Probing the CP nature of the Higgs coupling in $t\bar{t}h$ events at the LHC

D. Azevedo[1,2], A. Onofre[2], F. Filthaut[3,4], R. Gonçalo[5]

[1] *Department of Physics and Astronomy, Universidade do Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal*
[2] *LIP, Departamento de Física, Universidade do Minho, 4710-057 Braga, Portugal*
[3] *Department of Physics, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, Netherlands*
[4] *Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands*
[5] *LIP, Av. Prof. Gama Pinto, n°2, 1649-003 Lisboa, Portugal*

Abstract

The CP nature of the Higgs coupling to top quarks is addressed in this article, which focuses on single charged lepton final states of $t\bar{t}h$ events produced in proton-proton collisions at the LHC. Pure scalar ($h = H$) and pseudo-scalar ($h = A$) Higgs boson signal events, generated with MadGraph5_aMC@NLO, are fully reconstructed with a kinematic fit. Novel angular distributions of the decay products, as well as CP-sensitive asymmetries, are exploited to separate and gain sensitivity to possible pseudo-scalar components of the Higgs boson and reduce the contribution from the dominant irreducible background $t\bar{t}b\bar{b}$. Significant differences were found between the pure CP-even and -odd signal hypotheses, as well as with respect to the Standard Model (SM) background. Such differences survive the full kinematic reconstruction of the events, allowing to define optimal observables for a global fit of the Higgs couplings parameters. A dedicated analysis is applied to efficiently identify signal events and reject as much as possible the Standard Model expected background at the LHC. The results obtained were compared with a similar analysis in the dilepton channel. We show that the single lepton channel is more promising overall and can be used in combination for CP violation searches.

INTRODUCTION

The independent observation by ATLAS [1] and CMS [2] of a new scalar boson, predicted by the electroweak symmetry breaking mechanism [3], with a mass of approximately 125 GeV, triggered the discovery of the Higgs particle in July 2012. This motivated thorough analyses of the Higgs properties at the LHC, which have shown to be in agreement with the SM predictions [4]. Nevertheless, the SM does not solve the matter-antimatter asymmetry problem, which demands new sources of CP-violation beyond the SM (BSM). These sources can be accommodated through extended Higgs sectors, like the 2-Higgs doublet model (2HDM) by T. D. Lee [5] or the 3HDM by Weinberg [6], where the Higgs boson(s) have no definite CP quantum number resulting in a two component, one CP-even and one CP-odd, Yukawa coupling with the fermions (see for instance [7]).

The ATLAS and CMS collaborations have conducted analyses to measure the Higgs' spin and parity quantum numbers through its decays to photons, ZZ and WW, and through VH ($V = W, Z$) associated production [8–10]. The results agree with a scalar (spin 0), charge-parity even boson. The pure CP-odd case has been excluded at the 99.98% confidence level (CL). Nevertheless, there is the possibility of CP mixture in the Yukawa sector which remains to be probed directly. The top quark is expected to have the largest Yukawa coupling with the Higgs, considering only fermions, but so far sensitivity to this coupling was measured only indirectly through loop effects from $gg \to h$ and $h \to \gamma\gamma$. These processes suffer from large systematic uncertainties and require the assumption of no BSM contributions. The associated production of the Higgs boson with a pair of top quarks ($t\bar{t}h$)[11] overcomes this problem since it allows the direct measurement of top Yukawa couplings, as well as CP sensitivity due to its kinematic properties [20, 21, 44].

The main background for $t\bar{t}h$ events at the LHC is $pp \to t\bar{t}+$jets. If the main Higgs decay channel ($h \to b\bar{b}$) is considered for analysis, $t\bar{t}b\bar{b}$ is a challenging irreducible background. Several $t\bar{t}h$ decay channels have been studied [12–17]. Due to the complexity and huge spectrum of backgrounds this channel is particularly challenging to study at the LHC. Nonetheless, remarkable sensitivities have been reached by ATLAS and CMS collaborations with expected upper limits at 95% CL for the $t\bar{t}H$ signal strength, μ[1], below 2 in the background-only scenario. The best-fit values obtained for μ were 1.7 ± 0.8 by ATLAS [12] and 2.8 ± 1.0 by CMS [16]. Combined results from both collaborations and from the various Higgs analyses were used to fit the signal strengths of five Higgs production processes, while assuming SM-like Higgs branching ratios [18]. The best-fit value obtained for $\mu(t\bar{t}H)$ was $2.3^{+0.7}_{-0.6}$.

In this article, we present the single lepton final state of $t\bar{t}$ with the Higgs boson decaying through $h \to b\bar{b}$. The single lepton final state constitutes about 43.8% of the $t\bar{t}$ decays [19] making it a natural candidate for study. We search for deviations from the SM nature of the Higgs boson by comparing the kinematics of $t\bar{t}h$ signal samples with SM Higgs boson ($h = H$ and $J^{\mathrm{CP}} = 0^+$) to samples of $t\bar{t}h$ signal with pure pseudo-scalar Higgs boson ($h = A$ and $J^{\mathrm{CP}} = 0^-$). To do so, we consider the general

[1] The signal strength is defined as the ratio of the measured cross section, $\sigma \times \mathrm{Br}$, by the SM expectation, $(\sigma \times \mathrm{Br})_{\mathrm{SM}}$, $\mu = \frac{\sigma \times \mathrm{Br}}{(\sigma \times \mathrm{Br})_{\mathrm{SM}}}$

Yukawa coupling, which reads

$$\mathcal{L} = \kappa y_t \bar{t}(\cos\alpha + i\gamma_5 \sin\alpha) t h \quad (1)$$

where y_t is the SM Higgs Yukawa coupling and α represents a CP phase. The SM interaction is recovered for $|\cos\alpha| = 1$, while the pure pseudoscalar is obtained by setting $\cos\alpha = 0$.

Several observables in $t\bar{t}h$ events, sensitive to the CP nature of the top Yukawa coupling, have been proposed from which we will study in detail the ones presented in [20–22]. More general observables are obtained from the particles at production (t, $\bar{t}$ and h), only accessible experimentally through a reconstruction algorithm.

The four-momenta of the intermediary particles and of the undetected neutrino from the leptonically decaying W-boson are reconstructed by a full kinematic fit. In turn, a large set of angular observables is presented. Even after parton showering, detector simulation, event selection and event reconstruction the information in the matrix elements partially survives. Some studies [23, 24] consider that background discrimination can be done through angular distributions, since the spins of h in signal and g in $t\bar{t}b\bar{b}$ background (g being a gluon which splits into $b\bar{b}$) are different. In [22] a set of interesting observables for that effect are presented. Additional observables introduced in this article will be shown to have similar discriminating power. We will start by taking into account the irreducible $t\bar{t}b\bar{b}$ dominant background without a highly-optimized reconstruction method. Then, we also consider the full SM background and argue that our results are valid in a more general and realistic case. We also have distributions that are differently populated by samples of scalar Higgs $h = H$ and pseudoscalar Higgs $h = A$, thus some observables in our set can be used to probe the CP nature of the top Yukawa coupling.

An identical analysis is done in the dilepton channel of $t\bar{t}h$ events [44]. We compare the results obtained in both channels, arguing that the single lepton $t\bar{t}h$ final states presents better results and that it could be used in a joint analysis for CP-odd Yukawa coupling component searches.

EVENT GENERATION, SIMULATION AND RECONSTRUCTION

The generation of $t\bar{t}h$ and dominant background $t\bar{t}b\bar{b}$ events, at next-to-leading-order (NLO) in QCD, is done by the MadGraph5_aMC@NLO [25] package with the NNPDF2.3 PDF sets [32]. We use the default model, SM, for signal event generation that follows the Standard Model and to add a CP-odd component in the Yukawa couplings for signal events we use the HC_NLO_X0 model [27]. Signal samples were generated for two limit hypotheses for its coupling to top quarks: the pure CP-even case ($|\cos\alpha| = 1$) and the pure CP-odd case ($|\cos\alpha| = 0$). The other contributions from the SM are considered alongside the dominant background $t\bar{t}b\bar{b}$. These include samples of $t\bar{t}+$ jets (where *jets* stands for up to 3 additional c- or light-flavored quarks), $t\bar{t}V+$ jets (where $V = W^{\pm}, Z$ and *jets* can go up to one additional light quark), single top quark production (s-channel and t-channel), diboson ($W^+W^-, ZZ, W^{\pm}Z$ + jets with up to 3 additional light-quarks), $W^{\pm}+$ jets (with up to 4 additional light quarks) and $Wb\bar{b}+$ jets (with up to 2 additional light- quarks). These samples were also generated with MadGraph5_aMC@NLO but at leading-order (LO) in QCD. The cross-section of the $t\bar{t}+$ jets sample was normalized to the next-to-next-to leading-order (NNLO) in QCD with next-to-next-to leading-logarithm (NNLL) soft gluon resummation computation [28–32], the single top quark samples' cross sections were scaled to the approximate NNLO theoretical computations [33, 34], both considering the NNPDF2.3 PFD sets and rescaled to the top mass used for generation with the prescription given in [35].

LHC-like proton-proton collisions, with a centre-of-mass energy of 13 TeV, are considered with non-fixed renormalization and factorization scales set to the sum of the transverse masses of all final state particles and partons. The masses of the top quark (m_t), the W boson (m_W) and Higgs bosons (for both scalar, m_H, and pseudo-scalar, m_A) were set to 173 GeV, 80.4 GeV and 125 GeV, respectively.

Once the samples are generated, the MadSpin [24] package runs the decays of $t \to bW^+ \to b\ell^+\nu_\ell$ or $bq\bar{q}'$, $\bar{t} \to \bar{b}W^- \to \bar{b}\ell^-\bar{\nu}_\ell$ or $\bar{b}q''\bar{q}'''$ and $h \to b\bar{b}$, with $\ell^{\pm} \in \{e^{\pm}, \mu^{\pm}\}$ and $q^{(n)} \in \{u, d, c, s\}$, and preserves full spin correlation information. Parton showering and hadronization is achieved with the Pythia6 package [36]. For the generator and parton shower matching we use the MLM [40] scheme for LO events and the MC@NLO [37] matching for NLO events. The events were fed to the Delphes package [38] for a fast simulation of the ATLAS detector, using the default ATLAS card. The charged leptons and jets are reconstructed during detector simulation. The efficiencies and resolutions of the detector subsystems are parametrized in segments of p_T (or E) and η. In the $|\eta| < 2.5$ region particle tracking occurs and the efficiency is at least 83% (98%) for electrons (muons), with $p_T = 1$ GeV. The momentum resolution of a track is at most 5%. The efficiency for identifying electrons and muons is 95% in the central region $|\eta| \leq 1.5$, 85% in the intermediate region $1.5 < |\eta| \leq 2.5$ (2.7 for muons), and zero for $|\eta| > 2.5$ (2.7 for muons) or $p_T < 10$ GeV. For electrons with $E = 25$ GeV their energy resolution is 1.5%; for asymptotically high energy this decreases to 0.5%. On the other hand, muon momentum resolution is worse for higher transverse momentum and higher $|\eta|$, reaching a maximum at 10% for $p_T > 100$ GeV and $1.5 < |\eta| \leq 2.5$. The FastJet [45] package is used for jet reconstruction through the anti-k_t algorithm [39] with R

parameter set to 0.4. The efficiency for b-tagging b-jets is dependent on their transverse momentum (p_t) and given by

$$I(p_t) = 0.8 \tanh(0.003 p_t) \frac{30}{1 + 0.086 p_t} \tag{2}$$

in the region $p_T \geq 10$ GeV or $|\eta| \leq 2.5$. The efficiency is zero outside the previous region. For any other jet, the b-tagging misidentification rate is given by

$$M(p_t) = 0.002 + 7.3 * 10^6 p_t \tag{3}$$

A pre-analysis of the generated and simulated events was performed with MadAnalysis 5 [41] in the expert mode [42]. The main pre-analysis' goal was to define transfer functions (TFs) to parameterize the energy response of the jets and missing transverse momentum (MET). These TFs are used by the kinematic fit ahead. The prescription follows closely to [43]. On evaluating the TFs for jets, an energy and pseudo-rapidity dependence was verified. As such, we defined several TFs for each regime of jet energy (in blocks of 100 GeV) and of absolute value of jet pseudo-rapidity (in blocks of 0.5).

The kinematic fit of the events was accomplished with KLFitter [43]. Events are selected if one reconstructed lepton and at least six reconstructed jets are present. The leptons and b-tagged jets need to have their $P_t \geq 20$ GeV and $|\eta| \leq 2.5$, non b-tagged jets must have $P_t \geq 20$ GeV and $|\eta| \leq 4.5$. Furthermore, only events with $P_t^{\text{miss}} > 20$ GeV are accepted.

The kinematic fit of the events is carried out by maximizing the likelihood function

$$\begin{aligned} L(y|\theta) = & B(m_{(\text{bHad,LJ1,LJ2})}|m_{\text{top}}, \Gamma_{\text{top}}) \\ & \times B(m_{(\text{LJ1,LJ2})}|m_W, \Gamma_W) \\ & \times B(m_{(\text{bLep},l,\nu)}|m_{\text{top}}, \Gamma_{\text{top}}) \\ & \times B(m_{(l,\nu)}|m_W, \Gamma_W) \\ & \times B(m_{(\text{bH1,bH2})}|m_H, \Gamma_H) \\ & \times \prod_{i=1}^{6} W_i^{\text{jet}}(E_i^{\text{meas}}|E_i^{\text{parton}}) \\ & \times W_l(E_l^{\text{meas}}|E_l^{\text{parton}}) \\ & \times W_{\text{miss}}(E_{\text{miss,x}}^{\text{meas}}|E_{\nu,x}^{\text{parton}}) \\ & \times W_{\text{miss}}(E_{\text{miss,y}}^{\text{meas}}|E_{\nu,y}^{\text{parton}}) \end{aligned} \tag{4}$$

This likelihood function is divided in two parts. The first part, with several Breit-Wigner probability distribution functions (p.d.f.), $B(m_{x_1,x_2,\ldots}|m_{\text{pole}}, \Gamma)$, indicates how probable it is to obtain an invariant mass constructed from the final particle states $x_1, x_2, \ldots$ given the mass of the top quark ($m_t = 173$ GeV , $\Gamma_t = 1.5$ GeV), W boson ($m_W = 80.4$ GeV , $\Gamma_W = 2.1$ GeV) or Higgs boson ($m_h = 125$ GeV , $\Gamma_h = 3.512$ MeV). The second part, with several transfer functions (TFs) of the final particle state, $W(E^{\text{meas}}|E^{\text{parton}})$, account for radiation lost and smearing/resolution effects on the particles' energies once they pass through the detector. The TFs are obtained from the pre-analysis or *truth-match* analysis.

Permutations of the several final particles is done in order to fill the positions of the expected final parton topology which contains: two c- or light-flavored quarks (LJ1, LJ2) from the hadronically decaying W boson (W_{had}), one lepton and undetected neutrino from the leptonically decaying W boson (W_{lep}), a pair of $b\bar{b}$ quarks (bH1, bH2) from the decay of the Higgs boson and two additional b and $\bar{b}$ quarks (bHad, bLep) from the two top quarks (t_{had}, t_{lep}). The four latter partons inherit the name from their W boson decay mode.

The neutrino reconstruction is accomplished by considering a direct relation between the missing transverse momentum components with those of the neutrino's and solving for p_ν^z,

$$m_W^2 = (\vec{p}_\nu + \vec{p}_l)^2 \tag{5}$$

If there are two solutions, the one that maximizes the likelihood is taken. When there is no solution the event is discarded.

For each permutation the set of most likely parton energies (E_i^{parton}) can be retrieved from the estimator $\hat{\theta}$ defined as

$$\hat{\theta} : \frac{\partial}{\partial \theta} L(y|\theta) = 0 \tag{6}$$

where

$$\begin{aligned} \{\theta\} = \{ & E_{\text{bHad}}^{\text{parton}}, E_{\text{bLep}}^{\text{parton}}, E_{\text{LJ1}}^{\text{parton}}, E_{\text{LJ2}}^{\text{parton}}, E_{\text{bH1}}^{\text{parton}}, \\ & E_{\text{bH2}}^{\text{parton}}, E_l^{\text{parton}}, E_{\nu,x}^{\text{parton}}, E_{\nu,y}^{\text{parton}}, E_{\nu,z}^{\text{parton}} \} \end{aligned} \tag{7}$$

The set $\hat{\theta}^\star$, out of all permutations, that maximizes the value of the likelihood function is taken as the solution for that event. The partons' 4-momenta are reconstructed from the objects of that particular permutation ($\hat{\theta}^\star$) and in order to accommodate the momentum corrections from the transfer functions, their energy is changed to that obtained in $\hat{\theta}$ and the other components are rescaled according to

$$\vec{p}_i^{\,\text{parton}} = \xi_i \vec{p}_i^{\,\text{meas}} \tag{8}$$

with

$$\xi_i = \sqrt{\frac{(E_i^{\text{parton}})^2 - m_i^2}{(E_i^{\text{meas}})^2 - m_i^2}} \tag{9}$$

Only topologies with 6 to 8 jets (3 or 4 of which are b-tagged) were considered due to computational limitations. These topologies had the best reconstruction efficiency, defined as the fraction of matched events for which the chosen permutation was the correct one, of about 30.1 %.

Figure 1 shows the correlations between the reconstructed transverse momenta of the Higgs boson and top quark with the corresponding truth-level values. It is also shown the equivalent results from the dilepton final state studied in [44] for comparison, where there is a clear enhancement by using the KLFitter package.

$t\bar{t}H$, $t\bar{t}A$ AND $t\bar{t}b\bar{b}$ ANGULAR DISTRIBUTIONS

We follow the procedure given in [22] and let θ_Y^X be the angle between the direction of the Y system in the rest frame of X and the direction of the X system, in the rest frame of its parent system.

We refer to the decay chains of the signal by a numbering scheme, i.e, the $t\bar{t}h$ system is labeled 123, then it goes through successive two-body decays, i.e., $(123) \to 1+(23)$, $(23) \to 2+(3)$ and $3 \to 4+5$. In turn, we define the following functions: $f(\theta_1^{123})g(\theta_4^3)$, $f(\theta_1^{123})g(\theta_3^{23})$ and $f(\theta_3^{23})g(\theta_4^3)$, where f and g can either be the sine or cosine. The (123) system momentum direction is measured with respect to the laboratory frame. We established that particles 1 to 3 can be either the top, anti-top quark or the Higgs boson, without repetition. Any of the decay products of the top quarks, Higgs boson or W bosons can fill particle 4 position. We can boost particle 4 to the center-of-mass frame of particle 3 in two ways:

1. Using the laboratory four-momentum of both particles 3 and 4 (*direct* boost).

2. Boosting 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).

Both procedures lead to different results for the particle's 4 direction, due to Wigner rotations[2]. We consider the two procedures in this work.

Angular Distributions after Reconstruction

The signal distributions obtained deteriorate after the cuts from the selection of events and by the kinematic fit. We see that even though distributions distort and are affected by the reduction of number of events they still remain largely preserved. In Figure 2 and 3, 2D plots of $\theta_t^{t\bar{t}h}$, the angle between the direction of flight of the top quark (in the $t\bar{t}h$ system) and the $t\bar{t}h$ flight direction (in the lab frame), vs. $\theta_{b_h}^h$, the angle between the Higgs direction (in the $\bar{t}h$ frame) and the b quark from the Higgs (in the Higgs frame) are shown. There is a fourfold comparison in each figure: the first column are the semileptonic channel results to be compared with the results of the dileptonic channel in the second column. The first row shows the case where $h = H$ and on the second $h = A$. Figure 2 shows the results from the generator level. Figure 3 shows the same plots after event selection and kinematic reconstruction where the density of points shows a similar pattern of that from Figure 2. Even after kinematic reconstruction, clear differences between the different signal natures are visible. Also notable is the clear enhancement of quality in the single lepton channel compared to the dileptonic one.

For the trigonometric functions described in the previous section, six of these showed promising and, following the recipe from [22], we defined their forward-backward asymmetries. The asymmetries are evaluated at generator level and after kinematic fit. These functions (which will be taken as variables) and their forward-backward asymmetries are defined:
$\cos(\theta_h^{t\bar{t}h})\cos(\theta_{\ell^-}^h)$ for $A_{FB}^{\ell-(h)}$(*dir.* boost),
$\sin(\theta_h^{t\bar{t}h})\sin(\theta_{b_{\bar{t}}}^{\bar{t}})$ for $A_{FB}^{\bar{b}_{\bar{t}}(\bar{t})}$(*seq.* boost),
$\sin(\theta_h^{t\bar{t}h})\cos(\theta_{b_h}^{\bar{t}})$ for $A_{FB}^{b_h(\bar{t})}$(*seq.* boost),
$\sin(\theta_t^{t\bar{t}h})\sin(\theta_{W+}^h)$ for $A_{FB}^{W+(h)}$(*seq.* boost),
$\sin(\theta_{\bar{t}}^{t\bar{t}h})\sin(\theta_{b_h}^h)$ for $A_{FB}^{b_h(h)}$(*seq.* boost),
$\sin(\theta_h^{t\bar{t}h})\sin(\theta_{\bar{t}}^{t\bar{t}})$ for $A_{FB}^{\bar{t}(t\bar{t})}$(*dir.* boost) and
$b_4 = (p_t^z.p_{\bar{t}}^z)/(|\vec{p}_t|.|\vec{p}_{\bar{t}}|)$, as defined in [20], for $A_{FB}^{b_4}$.

The last being an additional variable that presented a big difference between the signals.

Some of these distributions are depicted in Figures 4 where we compare the *truth-match* distributions (with cuts) with the kinematic reconstructed ones. Table I shows the asymmetries values for generator events without any selection done and after full kinematic reconstruction.

Asymmetries	Generator (no cuts applied)		After selection and reconstruction	
	$t\bar{t}H/t\bar{t}A$	$t\bar{t}b\bar{b}$	$t\bar{t}H/t\bar{t}A$	$t\bar{t}b\bar{b}$
$A_{FB}^{\ell-(h)}$(*dir.* boost)	+0.33/+0.34	−0.01	+0.10/+0.17	−0.01
$A_{FB}^{\bar{b}_{\bar{t}}(\bar{t})}$(*seq.* boost)	+0.29/+0.37	−0.22	+0.20/+0.19	−0.09
$A_{FB}^{b_h(\bar{t})}$(*seq.* boost)	−0.65/−0.78	−0.22	−0.67/−0.72	−0.65
$A_{FB}^{W+(h)}$(*seq.* boost)	+0.02/−0.45	−0.67	−0.33/−0.51	−0.51
$A_{FB}^{b_h(h)}$(*seq.* boost)	+0.28/−0.08	+0.03	+0.18/+0.02	−0.05
$A_{FB}^{\bar{t}(t\bar{t})}$(*dir.* boost)	+0.16/+0.40	−0.26	+0.17/+0.15	−0.11
$A_{FB}^{b_4}$	+0.33/−0.12	+0.31	+0.17/−0.08	+0.06

TABLE I: Asymmetry values for $t\bar{t}H$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$ at generator level (without any cuts) and after applying the selection criteria and kinematic reconstruction, are shown.

[2] The generators of the Lorentz group do not commute.

OBSERVABLES SENSITIVE TO THE CP NATURE OF THE TOP YUKAWA COUPLING

We defined several angular observables that show important contrasts between distributions of $t\bar{t}b\bar{b}$ events and signal ($t\bar{t}H$ and $t\bar{t}A$) in the previous section. These can be sub-grouped in two sets: a first set of distributions for which $t\bar{t}H$ and $t\bar{t}A$ samples are very similar (see the plot on the middle of Figure 4 as an example), which can be used to search for the $t\bar{t}h$ production process. The second set is composed of distributions that discriminate between $t\bar{t}H$ and $t\bar{t}A$ samples at reconstruction level without truth-match, meaning that they are useful for the investigation into a pseudo-scalar component of the top Yukawa coupling.

Observables in $t\bar{t}h$ events with this same purpose have been previously proposed, for example, in [20, 23, 24]. The observables proposed in these works, for the $t\bar{t}H$ and $t\bar{t}A$ signal samples as well as for the $t\bar{t}b\bar{b}$ background, were studied in reconstructed events. Only the variable b_4 showed promising for the semileptonic case. This variable, as well as the ones presented in Figure 2 for instance, requires the knowledge of the intermediary particles like the top and anti-top quarks, only possible to obtain with a full kinematic reconstruction algorithm, like the one presented in this work.

ANALYSIS AND RESULTS

As mentioned, no further criteria have been applied beyond focusing on final topologies with 6-8 jets (of which 3-4 are b-tagged), due to the computational limitations. Conveniently, it was verified that most backgrounds, in particular $t\bar{t}$+jets, are suppressed by selecting events with at least 3 b-tagged jets.

The expected effective cross-sections (in fb) is showed in Table II, at several levels of the analysis, for single charged lepton signal and SM backgrounds. The $t\bar{t}A$ pseudo-scalar signal was scaled to the $t\bar{t}H$ scalar cross-section for comparison purposes.

In Figure 5, the expected number of events from the different SM processes are shown, including the Higgs signal, for a luminosity of 100 fb^{-1} at the LHC, for two angular distributions.

For completeness, we added up a pseudo-experiment, the fake data, from randomly created events following the expected Standard Model signal and background distributions. The goal is to have a guideline of the expected total number of events and related statistical uncertainties, at every stage of the analysis.

Expected limits at 95% confidence level (CL) for $\sigma \times BR(h \to b\bar{b})$ and for signal strength μ, in the background-only scenario, were extracted, using the b_4 output distribution. All the other variables yield the same results approximately. Two signal samples were used, with values of $|\cos(\alpha)| = \{0, 1\}$. Figure 6, shows the limits and signal strengths, for integrated luminosities of 100, 300 and 3000 fb^{-1}. Since data taking for large values of luminosity is expected to occur with $\sqrt{s}$=14 TeV, we show the results at 3000 fb^{-1} for comparison. Sensitivity to SM $t\bar{t}H$ production at the μ=1 should be attained shortly past the 100 fb^{-1} of luminosity, using this channel alone. Combining the semileptonic channel with other decay channels should allow to decrease significantly the necessary luminosity to probe the structure of the top quark Yukawa couplings to the Higgs boson.

	$N_{jets} \geq 6$ $N_{lep} = 1$	Kinematic Fit (all cuts)
$t\bar{t}$+$c\bar{c}$, $t\bar{t}$+lf	2488	565.5
$t\bar{t}$+$b\bar{b}$	898.4	165.6
$t\bar{t}$+$V(V$=$Z,W)$	74.9	4.1
Single t	492.2	4.9
W+jets	3293	0
W+$b\bar{b}$	709.7	3.7
Diboson	996.6	0.5
Total back.	8953	744.3
$t\bar{t}H$	26.6	8.85
$t\bar{t}A$	18.9	6.07

TABLE II: Expected cross-sections (in fb) at two selection stages, at 13 TeV, for signal and background events at the LHC.

CONCLUSIONS

In this paper, studies of $t\bar{t}h$ production, for scalar and pseudo-scalar Higgs bosons, at a centre-of-mass energy of 13 TeV at the LHC, are considered for different luminosities. Semileptonic final states from $t\bar{t}h$ decays $t \to bW^+ \to b(l^+\nu_l$ or $jj)$, $\bar{t} \to \bar{b}W^- \to \bar{b}(l^-\bar{\nu_l}$ or $jj)$, $h \to b\bar{b}$ are fully reconstructed by means of a kinematic fit that reconstructs the four momenta of all intermediary particles and the undetected neutrino. New angular distributions and asymmetries are proposed to allow better discrimination between signals of different nature (scalar or pseudo-scalar) and backgrounds at the LHC. It is possible to obtain relevant information through the measurements of new angular distributions and asymmetries using fully reconstructed $t\bar{t}h$ events. The reconstruction performance of these events is shown to be better with KLFitter than with the method used in the dileptonic analysis presented here for comparison. On the other hand, the KLFitter package carries a huge computational weight, something to consider when the amount of events to reconstruct is numerous. Nonethe-

less, both channels show that, even after event selection and full kinematic reconstruction, the spin information is largely preserved, opening a window for spin measurements and a better understanding of the nature of the top-Higgs Yukawa coupling and $t\bar{t}h$ production at the LHC. Expected limits at 95% CL were extracted on the $\sigma \times BR(h \to b\bar{b})$ and signal strength μ using these new angular variables. It should be stressed that some of the angular distributions investigated in this work were used in addition to other observables commonly discussed in the literature, yielding at least the same sensitivity to the nature of the top quark Yukawa coupling to Higgs boson, if not better. All results presented so far were obtained using the semileptonic final states of $t\bar{t}h$ events alone, which were found to fare better than in the dileptonic channel. Thus searches for a CP-odd component in the coupling of the Higgs boson to top quarks are expected to be improved when combined with this decay channel.

Acknowledgements

ADD PLOT Acknowledgements. Check contract numbers,etc.

This work was partially supported by Fundação para a Ciência e Tecnologia, FCT (projects CERN/FIS-NUC/0005/2015 and CERN/FP/123619/2011, grant SFRH/BPD/100379/2014 and contract IF/01589/2012/CP0180/CT0002). The work of R.S. is supported in part by HARMONIA National Science Center - Poland project UMO-2015/18/M/ST2/00518. The work of R.F. is supported by the Alexander von Humboldt Foundation in the framework of the Sofja Kovalevskaja Award Project "Event Simulation for the Large Hadron Collider at High Precision". Special thanks goes to our long term collaborator Filipe Veloso for the invaluable help and availability on the evaluation of the confidence limits discussed in this paper.

[1] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **716** (2012) 1, arXiv:1207.7214 [hep-ex];

[2] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **716** (2012) 30, arXiv:1207.7235 [hep-ex];

[3] P. W. Higgs, Phys. Lett. **12**, 132 (1964), Phys. Rev. Lett. **13**, 508 (1964) and Phys. Rev. **145**, 1156 (1964); F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964); G.S. Guralnik, C.R. Hagen and T.W. Kibble, Phys. Rev. Lett. **13**, 585 (1964).

[4] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **726** (2013) 88 [Erratum-ibid. B **734** (2014) 406], arXiv:1307.1427 [hep-ex]; G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **726**, 120 (2013), arXiv:1307.1432 [hep-ex]; G. Aad *et al.* [ATLAS Collaboration], arXiv:1501.04943 [hep-ex]; V. Khachatryan *et al.* [CMS Collaboration], arXiv:1412.8662 [hep-ex]; S. Chatrchyan *et al.* [CMS Collaboration], Nature Phys. **10**, 557 (2014) arXiv:1401.6527 [hep-ex];

[5] T. D. Lee, Phys. Lett. D **8 1226-1239** (1973) 1, doi:10.1103/PhysRevD.8.1226;

[6] S. Weinberg, Phys. Rev. Lett. **37 657** (1976) 1, doi:10.1103/PhysRevLett.37.657.;

[7] D. Fontes, J. C. Romão, R. Santos and J. P. Silva, JHEP **1506**, 060 (2015) doi:10.1007/JHEP06(2015)060 [arXiv:1502.01720 [hep-ph]].

[8] V. Khachatryan *et al.* [CMS Collaboration], arXiv:1411.3441 [hep-ex].

[9] V. Khachatryan *et al.* [CMS Collaboration], Phys. Lett. B **759**, 672 (2016) doi:10.1016/j.physletb.2016.06.004 [arXiv:1602.04305 [hep-ex]].

[10] G. Aad *et al.* [ATLAS Collaboration], Eur. Phys. J. C **75**, no. 10, 476 (2015) Erratum: [Eur. Phys. J. C **76**, no. 3, 152 (2016)] doi:10.1140/epjc/s10052-015-3685-1, 10.1140/epjc/s10052-016-3934-y [arXiv:1506.05669 [hep-ex]].

[11] J. N. Ng and P. Zakarauskas, Phys. Rev. D **29**, 876 (1984); Z. Kunszt, Nucl. Phys. B **247**, 339 (1984); W. J. Marciano and F. E. Paige, Phys. Rev. Lett. **66**, 2433 (1991); J. F. Gunion, Phys. Lett. B **261**, 510 (1991); J. Goldstein *et al.*, Phys. Rev. Lett. **86**, 1694 (2001), [hep-ph/0006311]; W. Beenakker *et al.*, Phys. Rev. Lett. **87**, 201805 (2001), [hep-ph/0107081] and Nucl. Phys. B **653**, 151 (2003), [hep-ph/0211352]; L. Reina and S. Dawson, Phys. Rev. Lett. **87**, 201804 (2001), [hep-ph/0107101]; S. Dawson *et al.*, Phys. Rev. D **67**, 071503 (2003), [hep-ph/0211438]; S. Dawson *et al.*, Phys. Rev. D **68**, 034022 (2003) [hep-ph/0305087]; S. Dittmaier, M. Kramer, and M. Spira, Phys. Rev. D **70**, 074010 (2004) [hep-ph/0309204]; R. Frederix *et al.*, Phys. Lett. B **701**, 427 (2011) [arXiv:1104.5613 [hep-ph]]; M. V. Garzelli *et al.*, Europhys. Lett. **96** (2011) 11001 [arXiv:1108.0387 [hep-ph]]; H. B. Hartanto *et al.*, arXiv:1501.04498 [hep-ph]; S. Frixione, V. Hirschi, D. Pagani, H. S. Shao and M. Zaro, JHEP **1409** (2014) 065 [arXiv:1407.0823 [hep-ph]]; Y. Zhang, W. G. Ma, R. Y. Zhang, C. Chen and L. Guo, Phys. Lett. B **738** (2014) 1 [arXiv:1407.1110 [hep-ph]].

[12] G. Aad *et al.* [ATLAS Collaboration], JHEP **1605**, 160 (2016) doi:10.1007/JHEP05(2016)160 [arXiv:1604.03812 [hep-ex]].

[13] G. Aad *et al.* [ATLAS Collaboration] Phys. Lett. B **740** (2015) 222, arXiv:1409.3122 [hep-ex].

[14] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **749**, 519 (2015) doi:10.1016/j.physletb.2015.07.079 [arXiv:1506.05988 [hep-ex]].

[15] G. Aad *et al.* [ATLAS Collaboration], arXiv:1503.05066 [hep-ex].

[16] V. Khachatryan *et al.* [CMS Collaboration], JHEP **1409**, 087 (2014) Erratum: [JHEP **1410**, 106 (2014)] doi:10.1007/JHEP09(2014)087, 10.1007/JHEP10(2014)106 [arXiv:1408.1682 [hep-ex]].

[17] V. Khachatryan *et al.* [CMS Collaboration], arXiv:1502.02485 [hep-ex].

[18] G. Aad *et al.* [ATLAS and CMS Collaborations], JHEP **1608**, 045 (2016) doi:10.1007/JHEP08(2016)045 [arXiv:1606.02266 [hep-ex]].

[19] C. Patrignani *et al.* [Particle Data Group], Chin. Phys. **C40** (2016), doi:10.1088/1674-1137/40/10/100001.

[20] J. F. Gunion and X. G. He, Phys. Rev. Lett. **76** (1996) 4468 doi:10.1103/PhysRevLett.76.4468 [hep-

ph/9602226].
[21] F. Boudjema, R. M. Godbole, D. Guadagnoli and K. A. Mohan, arXiv:1501.03157 [hep-ph];
[22] S. P. Amor dos Santos, J. P. Araque, R. Cantrill, N. F. Castro, M. C. N. Fiolhais, R. Frederix, R. Gonçalo and R. Martins *et al.*, Phys. Rev. D 92, 034021 (2015), arXiv:1503.07787 [hep-ph].
[23] J. Ellis *et al.*, JHEP **1404**, 004 (2014) [arXiv:1312.5736 [hep-ph]]; S. Biswas *et al.*, JHEP **1407**, 020 (2014) [arXiv:1403.1790 [hep-ph]]; F. Demartin *et al.*, Eur. Phys. J. C **74**, no. 9, 3065 (2014) [arXiv:1407.5089 [hep-ph]];
[24] P. Artoisenet *et al.*, JHEP **1303**, 015 (2013), arXiv:1212.3460 [hep-ph].
[25] Alwall, J. and Frederix, R. and Frixione, S. and Hirschi, V. and Maltoni, F. and Mattelaer, O. and Shao, H. -S. and Stelzer, T. and Torrielli, P. and Zaro, M., JHEP **07** (2014), doi:10.1007/JHEP07(2014)079.
[26] R. D. Ball *et al.*, Nucl. Phys. **B867** (2013), doi:10.1016/j.nuclphysb.2012.10.003.
[27] P. Artoisenet *et al.*, JHEP **11** (2013), doi:10.1007/JHEP11(2013)043, eprint:1306.6464.
[28] M. Czakon and A. Mitov, Comput. Phys. Commun. **185** (2014) 2930 doi:10.1016/j.cpc.2014.06.021 [arXiv:1112.5675 [hep-ph]].
[29] M. Botje *et al.*, arXiv:1101.0538 [hep-ph].
[30] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, Eur. Phys. J. C **64**, 653 (2009) [arXiv:0905.3531 [hep-ph]].
[31] J. Gao *et al.*, Phys. Rev. D **89**, no. 3, 033009 (2014) [arXiv:1302.6246 [hep-ph]].
[32] R. D. Ball *et al.*, Nucl. Phys. B **867**, 244 (2013) [arXiv:1207.1303 [hep-ph]].
[33] N. Kidonakis, Phys. Rev. D **81**, 054028 (2010) [arXiv:1001.5034 [hep-ph]].
[34] N. Kidonakis, Phys. Rev. D **83**, 091503 (2011) [arXiv:1103.2792 [hep-ph]].
[35] M. Czakon, P. Fiedler and A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013) arXiv:1303.6254 [hep-ph].
[36] T. Sjöstrand, S. Mrenna and P. Z. Skands, JHEP **0605**, 026 (2006), hep-ph/0603175.
[37] S. Frixione and B. R. Webber, JHEP **0206** (2002) 029 doi:10.1088/1126-6708/2002/06/029 [hep-ph/0204244]
[38] J. de Favereau *et al.* [DELPHES 3 Collaboration], JHEP **1402**, 057 (2014), arXiv:1307.6346 [hep-ex].
[39] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804** (2008) 063 doi:10.1088/1126-6708/2008/04/063 [arXiv:0802.1189 [hep-ph]].
[40] J. Alwall *et al.*, Eur. Phys. J. C **53** (2008) 473 doi:10.1140/epjc/s10052-007-0490-5 [arXiv:0706.2569 [hep-ph]].
[41] E. Conte, B. Fuks and G. Serret, Comput. Phys. Commun. **184**, 222 (2013), arXiv:1206.1599 [hep-ph].
[42] E. Conte *et al.*, Eur. Phys. J. C **74**, no. 10, 3103 (2014), arXiv:1405.3982 [hep-ph].
[43] P. Erdmann *et al.*, Nucl. Instrum. Meth. **A748** (2014), doi:10.1016/j.nima.2014.02.029, arXiv:1312.5595 [hep-ex].
[44] S. P. Amor dos Santos *et al.*, Phys. Rev. **D96** (2017), doi:10.1103/PhysRevD.96.013004, arXiv:1704.03565 [hep-ph].
[45] Cacciari, Matteo and Salam, Gavin P. and Soyez, Gregory, Eur. Phys. J. **C72** (2012), doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.609 [hep-ph].

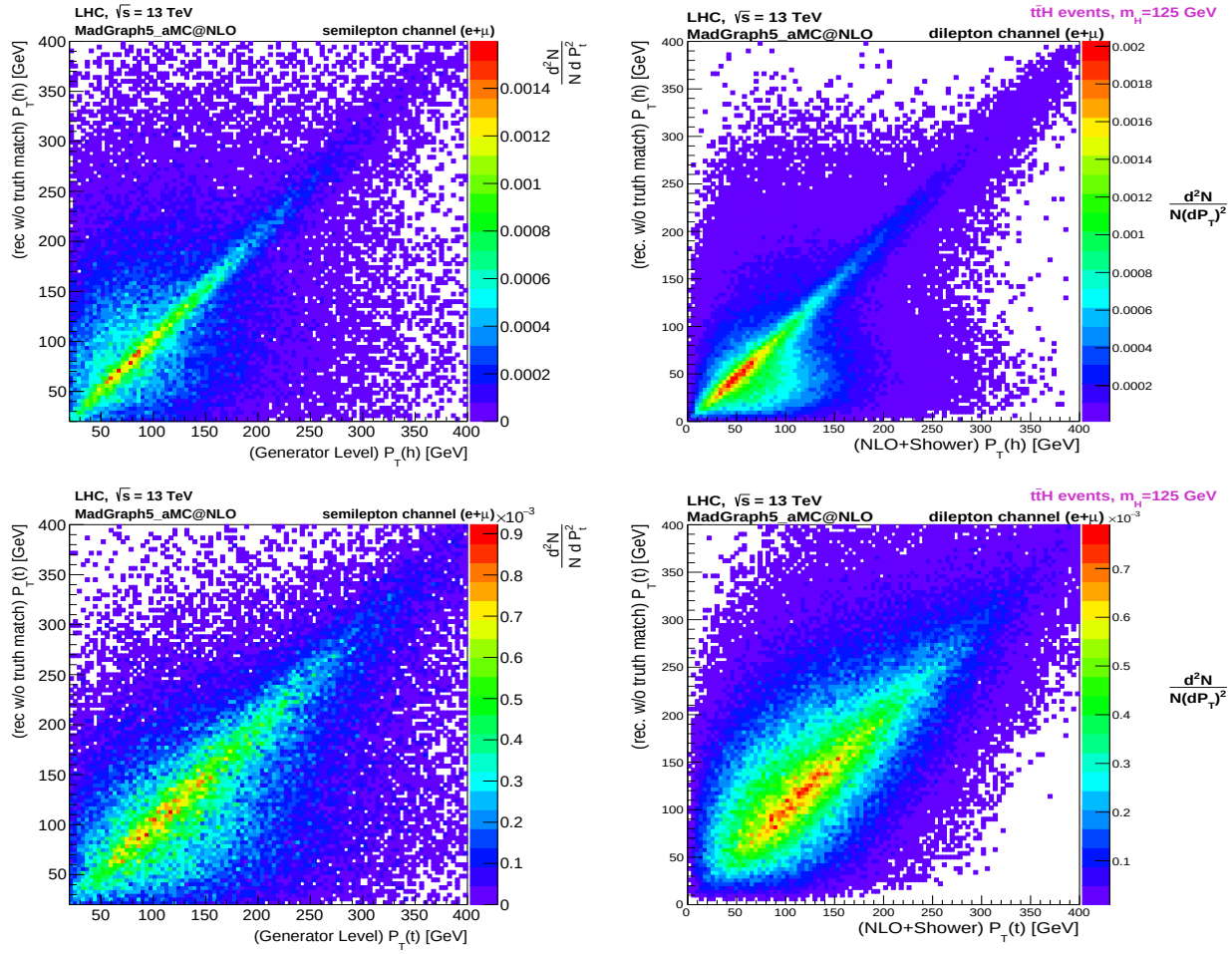

FIG. 1: Two-dimensional distributions of p_T in $t\bar{t}H$ events. The horizontal axes represent variables recorded at generator level, and the vertical axes are the corresponding variables recorded at reconstruction level without truth-match. Top: distribution for the semileptonic (left) and dileptonic (right) Higgs distributions. Bottom: distribution for the semileptonic (left) and dileptonic (right) top quark distributions

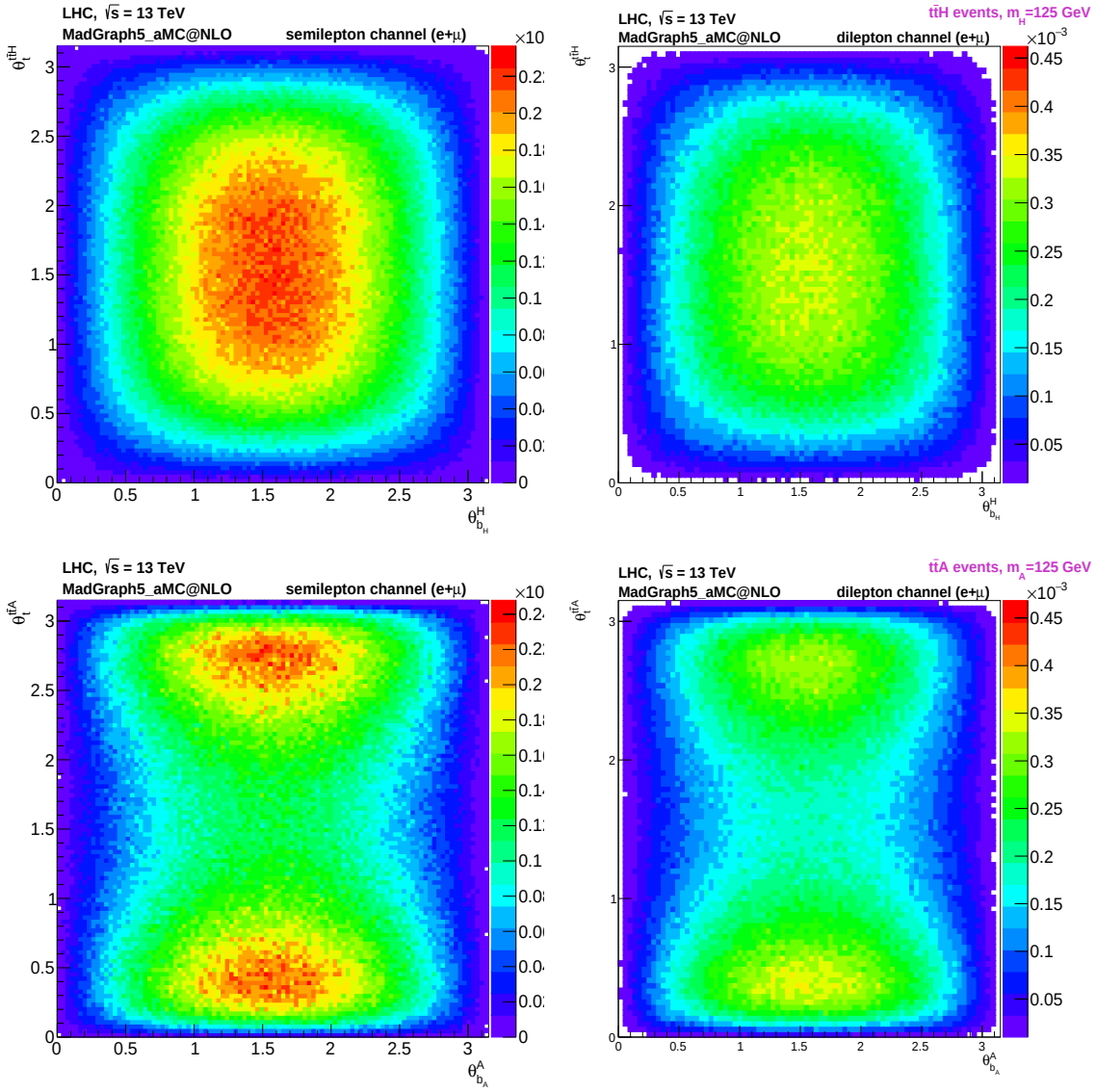

FIG. 2: Two dimensional distribution at generator level of the angle between the top quark, in the $t\bar{t}h$ centre-of-mass frame, and the $t\bar{t}h$ direction in the lab frame (y-axis) plotted against the angle between the Higgs direction, in the $\bar{t}h$ rest frame and the b quark from the Higgs (boosted to the Higgs center-of-mass, x-axis). Top: distributions for pure scalar Higgs ($h = H$) for the semileptonic (left) and dileptonic (right) channels. Bottom: distributions for pure pseudoscalar Higgs ($h = A$) for the semileptonic (left) and dileptonic (right) channels.

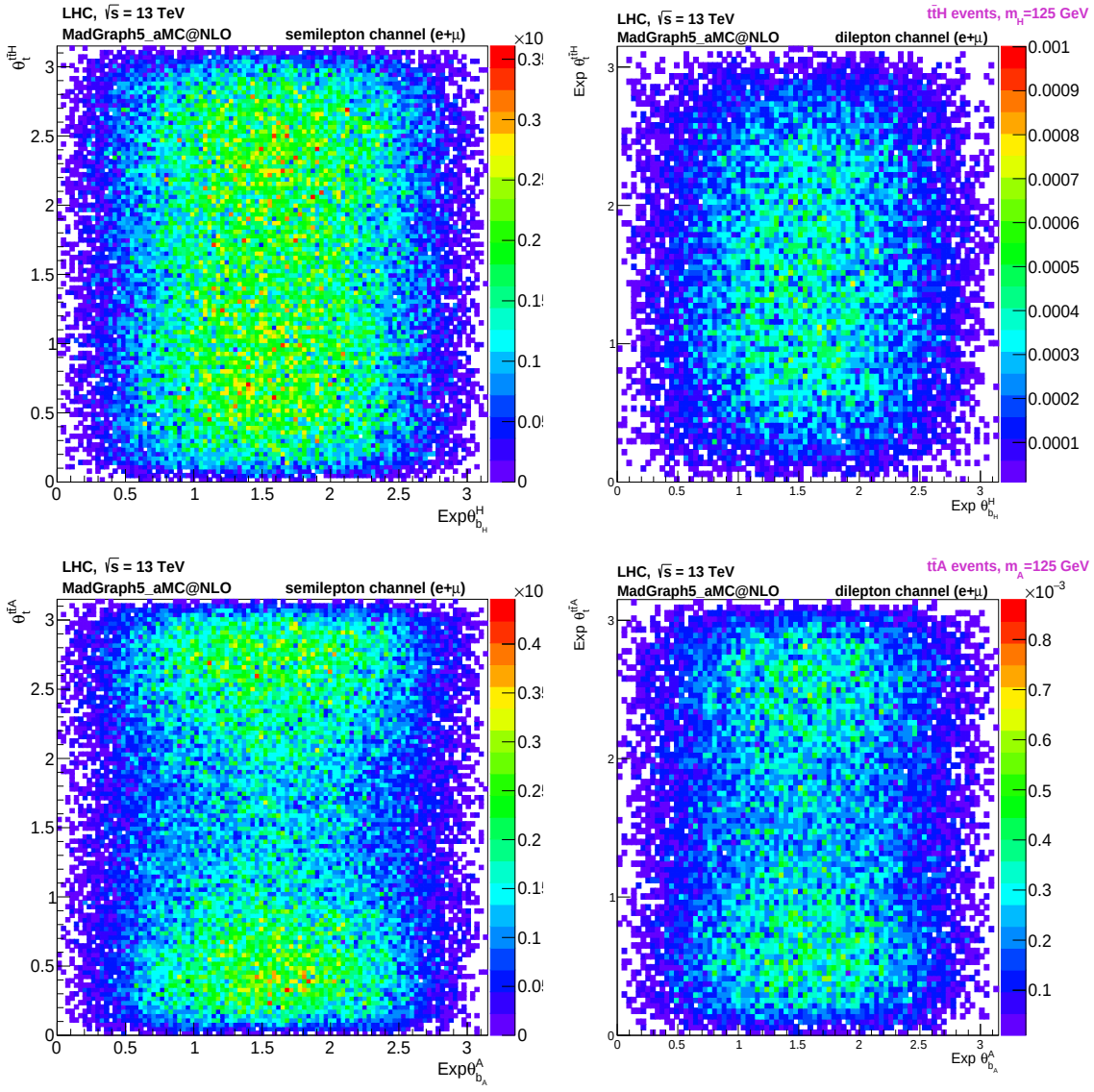

FIG. 3: The same as Figure 2, after all selection cuts and full kinematic reconstruction.

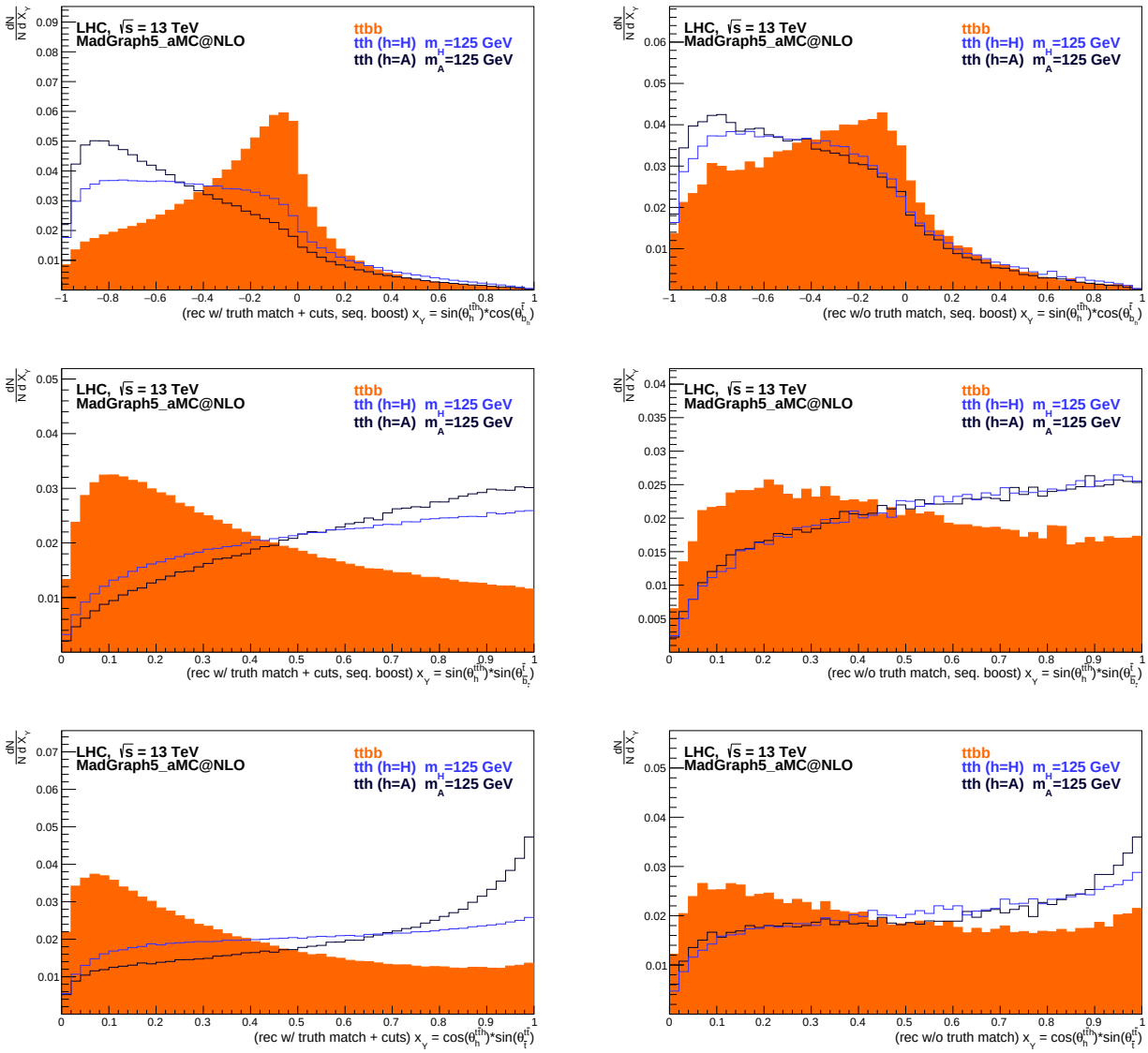

FIG. 4: Distributions of x_Y=sin $(\theta_h^{t\bar{t}h})$ sin $(\theta_{b_h}^{\bar{t}})$ (top), x_Y=sin $(\theta_h^{t\bar{t}h})$ cos $(\theta_{\bar{b}_{\bar{t}}}^{\bar{t}})$ (middle) and x_Y=cos $(\theta_h^{t\bar{t}h})$ sin $(\theta_{\bar{t}}^{t\bar{t}})$ (bottom). The distributions at truth-match level after cuts (left) and after full kinematic reconstruction (right), are shown. The light blue line represents the $t\bar{t}h$ SM model signal ($h = H$ and $CP = +1$) and the dark blue line corresponds to the pure pseudo-scalar distribution $t\bar{t}h$ ($h = A$ and $CP = -1$). The filled region corresponds to the $t\bar{t}b\bar{b}$ dominant background.

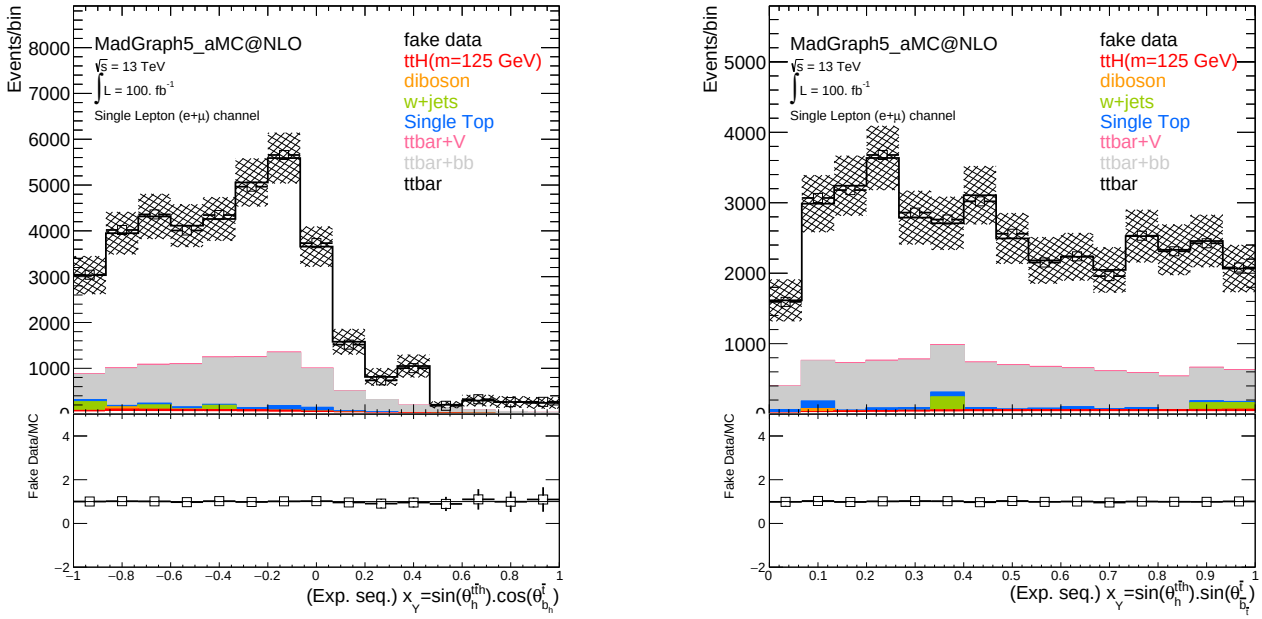

FIG. 5: Distributions of $x_Y{=}\sin(\theta_h^{t\bar{t}h})\sin(\theta_{b_h}^{\bar{t}})$ (left) and $x_Y{=}\sin(\theta_h^{t\bar{t}h})\cos(\theta_{b_{\bar{t}}}^{\bar{t}})$ (right) after final selection at 13 TeV for 100 fb^{-1} with the contributions from the full SM background and fake data.

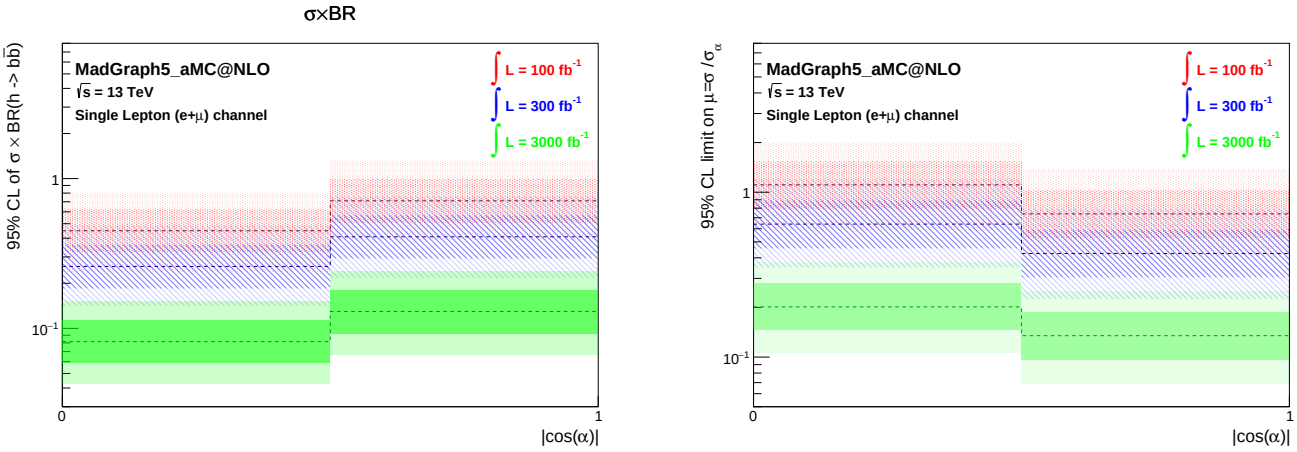

FIG. 6: Expected limits at 95% CL in the background-only scenario, for $|\cos(\alpha)| = 0, 1$. Limits on $\sigma \times BR(h \to b\bar{b})$ (left) and μ (right) obtained with the b_4 observable for integrated luminosities of 100, 300 and 3000 fb^{-1}. The lines correspond to the median, while the narrower (wider) bands correspond to the $1\sigma(2\sigma)$ intervals.