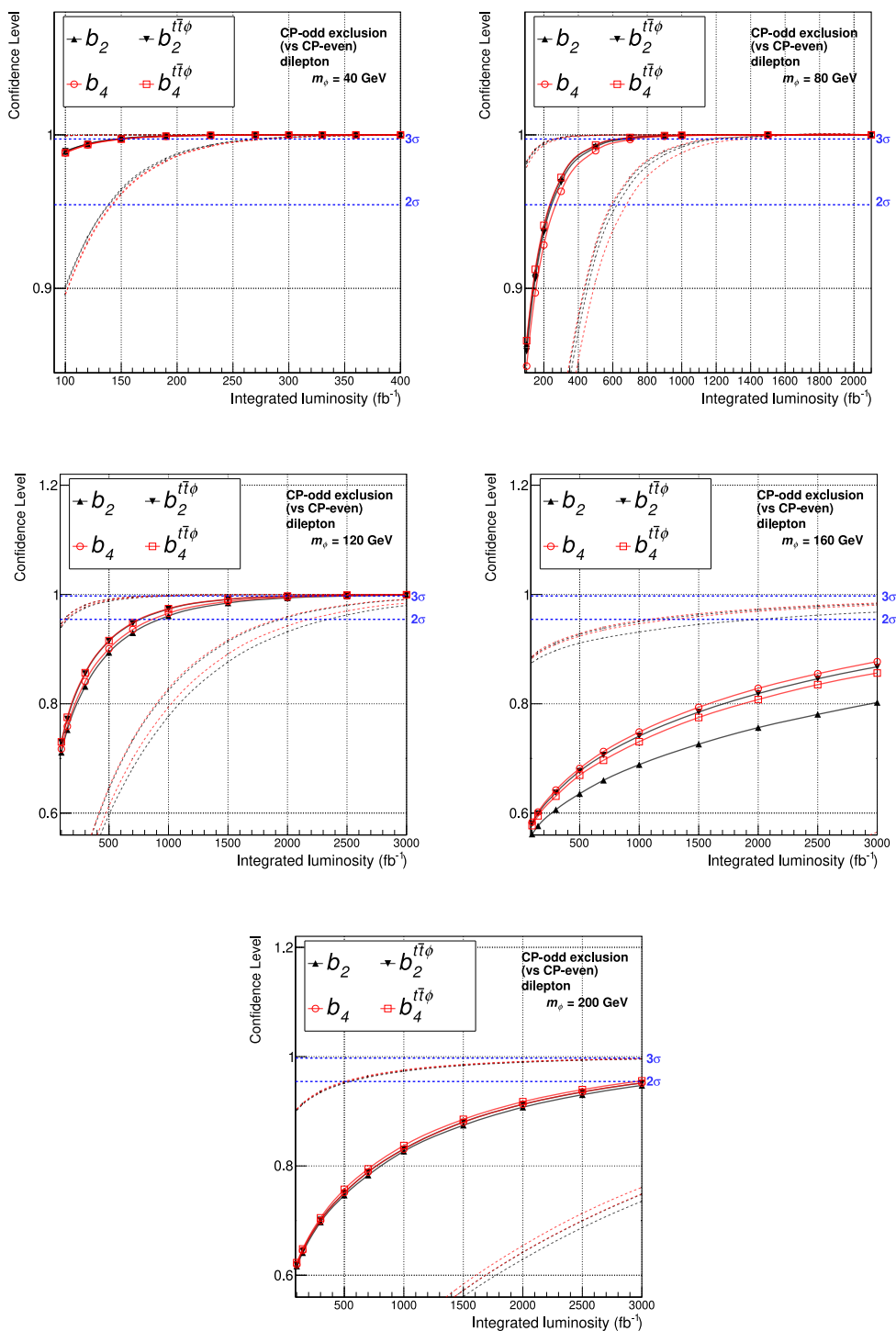
Figure 19. Expected CLs for CP-odd exclusion assuming the SM plus a new CP-even scalar particle (scenario 3), as a function of the integrated luminosity.
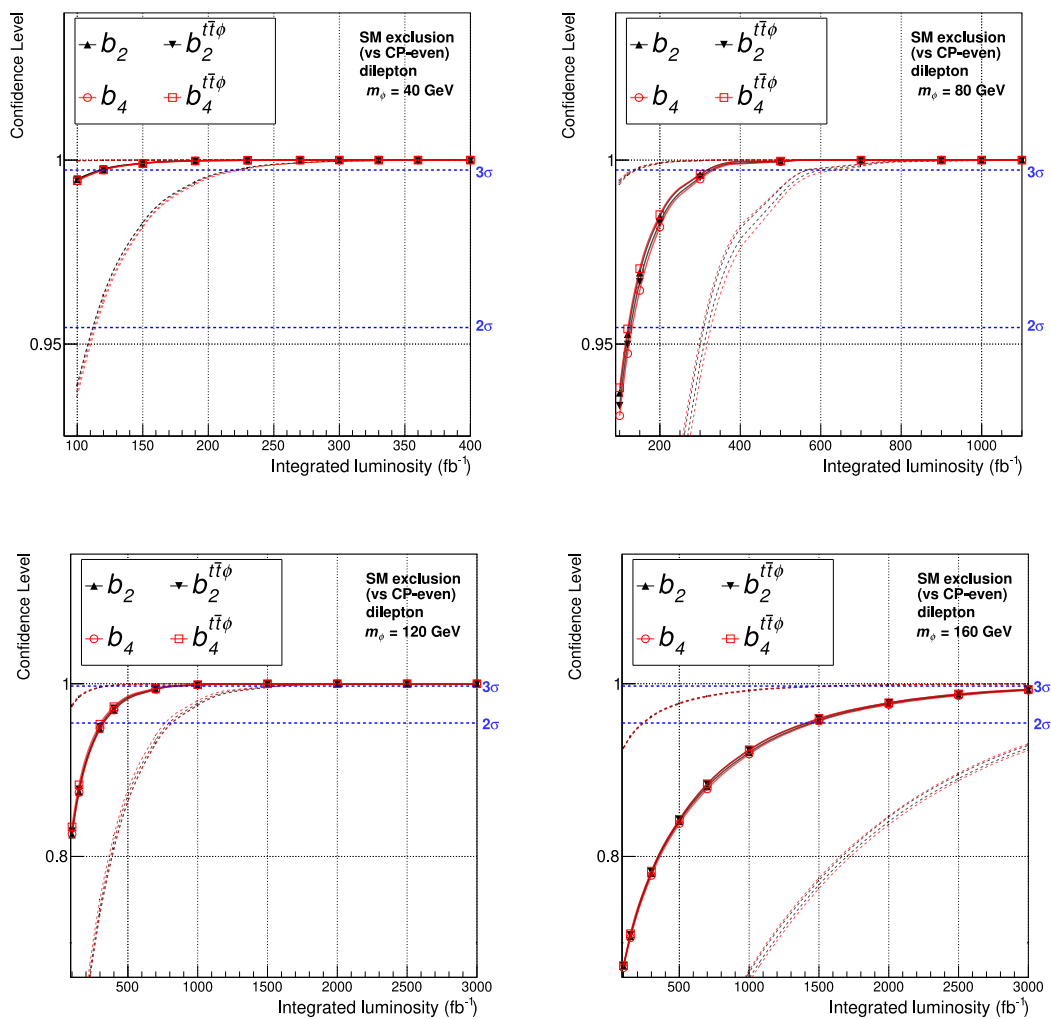
Figure 20. Expected CLs for SM exclusion assuming the SM plus a new CP-even scalar particle (scenario 4), as a function of the integrated luminosity.

References

[1] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B 716** (2012) 1 [arXiv:1207.7214] [inSPIRE].

[2] CMS collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett.* **B 716** (2012) 30 [arXiv:1207.7235] [inSPIRE].

[3] A.D. Sakharov, *Violation of CP Invariance, C asymmetry and baryon asymmetry of the universe*, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32 [inSPIRE].

[4] T.D. Lee, *A Theory of Spontaneous T Violation*, *Phys. Rev.* **D 8** (1973) 1226 [inSPIRE].

[5] I.F. Ginzburg, M. Krawczyk and P. Osland, *Two-Higgs-Doublet Models with CP-violation*, CERN-TH-2002-330 (2002) [hep-ph/0211371] [inSPIRE].

[6] W. Khater and P. Osland, *CP violation in top quark production at the LHC and two Higgs doublet models*, *Nucl. Phys.* **B 661** (2003) 209 [hep-ph/0302004] [inSPIRE].

[7] A.W. El Kaffas, P. Osland and O.M. Ogreid, *CP violation, stability and unitarity of the two Higgs doublet model*, *Nonlin. Phenom. Complex Syst.* **10** (2007) 347 [hep-ph/0702097] [inSPIRE].

[8] B. Grzadkowski and P. Osland, *Tempered Two-Higgs-Doublet Model*, *Phys. Rev.* **D 82** (2010) 125026 [arXiv:0910.4068] [inSPIRE].

[9] A. Arhrib, E. Christova, H. Eberl and E. Ginina, *CP violation in charged Higgs production and decays in the Complex Two Higgs Doublet Model*, *JHEP* **04** (2011) 089 [arXiv:1011.6560] [inSPIRE].

[10] A. Barroso, P.M. Ferreira, R. Santos and J.P. Silva, *Probing the scalar-pseudoscalar mixing in the 125 GeV Higgs particle with current data*, *Phys. Rev.* **D 86** (2012) 015022 [arXiv:1205.4247] [inSPIRE].

[11] S. Inoue, M.J. Ramsey-Musolf and Y. Zhang, *CP-violating phenomenology of flavor conserving two Higgs doublet models*, *Phys. Rev.* **D 89** (2014) 115023 [arXiv:1403.4257] [inSPIRE].

[12] K. Cheung, J.S. Lee, E. Senaha and P.-Y. Tseng, *Confronting Higgcision with Electric Dipole Moments*, *JHEP* **06** (2014) 149 [arXiv:1403.4775] [inSPIRE].

[13] D. Fontes, J.C. Romão and J.P. Silva, *$h \to Z\gamma$ in the complex two Higgs doublet model*, *JHEP* **12** (2014) 043 [arXiv:1408.2534] [inSPIRE].

[14] D. Fontes, J.C. Romão, R. Santos and J.P. Silva, *Large pseudoscalar Yukawa couplings in the complex 2HDM*, *JHEP* **06** (2015) 060 [arXiv:1502.01720] [inSPIRE].

[15] C.-Y. Chen, S. Dawson and Y. Zhang, *Complementarity of LHC and EDMs for Exploring Higgs CP-violation*, *JHEP* **06** (2015) 056 [arXiv:1503.01114] [inSPIRE].

[16] M. Mühlleitner, M.O.P. Sampaio, R. Santos and J. Wittbrodt, *Phenomenological Comparison of Models with Extended Higgs Sectors*, *JHEP* **08** (2017) 132 [arXiv:1703.07750] [inSPIRE].

[17] D. Fontes, M. Mühlleitner, J.C. Romão, R. Santos, J.P. Silva and J. Wittbrodt, *The C2HDM revisited*, *JHEP* **02** (2018) 073 [arXiv:1711.09419] [INSPIRE].

[18] J.F. Gunion and X.-G. He, *Determining the CP nature of a neutral Higgs boson at the LHC*, *Phys. Rev. Lett.* **76** (1996) 4468 [hep-ph/9602226] [INSPIRE].

[19] F. Boudjema, R.M. Godbole, D. Guadagnoli and K.A. Mohan, *Lab-frame observables for probing the top-Higgs interaction*, *Phys. Rev.* **D 92** (2015) 015019 [arXiv:1501.03157] [INSPIRE].

[20] S.P. Amor dos Santos et al., *Angular distributions in $t\bar{t}H(H \to b\bar{b})$ reconstructed events at the LHC*, *Phys. Rev.* **D 92** (2015) 034021 [arXiv:1503.07787] [INSPIRE].

[21] S. Amor Dos Santos et al., *Probing the CP nature of the Higgs coupling in $t\bar{t}h$ events at the LHC*, *Phys. Rev.* **D 96** (2017) 013004 [arXiv:1704.03565] [INSPIRE].

[22] F. Demartin, F. Maltoni, K. Mawatari, B. Page and M. Zaro, *Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction*, *Eur. Phys. J.* **C 74** (2014) 3065 [arXiv:1407.5089] [INSPIRE].

[23] 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 [arXiv:1603.03632] [INSPIRE].

[24] K. Cheung, A. Jueid, Y.-N. Mao and S. Moretti, *The 2-Higgs-Doublet Model with Soft CP-violation Confronting Electric Dipole Moments and Colliders*, arXiv:2003.04178 [INSPIRE].

[25] S. Berge, W. Bernreuther and J. Ziethe, *Determining the CP parity of Higgs bosons at the LHC in their tau decay channels*, *Phys. Rev. Lett.* **100** (2008) 171605 [arXiv:0801.2297] [INSPIRE].

[26] S. Berge and W. Bernreuther, *Determining the CP parity of Higgs bosons at the LHC in the tau to 1-prong decay channels*, *Phys. Lett.* **B 671** (2009) 470 [arXiv:0812.1910] [INSPIRE].

[27] S. Berge, W. Bernreuther, B. Niepelt and H. Spiesberger, *How to pin down the CP quantum numbers of a Higgs boson in its tau decays at the LHC*, *Phys. Rev.* **D 84** (2011) 116003 [arXiv:1108.0670] [INSPIRE].

[28] S. Berge, W. Bernreuther and S. Kirchner, *Determination of the Higgs CP-mixing angle in the tau decay channels at the LHC including the Drell-Yan background*, *Eur. Phys. J.* **C 74** (2014) 3164 [arXiv:1408.0798] [INSPIRE].

[29] S. Berge, W. Bernreuther and S. Kirchner, *Prospects of constraining the Higgs boson's CP nature in the tau decay channel at the LHC*, *Phys. Rev.* **D 92** (2015) 096012 [arXiv:1510.03850] [INSPIRE].

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

[31] D. Fontes, M. Mühlleitner, J.C. Romão, R. Santos, J.P. Silva and J. Wittbrodt, *Summary of "Extended Higgs Sector" group activities*, Talk presented by D. Fontes at *The 16th Workshop of the LHC Higgs Cross Section Working Group*, October, 2019.

[32] D. Fontes, M. Mühlleitner, J.C. Romão, R. Santos, J.P. Silva and J. Wittbrodt, *C2HDM Higgs pair production*, Talk presented by R. Santos at *The 16th Workshop of the LHC Higgs Cross Section Working Group*, October, 2019.

[33] A. Broggio, A. Ferroglia, M.C.N. Fiolhais and A. Onofre, *Pseudoscalar couplings in t$\bar{t}$H production at NLO+NLL accuracy*, *Phys. Rev.* **D 96** (2017) 073005 [arXiv:1707.01803] [INSPIRE].

[34] A. Broggio, A. Ferroglia, R. Frederix, D. Pagani, B.D. Pecjak and I. Tsinikos, *Top-quark pair hadroproduction in association with a heavy boson at NLO+NNLL including EW corrections*, *JHEP* **08** (2019) 039 [arXiv:1907.04343] [INSPIRE].

[35] A. Kulesza, L. Motyka, D. Schwartländer, T. Stebel and V. Theeuwes, *Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy*, *Eur. Phys. J.* **C 79** (2019) 249 [arXiv:1812.08622] [INSPIRE].

[36] A. Broggio, A. Ferroglia, B.D. Pecjak and L.L. Yang, *NNLL resummation for the associated production of a top pair and a Higgs boson at the LHC*, *JHEP* **02** (2017) 126 [arXiv:1611.00049] [INSPIRE].

[37] A. Broggio, A. Ferroglia, B.D. Pecjak, A. Signer and L.L. Yang, *Associated production of a top pair and a Higgs boson beyond NLO*, *JHEP* **03** (2016) 124 [arXiv:1510.01914] [INSPIRE].

[38] F. Maltoni, D. Pagani and I. Tsinikos, *Associated production of a top-quark pair with vector bosons at NLO in QCD: impact on t$\bar{t}$H searches at the LHC*, *JHEP* **02** (2016) 113 [arXiv:1507.05640] [INSPIRE].

[39] D. Azevedo, A. Onofre, F. Filthaut and R. Gonçalo, *CP tests of Higgs couplings in t$\bar{t}$h semileptonic events at the LHC*, *Phys. Rev.* **D 98** (2018) 033004 [arXiv:1711.05292] [INSPIRE].

[40] A. Ferroglia, M.C.N. Fiolhais, E. Gouveia and A. Onofre, *Role of the t$\bar{t}$h rest frame in direct top-quark Yukawa coupling measurements*, *Phys. Rev.* **D 100** (2019) 075034 [arXiv:1909.00490] [INSPIRE].

[41] A.V. Gritsan, R. Röntsch, M. Schulze and M. Xiao, *Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques*, *Phys. Rev.* **D 94** (2016) 055023 [arXiv:1606.03107] [INSPIRE].

[42] T. Ghosh, R. Godbole and X. Tata, *Determining the spacetime structure of bottom-quark couplings to spin-zero particles*, *Phys. Rev.* **D 100** (2019) 015026 [arXiv:1904.09895] [INSPIRE].

[43] P. Artoisenet et al., *A framework for Higgs characterisation*, *JHEP* **11** (2013) 043 [arXiv:1306.6464] [INSPIRE].

[44] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: Going Beyond*, *JHEP* **06** (2011) 128 [arXiv:1106.0522] [INSPIRE].

[45] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys.* **B 867** (2013) 244 [arXiv:1207.1303] [INSPIRE].

[46] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *JHEP* **03** (2013) 015 [arXiv:1212.3460] [INSPIRE].

[47] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[48] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[49] S. Frixione, F. Stoeckli, P. Torrielli, B.R. Webber and C.D. White, *The MCaNLO 4.0 Event Generator*, `arXiv:1010.0819` [ɪɴSPIRE].

[50] DELPHES 3 collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [`arXiv:1307.6346`] [ɪɴSPIRE].

[51] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [`arXiv:1111.6097`] [ɪɴSPIRE].

[52] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [`arXiv:0802.1189`] [ɪɴSPIRE].

[53] E. Conte, B. Fuks and G. Serret, *MadAnalysis 5, A User-Friendly Framework for Collider Phenomenology*, *Comput. Phys. Commun.* **184** (2013) 222 [`arXiv:1206.1599`] [ɪɴSPIRE].

[54] A. Hocker et al., *TMVA — Toolkit for Multivariate Data Analysis*, `physics/0703039` [ɪɴSPIRE].

[55] A.L. Read, *Presentation of search results: The CL(s) technique*, *J. Phys.* **G 28** (2002) 2693 [ɪɴSPIRE].

[56] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth.* **A 434** (1999) 435 [`hep-ex/9902006`] [ɪɴSPIRE].