

# Boosting probes of CP violation in the top Yukawa coupling with Deep Learning

Waleed Esmail<sup>a</sup>, A. Hammad<sup>b</sup>, Adil Jueid<sup>c</sup> and Stefano Moretti<sup>d,e</sup>

<sup>a</sup> Institut für Kernphysik, Universität Münster, Wilhelm-Klemm-Str. 9, 48149 Münster, Germany.

<sup>b</sup> Theory Center, IPNS, KEK, 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan.

<sup>c</sup> Particle Theory and Cosmology Group, Center for Theoretical Physics of the Universe,  
Institute for Basic Science (IBS), Daejeon, 34126, Republic of Korea.

<sup>d</sup> School of Physics & Astronomy, University of Southampton,  
Highfield, Southampton SO17 1BJ, U.K.

<sup>e</sup> Department of Physics & Astronomy, Uppsala University,  
Box 516, 75120 Uppsala, Sweden.

## Abstract

The precise measurement of the top-Higgs coupling is crucial in particle physics, offering insights into potential new physics Beyond the Standard Model (BSM) carrying CP Violation (CPV) effects. In this paper, we explore the CP properties of a Higgs boson coupling with a top quark pair, focusing on events where the Higgs state decays into a pair of  $b$ -quarks and the top-antitop system decays leptonically. The novelty of our analysis resides in the exploitation of two conditional Deep Learning (DL) networks: a Multi-Layer Perceptron (MLP) and a Graph Convolution Network (GCN). These models are trained for selected CPV phase values and then used to interpolate all possible values ranging from  $-\frac{\pi}{2}$  to  $\frac{\pi}{2}$ . This enables a comprehensive assessment of sensitivity across all CP phase values, thereby streamlining the process as the models are trained only once. Notably, the conditional GCN exhibits superior performance over the conditional MLP, owing to the nature of graph-based Neural Network (NN) structures. Specifically, for Higgs top coupling modifier set to 1, with  $\sqrt{s} = 13.6$  TeV and integrated luminosity of  $3 \text{ ab}^{-1}$  GCN excludes the CP phase larger than  $|5^\circ|$  at 95.4% Confidence Level (C.L). Our Machine Learning (ML) informed findings indicate that assessment of the CP properties of the Higgs coupling to the  $t\bar{t}$  pair can be within reach of the HL-LHC, quantitatively surpassing the sensitivity of more traditional approaches.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Setup</b>	<b>4</b>
2.1	The $pp \rightarrow t\bar{t}\phi_0$ process	4
2.2	Mass reconstructions of top (anti)quarks	5
<b>3</b>	<b>CP-sensitive observables</b>	<b>6</b>
<b>4</b>	<b>Conditional DNNs</b>	<b>10</b>
4.1	MLP	10
4.2	GNN	11
<b>5</b>	<b>DL analysis</b>	<b>13</b>
5.1	Signal and background kinematics	13
5.2	MC event generation and simulation tools	14
5.3	Training the network	15
<b>6</b>	<b>Results</b>	<b>15</b>
<b>7</b>	<b>Conclusions</b>	<b>19</b>
<b>A</b>	<b>Verify the network performance with a toy example</b>	<b>20</b>

---

## 1 Introduction

Since when first discovered in the long-lived  $K$ -meson rare decay channel  $K_L \rightarrow 2\pi$  back in 1964 [1], CPV has attracted conspicuous theoretical interest, given that studying its dynamics in laboratory experiments may eventually open up a window of understanding on the matter-antimatter asymmetry in the Universe. Unfortunately, all phenomena of the first kind (including those measured also in the  $D$ - and  $B$ -meson sectors) can be explained using the Kobayashi-Maskawa mechanism [2], which, while representing a success of the SM, is, however, not enough to explain the latter [3–5].

Over the past few decades, many Beyond the BSM scenarios that can accommodate additional CPV sources, whether spontaneous or explicit, have been proposed so as to remedy such a SM flaw. While theoretically viable, these are all strongly constrained by experiments. In particular, very precise measurements of the Electric Dipole Moments (EDMs) of, e.g., electron and neutron, [6–8] have already placed severe limits on many new CPV sources [9–12]. In fact, the sensitivities attained herein are far above the SM predictions [13–15], yet, EDM measurements, both those above and others, being very inclusive in their nature, are unlikely to determine the actual interactions affected by CPV. Conversely, collider experiments, despite having weaker sensitivities to CPV effects in comparison, can afford one, thanks to the huge variety of exclusive quantities that one can define in such settings, with an insight into the actual CPV dynamics.

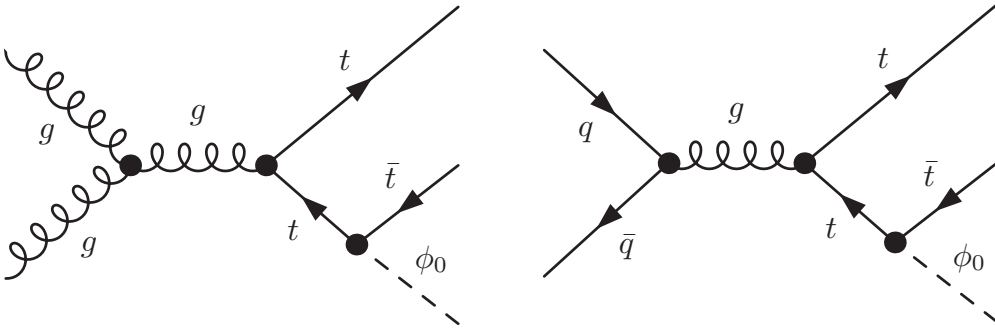


Figure 1: Representative Feynman diagrams corresponding to  $t\bar{t}H_1$  production at Leading Order. Here we show the production through  $gg$  fusion (left) and  $q\bar{q}$  annihilation (right).

This is particularly true for BSM frameworks with extended Higgs sectors [16–20], wherein non-zero EDMs are expected to be the first signal of CPV with collider effects instead being able to provide additional information on it, see, e.g., Refs. [21–38]. Indeed, since the discovery of the 125 GeV Higgs boson at the LHC in 2012 [39–41], testing its CP properties has been high on the ATLAS and CMS agendas, as the SM has a definite prediction in this respect, *i.e.*,  $0^+$ , so that any deviations from this would be a signal of BSM physics. At present, the status of such measurements is that they are all consistent with the CP-even (or scalar) state of the SM, yet, the possibility of a CP-odd (or pseudoscalar) component to it (*e.g.*, through mixing with another Higgs state) cannot be definitely excluded.

A simple BSM setup, exploiting the fact that Nature appears to privilege doublet representations of a Higgs field, is the 2-Higgs Doublet Model (2HDM) [19]. Herein, an effective method to test CPV effects is to study the Yukawa interactions between any Higgs boson (the SM-like one or others) and the top (anti)quark via the Lagrangian term

$$-\mathcal{L}_{t\bar{t}\phi_0} \propto \bar{t} \left( g_S + i g_P \gamma^5 \right) t \phi_0, \quad (1)$$

where  $\phi_0$  refers to a generic Higgs state with mixed CP quantum numbers and  $g_{S(P)}$  refers to its corresponding scalar(pseudoscalar) coupling to a  $t\bar{t}$  pair. This vertex is of importance because a top (anti)quark decays quickly enough so that the information emerging from such an interaction feeds into its final state distributions. Specifically, the CP quantum numbers (and also spin) of the Higgs state can be accessed through these, albeit on a statistical basis.

Phenomenologically, there are a lot of studies in the literature trying to test CPV in  $t\bar{t}\phi_0$  interactions at colliders [32–34, 36, 42–61]. It is the purpose of this paper the one of contributing to the endeavour of extracting CPV effects from  $pp \rightarrow t\bar{t}\phi_0$  at the HL-LHC [62] through a novel approach exploiting two alternative DL methods: a conditional MLP and a conditional GCN.

The paper is organised as follows. We describe our theoretical setup in Sec. 2, including its phenomenological manifestations through the  $pp \rightarrow t\bar{t}\phi_0$  process, while in Sec. 3 we introduce the kinematical observables that we will exploit in our numerical analysis. Sec. 4 is devoted to describe the aforementioned conditional Deep NNs (DNNs) in some detail. We then perform the DL analyses in Sec. 5 and produce our final results in Sec. 6. We then conclude in Sec. 7. (There is also an appendix where we test the DNN activities in terms of a toy example).

## 2 Setup

### 2.1 The $pp \rightarrow t\bar{t}\phi_0$ process

We consider the production of a scalar boson ( $\phi_0$ ) with a mass of  $m_{\phi_0} = 125$  GeV at the LHC in association with a top quark pair. We parameterize the  $t\bar{t}\phi_0$  Lagrangian as follows:

$$-\mathcal{L}_{t\bar{t}\phi_0} = \frac{m_t}{v} \kappa_{tt} \bar{t} \left( \cos \theta_{tt} + i\gamma_5 \sin \theta_{tt} \right) t \phi_0, \quad (2)$$

where  $\kappa_{tt}$  and  $\theta_{tt}$ , assumed to be real valued, are the free parameters of this simplified model of Yukawa interaction. (A pure SM Higgs boson is realised for  $\kappa_{tt} = 1$  and  $\theta_{tt} = 0^\circ$ .) We assume instead that the coupling modifiers of the Higgs boson to the other fermions and gauge bosons are SM-like while we allow for the same degree of  $CPV$  to be transferred from  $\phi_0$  to its decay products. Therefore, the Lagrangian for the interaction of  $\phi_0$  with the other fermions ( $f = d, u, s, c, b$  and  $\ell$ , with  $\ell = e, \mu, \tau$ ) and gauge bosons ( $V = W^\pm, Z$ ) of the SM is parameterized as

$$\mathcal{L} = - \sum_f \frac{m_f}{v} \bar{f} \left( \cos \theta_{ff} + i\gamma_5 \sin \theta_{ff} \right) f \phi_0 + \sum_V \frac{2m_V^2}{v} \cos \theta_{VV} V^\mu V_\mu \phi_0, \quad (3)$$

where we assume that  $\theta_{ff} = \theta_{VV} = \theta_{tt}$  in our analysis. The master formula for the production of  $t\bar{t}\phi_0$  final states at a  $pp$  collider is given by

$$\sigma(pp \rightarrow t\bar{t}\phi_0) \equiv \sum_{i,j} \int dx_i dx_j f_{i/p}(x_i, \mu_F^2) f_{j/p}(x_j, \mu_F^2) \hat{\sigma}_{ij \rightarrow t\bar{t}\phi_0}(\mu_R^2), \quad (4)$$

where  $f_{i/p}(x_i, \mu_F^2)$  is the probability for parton  $i$  to carry a momentum fraction  $x_i = p_i/P$  of the proton momentum  $P$  at a scale  $\mu_F^2$  (factorisation scale) and  $\hat{\sigma}_{ij \rightarrow t\bar{t}\phi_0}$  is the partonic cross section evaluated at a scale  $\mu_R^2$  (renormalisation scale). Representative Leading Order (LO) Feynman diagrams for such a process are shown in Fig. 1. The scattering amplitude receives two contributions at LO:  $gg$  fusion (left panel of Fig. 1) and  $q\bar{q}$  annihilation (right panel of Fig. 1). The gluon fusion contribution is important for small momentum fractions while the (anti)quark contribution is dominant at moderate and large momentum fractions. However, the total cross section is dominated by the contribution of gluons when integrated over all the momentum fractions (about 58% for  $\theta_{tt} = 0^\circ$ ).

We employ MADGRAPH5\_AMC@NLO version 3.4.1 [63, 64] to calculate the production cross section and using the LO PDF set NNPDF40\_lo\_as\_01180 [65] through the LHAPDF library [66]. As for the renormalization and factorization scales, we have adopted a common (dynamical) choice for the central theory predictions, i.e.,

$$\mu_F^0 = \mu_R^0 = \frac{1}{2} \sum_i \sqrt{p_{T,i}^2 + m_i^2}, \quad (5)$$

which is the sum of the transverse masses of all the final-state particles divided by two. We note that the theory uncertainties at LO are dominated by those arising from scale variation ( $\simeq 20\%$ – $30\%$ ) while PDF uncertainties are negligible ( $\simeq 2\%$ – $4\%$ ). The  $K$  factor measuring the size of QCD Next-to-LO (NLO) corrections can be of order  $\approx 1.1$ – $1.2$  (see Refs. [67, 68] for comprehensive analyses of QCD corrections), but we do not take it into account in our simulations for consistency reasons, as not all our backgrounds are known to NLO accuracy. In Fig. 2 we display the cross section defined as

$$\sigma \equiv \sigma(pp \rightarrow t\bar{t}\phi_0) \times \text{BR}(t \rightarrow b\ell\nu)^2 \times \text{BR}(\phi_0 \rightarrow b\bar{b}), \quad \ell = e, \mu, \quad (6)$$

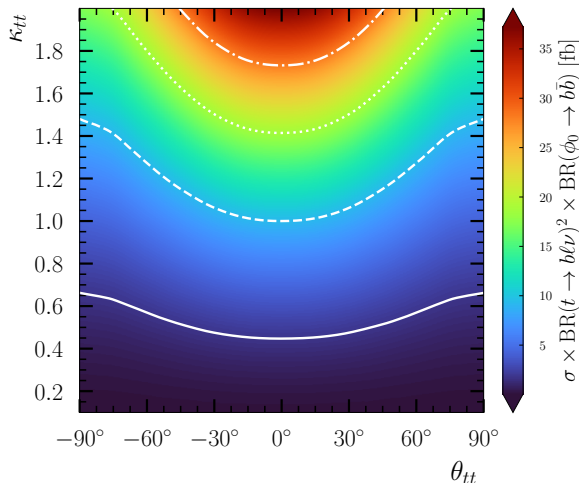


Figure 2: Production cross section of  $t\bar{t}\phi_0$  times the product of the Branching Ratios (BRs) assuming the top quark decaying leptonically and the scalar boson decaying into  $b\bar{b}$  projected on the plane of  $\kappa_{tt}$  and  $\theta_{tt}$ . We also show the contours corresponding to  $\sigma/\sigma_{\text{SM}} = 0.2, 1, 2, 3$  in solid, dashed, dotted and dash-dotted white lines, respectively. Here,  $\sigma_{\text{SM}}$  is the cross section for  $\kappa_{tt} = 1$  and  $\theta_{tt} = 0^\circ$ .

over the plane of  $\theta_{tt}$  and  $\kappa_{tt}$ . The partial width of  $\phi_0 \rightarrow b\bar{b}$  is calculated at Next-to-next NLO (NNNLO) using a full resummed running  $b$ -quark mass [69–72]. We can see that the cross section varies between 2 and 36 fb with the maximum being for large  $\kappa_{tt}$  and  $\theta_{tt} \approx 0$ . Notice that, given that we wish to study (charge and spin) correlations in the  $t\bar{t}$  system, we are considering here fully leptonic decays of it.

## 2.2 Mass reconstructions of top (anti)quarks

The dileptonic decays of the  $t\bar{t}$  system lead to two neutrinos in the final state which implies an ambiguity in the reconstruction of the (anti)top invariant mass. The reconstruction of the full  $t\bar{t}$  invariant mass is also very important to construct the  $CP$ -sensitive observables described in Sec. 3. At hadron colliders, the only handle to neutrinos in this process is through the total missing momentum given that the longitudinal momentum of the initial partons is unknown. The conservation of total momentum of the (anti)top quark and Higgs boson leads to the following constraints:

$$\begin{aligned}
M_{W^+}^2 &= (E_{\ell^+} + E_\nu)^2 - (p_{\ell^+}^x + p_\nu^x)^2 - (p_{\ell^+}^y + p_\nu^y)^2 - (p_{\ell^+}^z + p_\nu^z)^2, \\
M_{W^-}^2 &= (E_{\ell^-} + E_{\bar{\nu}})^2 - (p_{\ell^-}^x + p_{\bar{\nu}}^x)^2 - (p_{\ell^-}^y + p_{\bar{\nu}}^y)^2 - (p_{\ell^-}^z + p_{\bar{\nu}}^z)^2, \\
m_t^2 &= (E_b + E_{\ell^+} + E_\nu)^2 - (p_b^x + p_{\ell^+}^x + p_\nu^x)^2 - (p_b^y + p_{\ell^+}^y + p_\nu^y)^2 - (p_b^z + p_{\ell^+}^z + p_\nu^z)^2, \\
m_{\bar{t}}^2 &= (E_{\bar{b}} + E_{\ell^-} + E_{\bar{\nu}})^2 - (p_{\bar{b}}^x + p_{\ell^-}^x + p_{\bar{\nu}}^x)^2 - (p_{\bar{b}}^y + p_{\ell^-}^y + p_{\bar{\nu}}^y)^2 - (p_{\bar{b}}^z + p_{\ell^-}^z + p_{\bar{\nu}}^z)^2,
\end{aligned} \tag{7}$$

where  $m_t = m_{\bar{t}} = 172.5$  GeV and  $M_{W^+} = M_{W^-} = 80.4$  GeV are the pole masses of the top (anti)quark and  $W^\pm$  boson respectively. There are several methods to reconstruct the  $t\bar{t}$  rest frame [73–77]. In this analysis we use an analytical method which aims at solving the quartic equation in neutrino momentum with the help of the Sonnenschein method [78]<sup>1</sup>.

<sup>1</sup>We have implemented this method using Ref. [79]. We have written the entire implementation that can be found in this [github](#) repository in the MADANALYSIS 5 framework.

The number of unknowns in equation 8 are the components of the neutrino momenta which are eight. First, we use equation 8 to eliminate  $E_\nu$  and  $E_{\bar{\nu}}$ . The longitudinal momenta  $p_\nu^z$  and  $p_{\bar{\nu}}^z$  can be further eliminated by other rearrangements of the two equations. Finally, we can enforce the conservation of the transverse missing momentum to eliminate  $p_\nu^y$  and  $p_{\bar{\nu}}^y$ , so we get

$$h_4 p_x^4 + h_3 p_x^3 + h_2 p_x^2 + h_1 p_x + h_0 = 0, \quad (8)$$

where  $p_x = p_\nu^x$  and  $h_0, h_1, h_2, h_3, h_4$  are numerical coefficients that depend on the four-momentum components of the bottom jets and charged leptons. The solution of this equation leads to a fourfold ambiguity. Removing the  $b$ -tagged jets that may result from QCD or the Higgs boson decay, we get two other possible combinations from matching the charged leptons with the  $b$ -jets and leading to an eightfold ambiguity. We can then obtain  $p_\nu^z$  and  $p_{\bar{\nu}}^z$  from the arrangement we have used previously to eliminate these. The energy component can be then obtained from

$$\begin{aligned} E_\nu &= \sqrt{(p_\nu^x)^2 + (p_\nu^y)^2 + (p_\nu^z)^2}, \\ E_{\bar{\nu}} &= \sqrt{(p_{\bar{\nu}}^x)^2 + (p_{\bar{\nu}}^y)^2 + (p_{\bar{\nu}}^z)^2}. \end{aligned} \quad (9)$$

An important step is to choose the best suited solution for the neutrino momenta. There are different methods for making this choice: (*i*) solution that minimises the invariant mass of the  $t\bar{t}$  system; (*ii*) solution that uses the target mass of the  $t\bar{t}$  resonance (which may be more suited for resonance searches); (*iii*) characterisation of the top quark decay products using kinematical information. We employ the latter method as it seems to be process-independent and have small biases. To do so we use the angular information encoded in the ‘lego’ distance ( $\Delta R$ ) between the (anti)top quark and its decay products and between the decay products themselves plus the ratios of the neutrino momenta with respect of the visible object momenta used in the reconstruction process. On an event-by-event basis, we construct these variables for each neutrino solution and we take the solution that maximises the following Likelihood function:

$$L_{\text{prob.}} = \prod_i P(\Delta R(\nu, i)) \times P\left(\frac{p_T^\nu}{p_T^i}\right) \times P\left(\frac{E_\nu}{E^i}\right), \quad (10)$$

where  $i$  refers to all charged leptons and  $b$ -tagged jets used in equation 8 and  $P$  is the probability assigned to each combination. We have found that our implementation yields very good results for the top (anti)quark invariant mass, both the individual ones and the one of the pair. We show the  $m_t$  and  $m_{t\bar{t}}$  resolution of our reconstruction in Fig. 3, where we compare the reconstructed masses against Monte Carlo (MC) truth information.

### 3 CP-sensitive observables

In this section, we briefly review the different angular and energy observables that we use in this study. In addition to low-level and high-level kinematics variables that we discuss in Sec. 5.1 we also employ 39 angular and energy-ratio features (some distributions are shown in Fig. 4). Note that these features have been studied extensively in the literature in different physical applications onto the top (anti)quark sector (see e.g. Refs. [36, 43, 47, 56, 77, 80–90]).

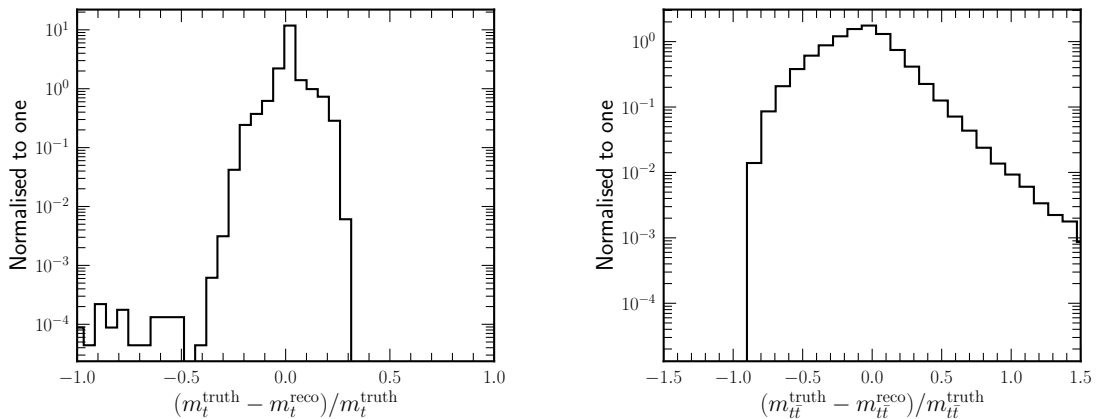


Figure 3: The  $m_t$  (left) and  $m_{t\bar{t}}$  (right) resolution of the analytical mass reconstruction method outlined in the text. Here,  $m_t^{\text{truth}}$  and  $m_{t\bar{t}}^{\text{truth}}$  refer to the invariant mass of the (anti)top and the  $t\bar{t}$  system in the last parton shower step before the decay of the (anti)top quark.

**Polarisation and spin-spin correlation observables.** They consist of some polar angles of the charged leptons from (anti)top quark decays. The generic expression can be written as follows [43, 80]:

$$\frac{1}{\sigma} \frac{d^2\sigma}{d\cos\theta_{\ell^a} d\cos\theta_{\ell^b}} = \frac{1}{4} \left( 1 + \alpha_{\ell^a} P_a \cos\theta_{\ell^a} + \alpha_{\ell^b} P_b \cos\theta_{\ell^b} + \alpha_{\ell^a} \alpha_{\ell^b} C_{ab} \cos\theta_{\ell^a} \cos\theta_{\ell^b} \right), \quad (11)$$

where  $\alpha_\ell = \pm 1$  is the spin analysing power of the charged lepton and  $\theta_{\ell^{a,b}} = \angle(\hat{\ell}^{a,b}, \hat{S}_{a,b})$ . Here,  $\hat{\ell}^{a,b}$  refers to the direction of flight of the charged lepton in the top (anti)quark rest frame and  $\hat{S}_{a,b}$  is the spin quantization axis in the basis  $a$ . We use three commonly studied bases<sup>2</sup>.

- Helicity basis ( $a = k$ ): The spin quantisation axis is defined as the direction of motion of the top (anti)quark in the  $t\bar{t}$  Zero-Momentum Frame (ZMF). In this case,  $\cos\theta_{\ell^+}^k$  is defined as

$$\cos\theta_{\ell^+}^k = \frac{\hat{p}_{\ell^+} \cdot \hat{p}_t^{\text{ZMF}}}{|\hat{p}_{\ell^+}| |\hat{p}_t^{\text{ZMF}}|} \quad (12)$$

and a similar expression holds for  $\cos\theta_{\ell^-}^k$ .

- Transverse basis ( $a = n$ ): In this case, the spin axis is defined as the three vector transverse to the production plane composed by the top quark direction of motion in the ZMF and the beam direction<sup>3</sup>. Therefore,  $\hat{S}_a$  is given by

$$\hat{S}_a = \hat{p}_t^{\text{ZMF}} \times \hat{p}_{\text{beam}}. \quad (13)$$

<sup>2</sup>See, e.g., Refs. [91, 92] for some recent measurements of spin-spin correlations in  $t\bar{t}$  production.

<sup>3</sup>As explained, in  $t\bar{t}\phi_0$  production, the contribution of gluon fusion dominates the production rate which makes the initial state Bose symmetric. Therefore, and by following the recommendation of Ref. [80], the value of  $\cos\theta_{\ell^{n,r}}$  is multiplied by the sign of the scattering angle  $\vartheta = \hat{\mathbf{p}} \cdot \hat{\mathbf{p}}_t$  with  $\hat{\mathbf{p}} = \mathbf{p}_t/|\mathbf{p}_t|$  the top quark direction of flight in the  $t\bar{t}$  ZMF and  $\hat{\mathbf{p}} = (0, 0, 1)$ .

- $r$ -basis ( $a = r$ ): This basis is defined as the transverse to the plane defined by the top quark direction of flight in the ZMF and the spin axis in the transverse basis, such that the three vectors form a complete orthogonal basis. We also weight  $\cos \theta_{\ell^\pm}^r$  by the sign of the scattering angle between the (anti)quark and a unit vector  $\hat{\mathbf{p}} = (0, 0, 1)$ .

The polarization of the (anti)top quark can be easily obtained by integrating 11 over the angle  $\theta_\ell^a$  (or  $\theta_\ell^b$ ):

$$\frac{1}{\sigma} \frac{d\sigma}{d \cos \theta_{\ell^\pm}^a} = \frac{1}{2} \left( 1 + \alpha_{\ell^\pm} P_{t,\bar{t}}^a \cos \theta_{\ell^\pm}^a \right). \quad (14)$$

**Laboratory frame observables.** We also consider some laboratory frame observables, i.e., observables that are constructed from the particle momenta in the laboratory frame. First, the difference in the azimuthal angles of the two charged leptons is a clean observable that is usually used to measure spin-spin correlations between the top and the antitop quarks in  $t\bar{t}$  production and decay. It was found in Ref. [36] that it can also serve as a good discriminator between the different CP hypotheses of the  $t\bar{t}\phi_0$  coupling. It is defined as

$$\Delta\phi_{\ell^+\ell^-} = |\phi_{\ell^+} - \phi_{\ell^-}|, \quad (15)$$

where  $\phi_{\ell^\pm}$  is the azimuthal angle of the charged leptons in the laboratory frame. We also study the sensitivity of an observable that relies on the reconstruction of the Higgs boson candidate [47]:

$$\cos \theta_{\ell\phi_0} = \frac{(\hat{p}_{\ell^+} \times \hat{p}_{\phi_0}) \cdot (\hat{p}_{\ell^-} \times \hat{p}_{\phi_0})}{|(\hat{p}_{\ell^+} \times \hat{p}_{\phi_0})| |(\hat{p}_{\ell^-} \times \hat{p}_{\phi_0})|}, \quad (16)$$

with  $\hat{p}_{\ell^+}$ ,  $\hat{p}_{\ell^-}$  and  $\hat{p}_{\phi_0}$  being the directions of flight of the positively-charged, negatively-charged lepton and of the reconstructed Higgs boson candidate in the laboratory frame. Thus,  $\theta_{\ell\phi_0}$  defines the angle spanned by the dilepton system projected on the plane that is orthogonal to the Higgs boson candidate momentum.

Another angle can be defined from  $\theta_{\ell\phi_0}$ , as follows:

$$\cos \tilde{\theta}_{\ell\phi_0} = \lambda \cos \theta_{\ell\phi_0}, \quad (17)$$

with  $\lambda = \text{sign}((\hat{p}_b - \hat{p}_{\bar{b}}) \cdot (\hat{p}_{\ell^-} \times \hat{p}_{\ell^+}))$  and  $\hat{p}_b$  and  $\hat{p}_{\bar{b}}$  the directions of flight of the  $b$ -(anti)quarks forming the Higgs boson candidate.

We also include some observables introduced in Ref. [56], from where we show the definition of the most sensitive observable (and to where we refer the reader for more details), as follows:

$$\cos \omega_1 = \frac{\hat{p}_{\phi_0} \cdot (\hat{p}_{\ell^+} \times \hat{p}_{\ell^-}) \hat{p}_{\phi_0} \cdot (\hat{p}_{\ell^+} - \hat{p}_{\ell^-})}{|\hat{p}_{\phi_0}|^2 |\hat{p}_{\ell^+} \times \hat{p}_{\ell^-}| |\hat{p}_{\ell^+} - \hat{p}_{\ell^-}|}. \quad (18)$$

**Energy-ratio observables.** Observables based on the ratios of the energies of the top (anti)quark and its decay products may carry some information on its polarisation/helicity state (see, e.g., Refs. [81–90]). We define these as

$$u = \frac{E_\ell}{E_\ell + E_b}, \quad z = \frac{E_b}{E_t}, \quad x_\ell = \frac{2E_\ell}{m_t}, \quad (19)$$

with  $E_\ell$ ,  $E_b$  and  $E_t$  being the energies of the charged lepton,  $b$ -quark and top quark in the laboratory frame.



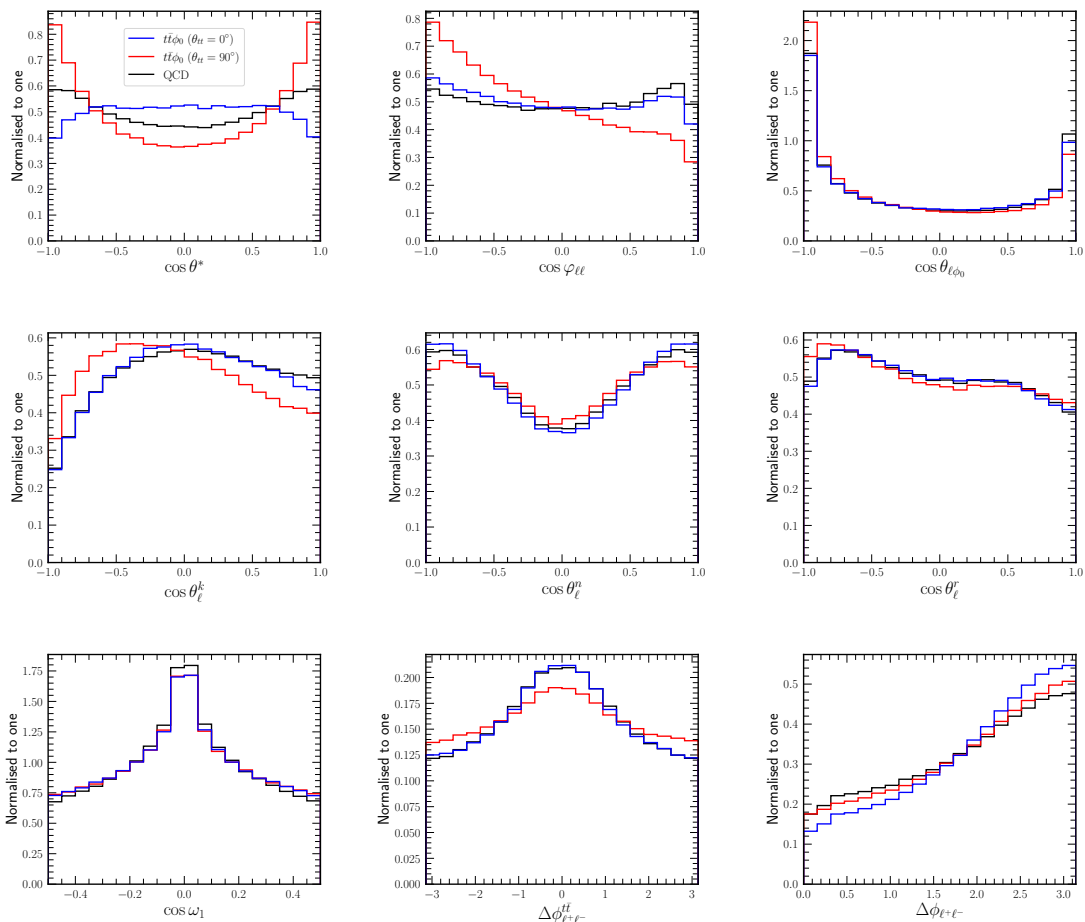


Figure 4: Examples of some distributions of the angles that are sensitive to the  $CP$  nature of the  $t\bar{t}\phi_0$  coupling for  $t\bar{t}\phi_0$  with  $\theta_t = 0^\circ$  (blue),  $t\bar{t}\phi_0$  with  $\theta_t = 90^\circ$  (red) and QCD backgrounds (black). In the upper panels we show  $\cos \theta^*$  (left),  $\cos \varphi_{\ell\ell}$  (middle) and  $\cos \theta_{\ell\phi_0}$  (right). In the middle panels we show the cosine of the polar angle in the helicity basis (left), in the transverse basis (middle) and in the  $r$ -basis (right). The lower panels show  $\cos \omega_1$  (left),  $\Delta\phi_{\ell^+\ell^-}^{t\bar{t}}$  (middle) and  $\Delta\phi_{\ell^+\ell^-}$  (right). All the distributions are normalised to unity and the calculations are done for  $\kappa_t = 1$ .

**Other observables.** We start by considering the opening angle between the two oppositely charged leptons which is defined as

$$\cos \varphi_{\ell_a \ell_b} = \frac{\hat{p}_{\ell^+} \cdot \hat{p}_{\ell^-}}{|\hat{p}_{\ell^+}| |\hat{p}_{\ell^-}|}, \quad (20)$$

where  $\hat{p}_{\ell^+}$  ( $\hat{p}_{\ell^-}$ ) is the direction of flight of the charged lepton  $\ell^+$  ( $\ell^-$ ) in the  $t\bar{t}$  ZMF. We also include two observables studied in Ref. [77]. The first observable is defined as the angle between the top quark direction of flight in the  $t\bar{t}$  ZMF and the beam three-momentum, i.e.,

$$\cos \theta^* = \frac{\hat{p}_t^{\text{ZMF}} \cdot \hat{p}_{\text{beam}}}{|\hat{p}_t^{\text{ZMF}}| |\hat{p}_{\text{beam}}|}. \quad (21)$$

The last observable, denoted by  $\Delta\phi_{\ell^+\ell^-}^{t\bar{t}}$ , which is defined as the angle spanned by the direction of the two leptons on the orthogonal plane to the top quark 3-momentum in the

$t\bar{t}$  ZMF

$$\Delta\phi_{\ell^+\ell^-}^{t\bar{t}} = \text{sign}[\hat{p}_t \cdot (\hat{p}_{\ell^+} \times \hat{p}_{\ell^-})] \arccos\left(\frac{(\hat{p}_{\ell^+}^{\text{ZMF}} \times \hat{p}_t^{\text{ZMF}}) \cdot (\hat{p}_{\ell^-}^{\text{ZMF}} \times \hat{p}_t^{\text{ZMF}})}{|\hat{p}_{\ell^+}^{\text{ZMF}} \times \hat{p}_t^{\text{ZMF}}| |\hat{p}_{\ell^-}^{\text{ZMF}} \times \hat{p}_t^{\text{ZMF}}|}\right). \quad (22)$$

## 4 Conditional DNNs

A conditional DNN for classification is a NN architecture where the classification process is conditioned on additional input information beyond the raw data. This additional information, often referred to as conditioning variables or features, can provide context or guidance to the NN, improving its ability to make accurate predictions. The network takes as input both the raw data to be classified and additional conditioning information. If the conditioning information is not directly compatible with the raw input data, it may need to be processed or transformed into a compatible format. This could involve feature extraction techniques such as encoding categorical variables, dimensionality reduction or any other preprocessing steps necessary to integrate the conditioning information with the input data. The network architecture is designed to incorporate the conditioning information into the classification process. This involves concatenating the conditioning information with the input data at later layers, passing it through additional conditioning layers to selectively focus on relevant parts of the input data based on the conditioning information, specifically, the angle  $\theta_{tt}$ . The network is trained to classify the input data into the appropriate classes while taking into account the provided conditioning information. This enables the network to interpolate the results of different values of the conditional variable,  $\theta_{tt}$ , which the model did not trained on. In general, a conditional DNN is trained on a set of feature variables  $x$  and a conditioned value  $\theta_{tt}$ , in which the network output is

$$\hat{y} = \mathcal{F}(x, \theta_{tt}), \quad (23)$$

where  $\mathcal{F}$  is the nonlinear function learned by the network to classify the input features  $x$  conditioned by the value of  $\theta_{tt}$ . The training objective typically involves minimizing a classification loss function, such as cross-entropy loss, computed between the predicted class probabilities and the true class labels, while also considering any regularization terms to prevent overfitting. In this paper, we utilize two conditional DNNs with different structure, namely, a conditional MLP and a conditional (multi-modal) GNN. Both networks are trained on four signal points with  $\theta_{tt} = 0^\circ, \pm 30^\circ, \pm 45^\circ, \pm 90^\circ$  and interpolate the results for  $\theta_{tt} = \pm 15^\circ, \pm 60^\circ, \pm 75^\circ$ .

### 4.1 MLP

Commencing with high-level kinematical distributions, we utilize a MLP model to enhance the discrimination between signal and background distributions. These distributions encapsulate distinctive information regarding the overall structure of both signal and background events. Consequently, the architecture of the MLP network, comprising fully connected layers, can discern global features effectively, resulting in robust classification power between signal and background events. Moreover, the fully connected layers that process the kinematical features are comprised by one linear layer which encodes the values of the condition parameter  $\theta_{tt}$ . In this case, the MLP is able to learn global features of the signal and background events that can be used to increase the signal to background yield assigned to each value of  $\theta_{tt}$ , which then enables the model to interpolates the values of  $\theta_{tt}$  that the network is not trained on it. Accordingly, the MLP is trained on specific values of the angle and is used to provide the sensitivity to all other values of  $\theta_{tt}$ .

Despite the MLP capability to achieve high classification performance, the similarity in kinematical structures between certain background distributions and signal ones diminishes the overall classification efficiency. However, by implementing initial cuts that maximize the signal-to-background yield prior to inputting the distributions into the MLP, one can augment the classification performance. However, the constructed kinematical spectra demonstrate significant intercorrelation, such that applying a cut to any distribution may influence the structure of others, consequently impeding the MLP classification performance. To mitigate the global impact of these initial cuts, it is imperative to decorrelate such dependencies across kinematical variables, either via the square-root of the covariance matrix or Gaussian transformation of variables, as elucidated in [93]. Ultimately, while initial cuts may bolster classification performance, we have chosen not to apply these, thereby affording the MLP with complete autonomy in identifying optimal classification boundaries.

The MLP structure consists of two input layers. The first input layer is used to encode the features with 103 neurons, which is the number of the used features. The second input layer encodes the value of the condition parameter  $\theta_{tt}$  and it has only one neuron. The first input layer is followed, sequentially, by three fully connected layers with 256, 128 and 64 neurons and ReLU activation function. Each fully connected layer is followed by a dropout layer with dropout rate of 20%. The second input layer is followed by one linear layer<sup>4</sup> with 64 neurons, to adjust the dimension with the last fully connected layer from the first stream. The final layers from the two streams are concatenated using a concatenation layer where the output is passed directly to an output layer with one neuron and sigmoid activation function.

## 4.2 GNN

GNNs represent a class of DL models specifically designed for processing graph-structured data, where a graph is a set of nodes/vertices connected by edges  $G(V, E)$ . By effectively leveraging the inherent connectivity and relational information within graphs, GNNs excel in capturing complex patterns that are not easily accessible to traditional NN architectures. Central to their operation is the message passing mechanism, where node representations are iteratively updated by aggregating features from their neighboring nodes, thus encoding both local and global graph structures into the learning process. The versatility of GNNs is evident in their wide range of applications across various tasks. In graph classification, GNNs aim to predict the labels of entire graphs based on their structure and node features, which is crucial in domains like chemical compound analysis and social network studies. Node classification, another prominent task, involves predicting the labels of individual nodes within a graph, often used in citation networks and social media to identify categories or communities. Additionally, GNNs are employed in link prediction to forecast potential connections between nodes, which has significant implications for recommender systems and network analysis. Through these tasks, GNNs demonstrate their powerful capability to model complex relational data effectively. GNNs can be broadly categorized into several types based on their architecture and the specific methods they employ for node information aggregation. For instance, Graph Convolutional Networks (GCNs) [94] utilize a convolutional approach, adapting the traditional convolution operations to work directly on graphs. GCNs can extract meaningful features from nodes and edges, achieving state-of-the-art performance in graph classification tasks. This adaptability and capability to model complex relationships make GCNs a versatile and promising approach in HEP analysis.

---

<sup>4</sup>It is important to keep the mapped conditioned parameter without any activation function.

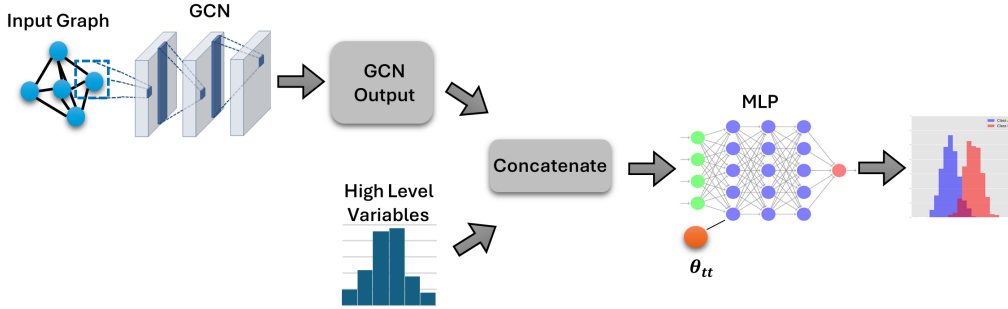


Figure 5: Schematic diagram of the (multi-modal) GNN, illustrating the flow from the input graph through the GCN to generate the final output. This output is concatenated with the high-level variables and processed through a MLP, this resulting in the final output classification.

For the purpose of constructing a conditional GNN, our approach integrates a multi-modal network combining GCN with MLP. The GCN component is particularly advantageous for incorporating the topological relationships among nodes and edges, thereby facilitating the learning of graph-structured data. In our methodology, we represent the reconstructed particles, their daughters, and parent particles as nodes within the graph. Following the methodology outlined in [95], each node  $i$  in the input layer is represented as a feature vector  $x = (I_1, I_2, I_3, I_4, p_T, E, \eta, \phi)$ . This vector encapsulates the properties of the corresponding particle, where  $p_T$  denotes transverse momentum,  $E$  is the energy,  $\eta$  represents pseudorapidity and  $\phi$  is the azimuthal angle. Initially, the values of  $I_1$  through  $I_4$  are set to zero. The indicator  $I_1$  is set to 1 if the particle is a lepton,  $I_2$  is set to 1 if the particle is a  $b$ -jet,  $I_3$  is set to 1 if it is a neutrino and  $I_4$  is set to 1 if it is a reconstructed top (anti)quark (and 0 otherwise, in all cases). The input graph to the GCN consists of 11 nodes, specifically 2 leptons, 4  $b$ -jets, 2 neutrinos, 2 top (anti)quarks and the Higgs boson, each characterized by the aforementioned feature vector  $x$ . The graph is fully connected, with edges weighted by the angular distance  $\Delta R_{(x_i, x_j)}$  between the particles in nodes  $i$  and  $j$ , thereby enabling the model to capture intricate spatial relationships between such particles. Our model consists of 3 GCN layers, as first introduced in [94], with ReLU activation. This is followed by max pooling to aggregate the node embedding. The other part of the model is a basic set of 3 fully connected layers (another MLP), which takes the high-level variables mentioned earlier that are used to train the basic MLP model. In addition, it also takes the output of the GCN model. Both inputs are concatenated and fed into the 3 fully connected layers. Our hybrid approach is trained conditionally similar to the basic MLP on specific values of the angle and is used to provide the sensitivity to all other values of  $\theta_{tt}$ . The architecture schematic diagram, depicting all inputs and outputs, is shown in Fig. 5.

For the optimization process, we employed a learning rate of  $1 \times 10^{-4}$  and a weight decay parameter of  $1 \times 10^{-4}$ . These values were selected to balance efficient learning with the stability of the model. All models developed in our study were constructed using the PYTORCH GEOMETRIC framework [96], a powerful and efficient library designed to facilitate the implementation of graph-based DL models.

## 5 DL analysis

In this section we discuss the analysis taking place through the MLP and GNN models described. Starting with low level information about the reconstructed final state particles, we discuss how all features for the two DNNs are considered.

### 5.1 Signal and background kinematics

Considering the leptonic decays of the top (anti)quark and those of the Higgs boson into  $b\bar{b}$ , the final state will consist of two charged leptons ( $\ell = e, \mu$ ) with opposite electric charge, at least four  $b$ -tagged jets and missing energy. The main background processes arise from the QCD production of  $t\bar{t}$  in association with two jets ( $t\bar{t}b\bar{b}$ ,  $t\bar{t}jj$ ). As our analysis relies on exactly four  $b$ -tagged jets, background processes such as  $t\bar{t}$ , multi-jets and  $V$ +jets are subleading. We first apply some basic generator-level cuts on parton-level objects like electrons, muons, partons and missing energy, i.e.:

$$p_T(\ell) > 20 \text{ GeV}, |\eta_\ell| < 2.5, p_T(j) > 25 \text{ GeV}, |\eta_j| < 2.5, E_T^{\text{miss}} > 20 \text{ GeV}. \quad (24)$$

At the reconstruction level, we require that events do not contain any isolated hadronically decaying  $\tau$  lepton with  $p_T > 30 \text{ GeV}$  and  $|\eta| < 2.5$ . Then we require exactly two charged leptons (electrons or muons) with  $p_T > 25 \text{ GeV}$  and  $|\eta| < 2.4$  excluding electrons in the transition region in the calorimeter (i.e., those with  $1.44 < |\eta| < 1.57$ ). The charged leptons are required to be isolated using tight isolation criteria. We then require that  $E_T^{\text{miss}} > 20 \text{ GeV}$ . We impose that events should contain at least four  $b$ -tagged jets with  $p_T > 30 \text{ GeV}$  and  $|\eta| < 2.5$ . The combination of the two  $b$ -tagged jets in the event that match the charged leptons will be used for the top quark reconstruction as described in Sec. 2.2 while the remaining  $b$ -tagged jets are ordered in transverse momentum for which case the two leading  $b$ -tagged jets are used to form  $\phi_0$  candidates. We require that the  $t\bar{t}$  invariant mass to lie in the window of  $[132.5, 210.5] \text{ GeV}$ . We, however, do not impose any requirement on the invariant mass of the  $\phi_0$  candidate. After all the events pass the basic selection criteria, we explore the following variables for more sophisticated DL analyses.

**CP-sensitive variables.** They consist of 39 variables which were described in detail in Sec. 3.

**Low-level variables.** Low-level variables consist of the four components of the momenta of the top (anti)quarks and Higgs boson decay products. There are 32 variables in total, as follows.

- The two neutrino candidates from the solution of the Likelihood-based top (anti)quark reconstruction method (see Sec. 2.2):

$$\eta(\nu_1), p_T(\nu_1), E(\nu_1), \phi(\nu_1), \eta(\nu_2), p_T(\nu_2), E(\nu_2), \phi(\nu_2).$$

- The four momenta of the two charged leptons:

$$\eta(\ell_1), p_T(\ell_1), E(\ell_1), \phi(\ell_1), \eta(\ell_2), p_T(\ell_2), E(\ell_2), \phi(\ell_2).$$

- The four momenta of the four  $b$ -tagged jets:

$$\begin{aligned} &\eta(b_1), p_T(b_1), E(b_1), \phi(b_1), \eta(b_2), p_T(b_2), E(b_2), \phi(b_2), \\ &\eta(b_3), p_T(b_3), E(b_3), \phi(b_3), \eta(b_4), p_T(b_4), E(b_4), \phi(b_4). \end{aligned}$$

**High-level variables.** They consist of more complicated variables which are built upon the low-level variables. There are 30 variables in total, as follows.

- Invariant mass of the top quark, top antiquark and Higgs boson candidates:  $m_{t_1}, m_{t_2}$  and  $m(b_3b_4)$ .
- The components of the four momenta of the top quark, top antiquark and Higgs boson candidates:

$$\begin{aligned} &\eta(t_1), p_T(t_1), E(t_1), \phi(t_1), \\ &\eta(t_2), p_T(t_2), E(t_2), \phi(t_2), \\ &\eta(b_3b_4), p_T(b_3b_4), E(b_3b_4), \phi(b_3b_4). \end{aligned}$$

- The invariant mass, the energy and the transverse momentum of the  $t\bar{t}$  system:

$$m(t_1t_2), p_T(t_1t_2), E(t_1t_2).$$

- The invariant mass of the  $t\bar{t}\phi_0$  system:  $m(t_1t_2b_3b_4)$ .
- The transverse of the dilepton system:  $p_T(\ell_1\ell_2)$ .
- The scalar sum of the jet transverse momenta including all the  $b$ -tagged jets in the events

$$H_T^b \equiv \sum_i p_T(b_i).$$

- The effective mass

$$M_{\text{eff}} \equiv H_T^b + p_T(\ell_1) + p_T(\ell_2).$$

- The minimum and maximum of the transverse momentum and the invariant of the top (anti)quark and the  $b$ -tagged jets forming the Higgs boson candidates:

$$p_T^1 \equiv \max_i \{p_T(b_i, t_1), p_T(b_i, t_2)\}, \quad p_T^2 \equiv \min_i \{p_T(b_i, t_1), p_T(b_i, t_2)\}.$$

## 5.2 MC event generation and simulation tools

The effective Lagrangian of Sec. 2 is implemented in FEYNRULES [97]. The output file in UFO format [98] is used as an input to MADGRAPH5\_AMC@NLO [64] to generate parton-level samples for both the signal and background events. As mentioned, all signal and background processes are simulated at LO in QCD. To keep track of spin and correlation effects, we decay the intermediate resonances by using MADSPIN [99]. PYTHIA version 8309 is used to add parton showering, hadronisation and heavy hadron decays to the event samples [100]. The detector response is modelled using the Simplified Fast-detector Simulator (SFS) [101] in the MADANALYSIS 5 framework [102–105]. The simulation of the reconstructed objects such as tracks, isolated electrons, jets, hadronically-decaying  $\tau$  lepton and missing transverse energy ( $E_T^{\text{miss}}$ ) is done following the same lines of Ref. [106], which is based on a CMS analysis targeting the search of right-handed gauge bosons in the two leptons and two jet events [107]. We have, however, slightly modified the detector card in this analysis by assuming a flat 85%  $b$ -tagging and 10% mis-tagging efficiencies across all the pseudorapidity and transverse momentum values. Jets are clustered with the anti- $k_t$  algorithm [108] with a jet radius of  $R = 0.4$  using FASTJET version 3.4.1 [109].

As mentioned, for the DNN analysis, we use PYTORCH GEOMETRIC[96] for building the GNN network, while standard PYTORCH [110] is used for the MLP. Finally, the SIKIT-LEARN package [111] is used to facilitate network training and evaluation.

### 5.3 Training the network

Once the datasets are prepared, we commence training the networks to understand the complex, non-linear relationships between the input data and their corresponding labels. Additionally, the network learns to interpolate between the trained values of the conditioned parameter. Events are organized into a feature dataset with dimensions  $(n, 103)$ , where  $n$  represents the number of events in the training dataset, and a conditional input vector with the value of  $\theta_{tt}$ . The training dataset is composed of equally-sized events generated with  $\theta_{tt} = 0^\circ, \pm 30^\circ, \pm 45^\circ$ , and  $\pm 90^\circ$ , each containing 100000 events, resulting in a training signal dataset of size 400000 events. The conditional parameter vector is prepared with the exact value of  $\theta_{tt}$  in radians to facilitate better convergence of the network to the minimum of the loss function.

For the background dataset, the conditional parameter input is a vector of the same length as the training background dataset, containing random values between 0 and  $\pi/2$ . Since the primary objective of the network is to learn a global pattern of the signal events and interpolate between the trained values, we utilize training datasets of size 400000 and 300000 for both signal and background, respectively. This strategy is akin to assigning higher weights to the signal events, enabling the network to focus more on learning the features of the signal events.

For network evaluation, we utilize equal-sized, new, unseen datasets for the signal and background, each consisting of 100000 events. For each of the training datasets we assign the label  $Y = 1$  for signal events while for background events we assign the label  $Y = 0$ . In order to remove the network dependence on the position of the signal and background events, we stack the signal and background events in one data set and shuffle it together with the assigned labels. During the network training stage, during each epoch (defined as number of passes of the entire data sets), the network updates the weights assigned to the neurons for each event via backward propagation of errors. The network then tries to minimize the error between its predictions and the true labels by reaching a global minimum of some loss function. For this purpose we use a binary cross-entropy as a loss function and a Adam optimizer to optimize the network convergence to the global minimum of the loss function. Finally, the network repeats the process until it reaches the desired accuracy. Once the model is trained, we test it by using completely unseen new data sets to measure the network performance.

We prepared seven test datasets according to the value of  $\theta_{tt} = 0^\circ, \pm 15^\circ, \pm 30^\circ, \pm 45^\circ, \pm 60^\circ, \pm 75^\circ$  and  $\pm 90^\circ$ . These datasets are prepared of equal size of the signal events and background events. We stress here, that the network is trained on signal events with  $\theta_{tt} = 0^\circ, \pm 30^\circ, \pm 45^\circ$  and  $\pm 90^\circ$ , and it used to interpolate the points in between.

For all networks, we train the model with 20 epochs with batch size equalling a 500 sample. The dimension of the final output probability,  $\hat{Y}$ , is  $1 \times 2$ ,  $(P_{sig}, P_{bkg})$ , with  $P$  ranging between  $[0, 1]$ . If  $P_{sig} > 0.5$  ( $P_{bkg} < 0.5$ ), the corresponding event is classified as most likely being a signal event and if  $P_{sig} < 0.5$  ( $P_{bkg} > 0.5$ ) the corresponding event is classified as most likely being a background event.

## 6 Results

In this section we present the results of using a conditional MLP and GNN to probe the CPV phase in  $pp \rightarrow t\bar{t}\phi_0$  production, followed by  $t\bar{t} \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}l\bar{\nu}l\nu$  and  $\phi_0 \rightarrow b\bar{b}$ , at the HL-LHC with center-of-mass energy of 13.6 TeV. The discriminative ability of each network determines how effectively it distinguishes between signal and background features, a metric quantified by the Receiver Operating Characteristic (ROC) curve. Enhanced discrimination

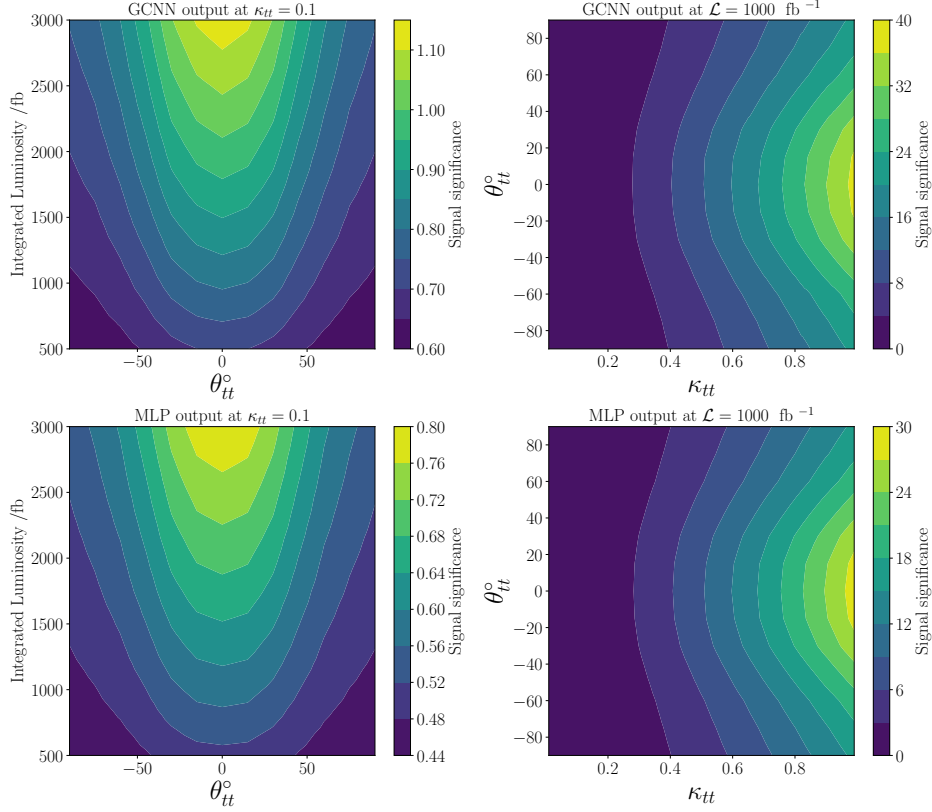


Figure 6: Left: contour plot for varying integrated luminosity and  $\theta_{tt}$  at fixed value of  $\kappa_{tt} = 0.1$ . Right: contour plot for varying  $\theta_{tt}$  and  $\kappa_{tt}$  at fixed integrated luminosity  $L = 1000 \text{ fb}^{-1}$ . The contour colors represent the value of the signal significance. Finally, the upper row represents the GNN results while the lower one represents the MLP results.

performance is indicated by a higher true positive rate compared to the false positive rate in the ROC curve. To optimize the performance of each network, we individually adjust the cut on the ROC curve, aiming to boost the signal-to-background yield by calculating  $\frac{S}{\sqrt{S+B}}$  at each bin of the ROC curves. Post-application of these cuts, the number of signal and background events is taken into account to compute the signal significance and ensuing limits.

This enables us to test the signal discovery hypothesis or setting an upper limit on the total cross section under the non-observation hypothesis. These can be determined by optimizing the signal-to-background cut on the DNN output, achieved using the following significance formula [112–115]:

$$\sigma_{sys} = \left[ 2 \left( (N_s + N_b) \ln \frac{(N_s + N_b)(N_b + \sigma_b^2)}{N_b^2 + (N_s + N_b)\sigma_b^2} - \frac{N_b^2}{\sigma_b^2} \ln \left( 1 + \frac{\sigma_b^2 N_s}{N_b(N_b + \sigma_b^2)} \right) \right) \right]^{1/2}, \quad (25)$$

where  $N_s$  and  $N_b$  represent the counts of signal and background events, respectively, and  $\sigma_b$  denotes the total uncertainty in the background events. Fig. 6 shows the signal significance on the  $\kappa_{tt}$  for  $\theta_{tt}$  values ranges from  $-\frac{\pi}{2}$  to  $\frac{\pi}{2}$  for the GNN (top) and MLP (bottom). Left plots display the contours for varying integrated luminosity and  $\theta_{tt}$  at fixed value of  $\kappa_{tt} = 0.1$  while the right plots display the contours for varying  $\theta_{tt}$  and  $\kappa_{tt}$  at fixed integrated luminosity  $L = 1000 \text{ fb}^{-1}$ . For all results the GNN shows an improvement over the MLP.

As mentioned previously, the conditional network is used to interpolate the significance



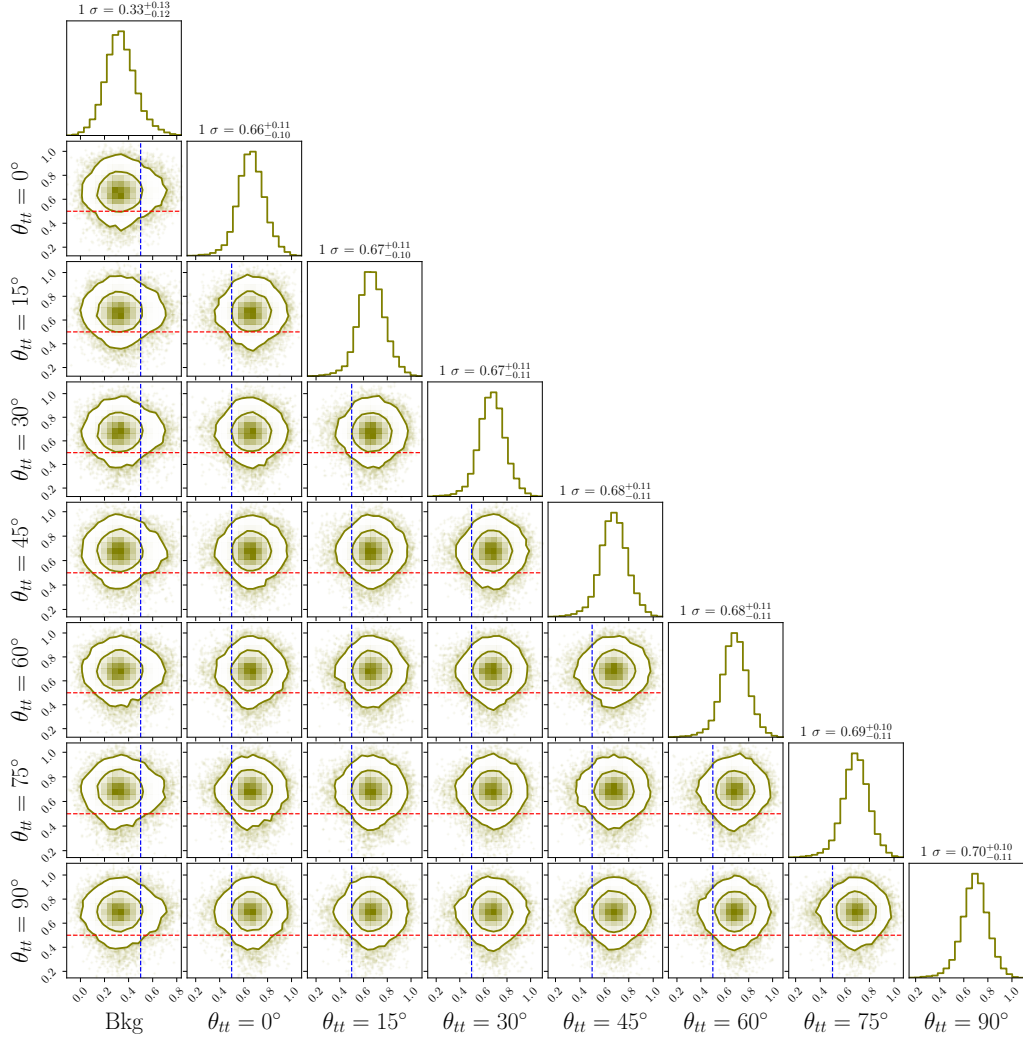


Figure 7: Contour plots for 1 (68% CL) and 2 sigma (95% CL) of the MLP output when tested on different signal benchmark points using 30000 test events. The model is trained on  $\theta_{tt} = 0^\circ, 30^\circ, 45^\circ, 90^\circ$  and interpolates the results for  $\theta_{tt} = 15^\circ, 60^\circ, 75^\circ$ . The 50% quantile and  $1\sigma$  error are written on top of each histogram. Events with MLP output near 1 are considered as most likely signal events while those with MLP output near 0 are considered as most likely background events. Blue and red dashed lines indicates the 0.5 value of the network output. For background correlation with all signal points, the upper left corner represents the true classified signal and background events while other corners represent the mis-identified rates. For signal-to-signal correlations the upper right corner represents the true classified events.

for new signal points with different  $\theta_{tt}$  values. To test the network performance for the interpolation we test the correlation of the network output of the background and all tested signal events. Fig. 7 shows the contour plots for 1 (68% CL) and 2 sigma (95% CL) of the MLP output when tested on different signal benchmark points using 30000 test events<sup>5</sup>. The model is trained on  $\theta_{tt} = 0^\circ, \pm 30^\circ, \pm 45^\circ, \pm 90^\circ$  and interpolates the results for  $\theta_{tt} = \pm 15^\circ, \pm 60^\circ, \pm 75^\circ$ . The 50% quantile and  $1\sigma$  error is written on top of each histogram. Events with MLP output near 1 is considered as most likely signal events,

<sup>5</sup>We opted to present the MLP results only, with the positive phase values, as the GNN output has similar response for interpolation.

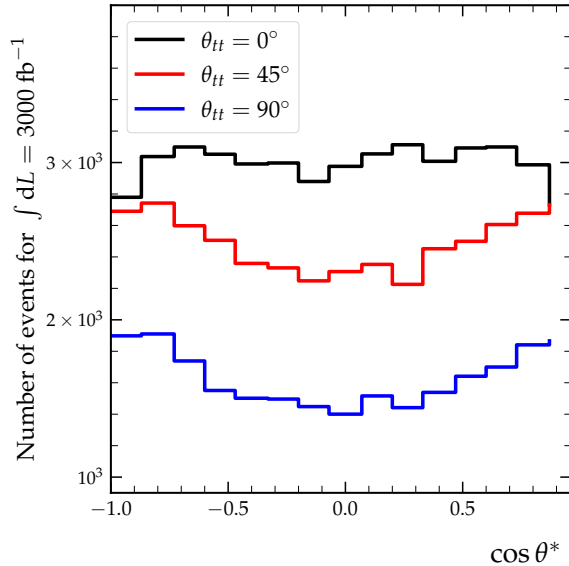


Figure 8:  $\cos\theta^*$  distribution, defined in eq. 21, in the signal region defined from GCN training for  $\theta_{tt} = 0^\circ$  (black),  $\theta_{tt} = 45^\circ$  (red) and  $\theta_{tt} = 90^\circ$  (blue). For all the cases, the background contribution is added. Here, we assume that  $\epsilon_b = 70\%$ .

while those with MLP output near 0 are considered as most likely background events. Blue and red dashed lines indicates the 0.5 value of the network output. For background correlation with all signal points, the upper left corner represents the true classified signal and background events while other corners represent the mis-identified rates. For signal-to-signal correlations the upper right corner represents the true classified events. In short, it is clearly shown that the network is able to correctly interpolate to new points within  $1\sigma$  level. Furthermore, this figure also indicates that conditional MLP has a good interpolation performance to new points within  $1\sigma$  level.

We close this section by showing the results of the shape analysis of the observables after the MLP and GNN optimization. Events passing the signal region definition, are used for the calculation of the Log-Likelihood function defined as

$$\log(\mathcal{L}) = - \sum_i \left[ n_i \log \left( \frac{n_i}{\nu_i} \right) + \nu_i - n_i \right], \quad (26)$$

with  $n_i$  and  $\nu_i$  being the number of events for the expected ( $\theta_{tt} = 0^\circ + \text{QCD}$ ) and the alternative ( $\theta_{tt} \neq 0^\circ + \text{QCD}$ ) hypotheses respectively. The sum runs over all the bins and all the observables being used. For the MLP case, we have used four observables as input:  $\Delta\phi_{\ell+\ell^-}$ ,  $\cos\varphi_{\ell\ell}$ ,  $\Delta\phi_{\ell\ell}^{tt}$  and  $\cos\theta^*$  while for the GNN case we have used three observables:  $\cos\varphi_{\ell\ell}$ ,  $\cos\theta_\ell^k$  and  $\cos\theta^*$  (an example is shown in Fig. 8). It is found that these observables are the most sensitive to the CP nature and show the best feature importance in the training stage of the NN algorithms. The results for the binned log-likelihood are shown in Fig. 9 for  $300 \text{ fb}^{-1}$  (left) and  $3000 \text{ fb}^{-1}$  (right) for the two networks being considered in our analysis. For  $\kappa_{tt} = 1$ , the GNN algorithm shows a superior performance as compared to the MLP algorithm as we can see that already  $300 \text{ fb}^{-1}$  is enough to probe  $\theta_{tt}$  of order  $20^\circ$ . We summarize below the expected exclusions at 95.4% CL:

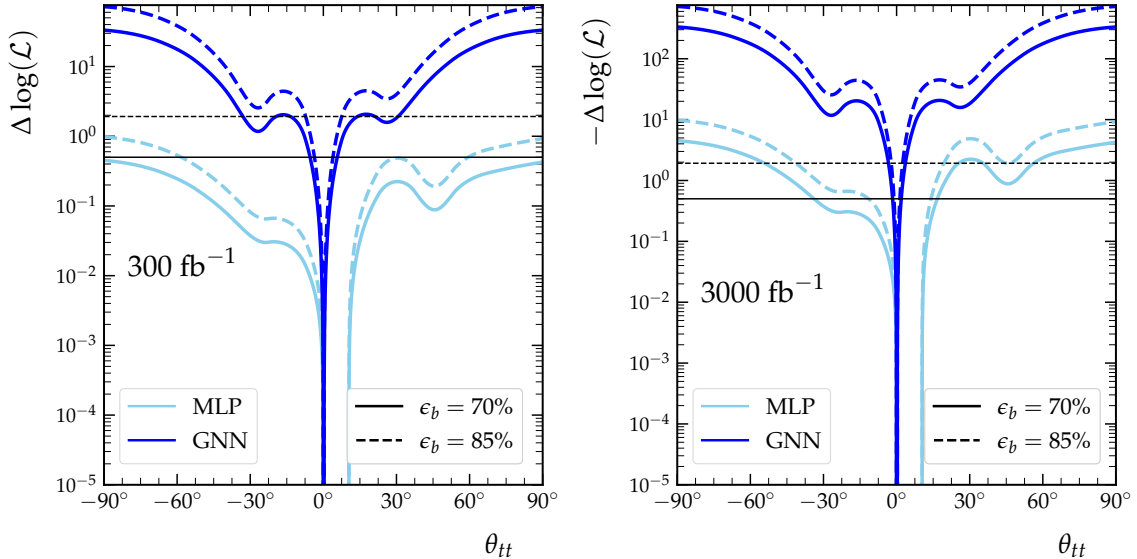


Figure 9: The log likelihood as a function of  $\theta_{tt}$  for the two networks: MLP (cyan) and GNN (blue) of  $300 \text{ fb}^{-1}$  (left) and  $3000 \text{ fb}^{-1}$  (right). In each case we show the results for two nominal  $b$ -tagging efficiencies of 70% (solid) and 85% (dashed). The solid and dashed black lines correspond to the 68.4% and 95.3% confidence levels respectively. All the results are shown for  $\kappa_{tt} = 1$ .

$$\begin{aligned}
 \theta_{tt}^{\text{MLP}} &\in \begin{cases} [-60^\circ, 60^\circ] & \text{for } \mathcal{L} = 300 \text{ fb}^{-1}, \\ [-45^\circ, 30^\circ] & \text{for } \mathcal{L} = 3000 \text{ fb}^{-1}. \end{cases} \\
 \theta_{tt}^{\text{GNN}} &\in \begin{cases} [-10^\circ, 10^\circ] & \text{for } \mathcal{L} = 300 \text{ fb}^{-1}, \\ [-5^\circ, 5^\circ] & \text{for } \mathcal{L} = 3000 \text{ fb}^{-1}, \end{cases} \quad (27)
 \end{aligned}$$

where we assumed  $\epsilon_b = 85\%$ . We note that the results for case of  $\epsilon_b = 70\%$  are not extremely different as can be seen in Fig. 9. Moreover, GNN results surpass the combined analysis in [59] in which the CP phase is excluded for  $|\theta_{tt}| \geq 18^\circ$  at  $\sqrt{s} = 14 \text{ TeV}$  and  $\mathcal{L} = 3000 \text{ fb}^{-1}$ , assuming  $\kappa_{tt} = 1$ .

## 7 Conclusions

DL is currently the state-of-the-art approach in many ML applications to particle physics (amongst other disciplines), yet the evaluation and training of DL models is generally still quite time-consuming and altogether computationally expensive. The so-called conditional computation approach has been thus proposed to tackle such a problem, as it operates by selectively activating only parts of the concerned network at a time. We have thus embraced it here in two different implementations: MLP and GNN.

Armed with such computational tools, we have chosen a particle physics problem which is particularly suited to these approaches, given the large multiplicity of the final state (having eight particles to start with), the necessity of reconstructing masses of intermediate objects (five of these) and, finally, the need of extracting subleading CPV effects from a large variety of kinematical observables (nine of these). The target process was  $pp \rightarrow t\bar{t}\phi_0$ , where  $\phi_0$  is a generic neutral Higgs boson, which is currently being targeted by the ATLAS

and CMS collaborations for the purpose of measuring the Yukawa coupling between the top (anti)quark and the SM-like Higgs boson discovered in 2012 ( $h_{\text{SM}}$ ). In such a process, this is done in so-called ‘open production’ (wherein the top (anti)quark is produced as a real object in the final state), so as to be able to compare it against the same coupling measured in so-called ‘closed production’ (wherein the top (anti)quark is produced as a virtual object in the loop of gluon-gluon fusion). Possibly more importantly, the extraction of such a coupling in the former case, unlike the latter one, offers the unique possibility of testing the CP properties of the interaction between  $\phi_0$  and  $t\bar{t}$ , by exploiting correlations amongst the momenta of the decay products of both the top (anti)quark pair and Higgs boson. In particular, in our analysis, we have assumed leptonic decays of the  $t\bar{t}$  system and  $b\bar{b}$  ones of the  $\phi_0$  state. Given that such CP properties can only be accessed through differential distributions (rather than inclusively at integrated cross section level), specifically, through their characteristic line-shapes, a significant number of events is necessary for this purpose, hence, as collider setup, we have chosen here the HL-LHC.

Following a sophisticated MC analysis based on, again, state-of-the-art event generation down to the detector level, we have been able to prove the superiority of our (conditional) MLP and GNN approaches with respect to more traditional ones, wherein either a cut-and-count selection is solely exploited or else this is used in combination with more trivial ML informed methods. In particular, assuming foreseen energy and luminosity of the HL-LHC, we have proven that one can establish sensitivity to all CPV phase values between  $-\frac{\pi}{2}$  and  $\frac{\pi}{2}$ , including distinguishing between CP-even and -odd components in the signal sample. Furthermore, we have also tensioned the MLP against the GNN implementation and found that the latter exhibits better performance than the former. Finally, notice that, in order to enable validation of the results obtained here, our code and data are available on GitHub at [https://github.com/AHamamd150/Conditional\\_GNN](https://github.com/AHamamd150/Conditional_GNN).

## Acknowledgments

AH is funded by the Grant Number 22H05113 from the ‘‘Foundation of Machine Learning Physics’’, the ‘‘Grant in Aid for Transformative Research Areas’’ 22K03626 and the Grant-in-Aid for Scientific Research (C). The work of AJ is supported by the Institute for Basic Science (IBS) under Project Code IBS-R018-D1. SM is supported in part through the NExT Institute and the STFC Consolidated Grant ST/L000296/1. WE is funded by the ErUM-WAVE project 05D2022 ‘‘ErUM-Wave: Antizipation 3-dimensionaler Wellenfelder’’, which is supported by the German Federal Ministry of Education and Research (BMBF).

## A Verify the network performance with a toy example

Following the methodology outlined in [116] for conditional DNN, we validate our network architecture by replicating the illustrative toy example presented in the aforementioned paper. This example involves a simplified scenario featuring a single feature  $x$  and a corresponding conditional parameter  $\theta$ . The input features are Gaussian distributions with mean values equal to  $\theta$  and a standard deviation of  $\sigma = 2.5$ . Specifically, we examine signal points corresponding to  $\theta = -2, -1, 0, 1, 2$ , while the background points follow a uniform distribution, as depicted in Figure 10 (left).

In preparing the training dataset, we generate 500,000 signal points by stacking equally sized features for each  $\theta$  value. Since the primary objective of the network is to learn the distinguishing features of the signal events and interpolate between them, we incorporate smaller number of background points specifically 150,000 points, but with equal size of

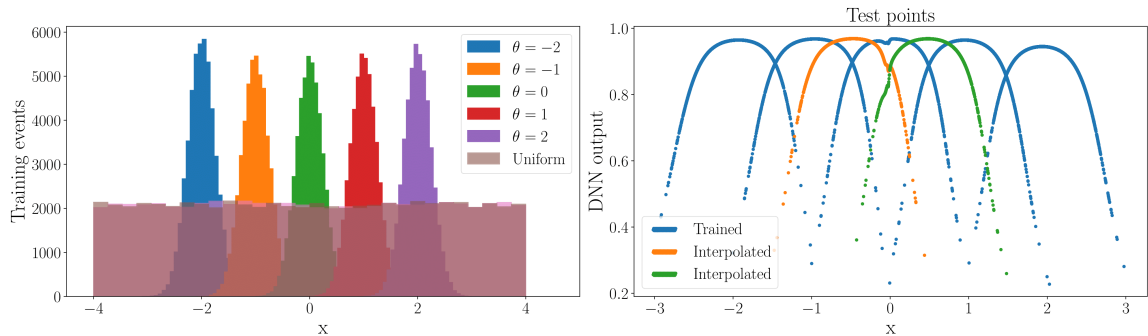


Figure 10: Left: training distributions in which the signal are represented by Gaussian and the background represented by a uniform distribution. Right: the network output as a function of the value of the input feature  $x$ , for various choices of the input parameter  $\theta$ . Orange and green distributions represent the points that the network has not seen during the training with  $\theta = -0.5, 0.5$ , respectively.

signal and background for test. For the conditional input vector,  $\theta$ , we concatenate the five values corresponding to each  $x$ , resulting in a single vector of length 500,000. For background events, we create a vector of random variables ranging between  $-2$  and  $2$ . The labels assigned to signal points are  $Y = 1$ , while background points are labeled as  $Y = 0$ .

For this example, we consider a conditional MLP that has two input layers, one for the feature  $x$  and one for the conditional parameter  $\theta$ . Input layer of the feature  $x$  is followed by three FC layers with number of neurons 300, 300, 100 and ReLU activation function; while the second input layer is followed by a linear layer with 100 neurons and no activation function. The two layers with 100 neurons are concatenated and passed to an output layer with one neuron and sigmoid activation function. The model is trained with 5 epochs and batch of size 500 points. We use a mean squared error function and Adam optimizer with learning rate of  $10^{-4}$ .

The network attains a training accuracy of 95.8% and a test accuracy of 95.7%. The network output is shown in Figure 10 (right plot) in blue. To evaluate its performance further, we introduce new points with  $\theta$  values not included in the training set, specifically  $-0.5$  and  $0.5$ , represented by the orange and green distributions, respectively. Remarkably, the network demonstrates its capability to interpolate to these novel  $\theta$  values, despite not being explicitly trained on them.

## References

- [1] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the  $2\pi$  Decay of the  $K_2^0$  Meson*, *Phys. Rev. Lett.* **13** (1964) 138–140.
- [2] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973) 652–657.
- [3] A. G. Cohen, D. B. Kaplan, and A. E. Nelson, *Spontaneous baryogenesis at the weak phase transition*, *Phys. Lett. B* **263** (1991) 86–92.
- [4] A. G. Cohen, D. B. Kaplan, and A. E. Nelson, *Progress in electroweak baryogenesis*, *Ann. Rev. Nucl. Part. Sci.* **43** (1993) 27–70, [[hep-ph/9302210](#)].

- [5] D. E. Morrissey and M. J. Ramsey-Musolf, *Electroweak baryogenesis*, *New J. Phys.* **14** (2012) 125003, [[arXiv:1206.2942](#)].
- [6] **ACME** Collaboration, V. Andreev et al., *Improved limit on the electric dipole moment of the electron*, *Nature* **562** (2018), no. 7727 355–360.
- [7] W. B. Cairncross, D. N. Gresh, M. Grau, K. C. Cossel, T. S. Roussy, Y. Ni, Y. Zhou, J. Ye, and E. A. Cornell, *Precision Measurement of the Electron’s Electric Dipole Moment Using Trapped Molecular Ions*, *Phys. Rev. Lett.* **119** (2017), no. 15 153001, [[arXiv:1704.07928](#)].
- [8] **nEDM** Collaboration, C. Abel et al., *Measurement of the permanent electric dipole moment of the neutron*, *Phys. Rev. Lett.* **124** (2020) 081803, [[arXiv:2001.11966](#)].
- [9] J. Engel, M. J. Ramsey-Musolf, and U. van Kolck, *Electric Dipole Moments of Nucleons, Nuclei, and Atoms: The Standard Model and Beyond*, *Prog. Part. Nucl. Phys.* **71** (2013) 21–74, [[arXiv:1303.2371](#)].
- [10] N. Yamanaka, B. K. Sahoo, N. Yoshinaga, T. Sato, K. Asahi, and B. P. Das, *Probing exotic phenomena at the interface of nuclear and particle physics with the electric dipole moments of diamagnetic atoms: A unique window to hadronic and semi-leptonic CP violation*, *Eur. Phys. J. A* **53** (2017), no. 3 54, [[arXiv:1703.01570](#)].
- [11] M. S. Safronova, D. Budker, D. DeMille, D. F. J. Kimball, A. Derevianko, and C. W. Clark, *Search for New Physics with Atoms and Molecules*, *Rev. Mod. Phys.* **90** (2018), no. 2 025008, [[arXiv:1710.01833](#)].
- [12] T. Chupp, P. Fierlinger, M. Ramsey-Musolf, and J. Singh, *Electric dipole moments of atoms, molecules, nuclei, and particles*, *Rev. Mod. Phys.* **91** (2019), no. 1 015001, [[arXiv:1710.02504](#)].
- [13] M. Pospelov and A. Ritz, *Electric dipole moments as probes of new physics*, *Annals Phys.* **318** (2005) 119–169, [[hep-ph/0504231](#)].
- [14] Y. Yamaguchi and N. Yamanaka, *Large long-distance contributions to the electric dipole moments of charged leptons in the standard model*, *Phys. Rev. Lett.* **125** (2020) 241802, [[arXiv:2003.08195](#)].
- [15] Y. Yamaguchi and N. Yamanaka, *Quark level and hadronic contributions to the electric dipole moment of charged leptons in the standard model*, *Phys. Rev. D* **103** (2021), no. 1 013001, [[arXiv:2006.00281](#)].
- [16] L. Bento, G. C. Branco, and P. A. Parada, *A Minimal model with natural suppression of strong CP violation*, *Phys. Lett. B* **267** (1991) 95–99.
- [17] T. D. Lee, *A Theory of Spontaneous T Violation*, *Phys. Rev.* **D8** (1973) 1226–1239. [,516(1973)].
- [18] T. D. Lee, *CP Nonconservation and Spontaneous Symmetry Breaking*, *Phys. Rept.* **9** (1974) 143–177. [,124(1974)].
- [19] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, *Theory and phenomenology of two-Higgs-doublet models*, *Phys. Rept.* **516** (2012) 1–102, [[arXiv:1106.0034](#)].



- [20] S. Weinberg, *Gauge Theory of CP Violation*, *Phys. Rev. Lett.* **37** (1976) 657.
- [21] A. W. El Kaffas, W. Khater, O. M. OGREID, and P. OSLAND, *Consistency of the two Higgs doublet model and CP violation in top production at the LHC*, *Nucl. Phys. B* **775** (2007) 45–77, [[hep-ph/0605142](#)].
- [22] 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](#)].
- [23] J. Shu and Y. Zhang, *Impact of a CP Violating Higgs Sector: From LHC to Baryogenesis*, *Phys. Rev. Lett.* **111** (2013), no. 9 091801, [[arXiv:1304.0773](#)].
- [24] Y.-n. Mao and S.-h. Zhu, *Lightness of Higgs boson and spontaneous CP violation in the Lee model*, *Phys. Rev. D* **90** (2014), no. 11 115024, [[arXiv:1409.6844](#)].
- [25] 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](#)].
- [26] V. Keus, S. F. King, S. Moretti, and K. Yagyu, *CP Violating Two-Higgs-Doublet Model: Constraints and LHC Predictions*, *JHEP* **04** (2016) 048, [[arXiv:1510.04028](#)].
- [27] 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](#)].
- [28] L. Bian and N. Chen, *Higgs pair productions in the CP-violating two-Higgs-doublet model*, *JHEP* **09** (2016) 069, [[arXiv:1607.02703](#)].
- [29] Y.-n. Mao and S.-h. Zhu, *Lightness of a Higgs Boson and Spontaneous CP-violation in the Lee Model: An Alternative Scenario*, *Phys. Rev. D* **94** (2016), no. 5 055008, [[arXiv:1602.00209](#)]. [Addendum: *Phys. Rev. D* **94**, no. 5, 059904 (2016)].
- [30] K. Hagiwara, K. Ma, and S. Mori, *Probing CP violation in  $h \rightarrow \tau^- \tau^+$  at the LHC*, *Phys. Rev. Lett.* **118** (2017), no. 17 171802, [[arXiv:1609.00943](#)].
- [31] C.-Y. Chen, H.-L. Li, and M. Ramsey-Musolf, *CP-Violation in the Two Higgs Doublet Model: from the LHC to EDMs*, *Phys. Rev. D* **97** (2018), no. 1 015020, [[arXiv:1708.00435](#)].
- [32] Q.-H. Cao, K.-P. Xie, H. Zhang, and R. Zhang, *A New Observable for Measuring CP Property of Top-Higgs Interaction*, *Chin. Phys. C* **45** (2021), no. 2 023117, [[arXiv:2008.13442](#)].
- [33] D. Azevedo, R. Capucha, E. Gouveia, A. Onofre, and R. Santos, *Light Higgs searches in  $t\bar{t}\phi$  production at the LHC*, *JHEP* **04** (2021) 077, [[arXiv:2012.10730](#)].
- [34] D. Azevedo, R. Capucha, A. Onofre, and R. Santos, *Scalar mass dependence of angular variables in  $t\bar{t}\phi$  production*, *JHEP* **06** (2020) 155, [[arXiv:2003.09043](#)].
- [35] S. Antusch, O. Fischer, A. Hammad, and C. Scherb, *Testing CP Properties of Extra Higgs States at the HL-LHC*, *JHEP* **03** (2021) 200, [[arXiv:2011.10388](#)].
- [36] K. Cheung, A. Jueid, Y.-n. Mao, and S. Moretti, *Two-Higgs-doublet model with soft CP violation confronting electric dipole moments and colliders*, *Phys. Rev. D* **102** (2020), no. 7 075029, [[arXiv:2003.04178](#)].

- [37] S. Kanemura, M. Kubota, and K. Yagyu, *Testing aligned CP-violating Higgs sector at future lepton colliders*, *JHEP* **04** (2021) 144, [[arXiv:2101.03702](#)].
- [38] I. Low, N. R. Shah, and X.-P. Wang, *Higgs alignment and novel CP-violating observables in two-Higgs-doublet models*, *Phys. Rev. D* **105** (2022), no. 3 035009, [[arXiv:2012.00773](#)].
- [39] **ATLAS** Collaboration, G. Aad et al., *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) 1–29, [[arXiv:1207.7214](#)].
- [40] **CMS** Collaboration, S. Chatrchyan et al., *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30–61, [[arXiv:1207.7235](#)].
- [41] **ATLAS, CMS** Collaboration, G. Aad et al., *Combined Measurement of the Higgs Boson Mass in pp Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Experiments*, *Phys. Rev. Lett.* **114** (2015) 191803, [[arXiv:1503.07589](#)].
- [42] C. R. Schmidt and M. E. Peskin, *A Probe of CP violation in top quark pair production at hadron supercolliders*, *Phys. Rev. Lett.* **69** (1992) 410–413.
- [43] G. Mahlon and S. J. Parke, *Angular correlations in top quark pair production and decay at hadron colliders*, *Phys. Rev.* **D53** (1996) 4886–4896, [[hep-ph/9512264](#)].
- [44] E. Asakawa and K. Hagiwara, *Probing the CP nature of the Higgs bosons by  $t$  anti- $t$  production at photon linear colliders*, *Eur. Phys. J. C* **31** (2003) 351–364, [[hep-ph/0305323](#)].
- [45] P. S. Bhupal Dev, A. Djouadi, R. M. Godbole, M. M. Muhlleitner, and S. D. Rindani, *Determining the CP properties of the Higgs boson*, *Phys. Rev. Lett.* **100** (2008) 051801, [[arXiv:0707.2878](#)].
- [46] X.-G. He, G.-N. Li, and Y.-J. Zheng, *Probing Higgs boson CP Properties with  $t\bar{t}H$  at the LHC and the 100 TeV pp collider*, *Int. J. Mod. Phys. A* **30** (2015), no. 25 1550156, [[arXiv:1501.00012](#)].
- [47] F. Boudjema, R. M. Godbole, D. Guadagnoli, and K. A. Mohan, *Lab-frame observables for probing the top-Higgs interaction*, *Phys. Rev.* **D92** (2015), no. 1 015019, [[arXiv:1501.03157](#)].
- [48] N. Mileo, K. Kiers, A. Szykman, 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](#)].
- [49] M. R. Buckley and D. Goncalves, *Boosting the Direct CP Measurement of the Higgs-Top Coupling*, *Phys. Rev. Lett.* **116** (2016), no. 9 091801, [[arXiv:1507.07926](#)].
- [50] 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.* **D96** (2017), no. 1 013004, [[arXiv:1704.03565](#)].
- [51] D. Azevedo, A. Onofre, F. Filthaut, and R. Goncalo, *CP tests of Higgs couplings in  $t\bar{t}h$  semileptonic events at the LHC*, *Phys. Rev. D* **98** (2018), no. 3 033004, [[arXiv:1711.05292](#)].



- [52] W. Bernreuther, L. Chen, I. García, M. Perelló, R. Poeschl, F. Richard, E. Ros, and M. Vos, *CP-violating top quark couplings at future linear  $e^+e^-$  colliders*, *Eur. Phys. J. C* **78** (2018), no. 2 155, [[arXiv:1710.06737](#)].
- [53] K. Hagiwara, H. Yokoya, and Y.-J. Zheng, *Probing the CP properties of top Yukawa coupling at an  $e^+e^-$  collider*, *JHEP* **02** (2018) 180, [[arXiv:1712.09953](#)].
- [54] K. Ma, *Enhancing CP Measurement of the Yukawa Interactions of Top-Quark at  $e^-e^+$  Collider*, *Phys. Lett. B* **797** (2019) 134928, [[arXiv:1809.07127](#)].
- [55] M. Cepeda et al., *Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC*, *CERN Yellow Rep. Monogr.* **7** (2019) 221–584, [[arXiv:1902.00134](#)].
- [56] D. A. Faroughy, J. F. Kamenik, N. Košnik, and A. Smolkovič, *Probing the CP nature of the top quark Yukawa at hadron colliders*, *JHEP* **02** (2020) 085, [[arXiv:1909.00007](#)].
- [57] H. Bahl, P. Bechtle, S. Heinemeyer, J. Katzy, T. Klingl, K. Peters, M. Saimpert, T. Stefaniak, and G. Weiglein, *Indirect CP probes of the Higgs-top-quark interaction: current LHC constraints and future opportunities*, *JHEP* **11** (2020) 127, [[arXiv:2007.08542](#)].
- [58] H. Bahl and S. Brass, *Constraining CP-violation in the Higgs-top-quark interaction using machine-learning-based inference*, *JHEP* **03** (2022) 017, [[arXiv:2110.10177](#)].
- [59] R. K. Barman, D. Gonçalves, and F. Kling, *Machine learning the Higgs boson-top quark CP phase*, *Phys. Rev. D* **105** (2022), no. 3 035023, [[arXiv:2110.07635](#)].
- [60] H. Bahl, E. Fuchs, S. Heinemeyer, J. Katzy, M. Menen, K. Peters, M. Saimpert, and G. Weiglein, *Constraining the CP structure of Higgs-fermion couplings with a global LHC fit, the electron EDM and baryogenesis*, *Eur. Phys. J. C* **82** (2022), no. 7 604, [[arXiv:2202.11753](#)].
- [61] H. Bahl, E. Fuchs, M. Hannig, and M. Menen, *Classifying the CP properties of the  $ggH$  coupling in  $H+2j$  production*, [arXiv:2309.03146](#).
- [62] F. Gianotti et al., *Physics potential and experimental challenges of the LHC luminosity upgrade*, *Eur. Phys. J. C* **39** (2005) 293–333, [[hep-ph/0204087](#)].
- [63] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, *JHEP* **06** (2011) 128, [[arXiv:1106.0522](#)].
- [64] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *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](#)].
- [65] **NNPDF** Collaboration, R. D. Ball et al., *The path to proton structure at 1% accuracy*, *Eur. Phys. J. C* **82** (2022), no. 5 428, [[arXiv:2109.02653](#)].
- [66] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rufenacht, M. Schönherr, and G. Watt, *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J. C* **75** (2015) 132, [[arXiv:1412.7420](#)].

- [67] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau, and P. Torrielli, *Scalar and pseudoscalar Higgs production in association with a top–antitop pair*, *Phys. Lett. B* **701** (2011) 427–433, [[arXiv:1104.5613](#)].
- [68] F. Maltoni, E. Vryonidou, and C. Zhang, *Higgs production in association with a top-antitop pair in the Standard Model Effective Field Theory at NLO in QCD*, *JHEP* **10** (2016) 123, [[arXiv:1607.05330](#)].
- [69] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, *Phys. Rept.* **457** (2008) 1–216, [[hep-ph/0503172](#)].
- [70] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model*, *Phys. Rept.* **459** (2008) 1–241, [[hep-ph/0503173](#)].
- [71] K. G. Chetyrkin, *Correlator of the quark scalar currents and  $\Gamma(\text{tot}) (H \rightarrow \text{hadrons})$  at  $O(\alpha_s^3)$  in pQCD*, *Phys. Lett.* **B390** (1997) 309–317, [[hep-ph/9608318](#)].
- [72] K. G. Chetyrkin and M. Steinhauser, *Complete QCD corrections of order  $O(\alpha_s^3)$  to the hadronic Higgs decay*, *Phys. Lett.* **B408** (1997) 320–324, [[hep-ph/9706462](#)].
- [73] A. J. Barr and C. G. Lester, *A Review of the Mass Measurement Techniques proposed for the Large Hadron Collider*, *J. Phys. G* **37** (2010) 123001, [[arXiv:1004.2732](#)].
- [74] C. G. Lester and D. J. Summers, *Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders*, *Phys. Lett. B* **463** (1999) 99–103, [[hep-ph/9906349](#)].
- [75] H.-C. Cheng and Z. Han, *Minimal Kinematic Constraints and  $m(T_2)$* , *JHEP* **12** (2008) 063, [[arXiv:0810.5178](#)].
- [76] W. S. Cho, K. Choi, Y. G. Kim, and C. B. Park,  *$M(T_2)$ -assisted on-shell reconstruction of missing momenta and its application to spin measurement at the LHC*, *Phys. Rev. D* **79** (2009) 031701, [[arXiv:0810.4853](#)].
- [77] D. Gonçalves, K. Kong, and J. H. Kim, *Probing the top-Higgs Yukawa CP structure in dileptonic  $t\bar{t}h$  with  $M_2$ -assisted reconstruction*, *JHEP* **06** (2018) 079, [[arXiv:1804.05874](#)].
- [78] L. Sonnenschein, *Algebraic approach to solve  $t\bar{t}$  dilepton equations*, *Phys. Rev. D* **72** (2005) 095020, [[hep-ph/0510100](#)].
- [79] D. A. Millar, *Phenomenology of searches for new neutral resonances in top quark pair production at the LHC*. PhD thesis, Southampton U., 2018.
- [80] W. Bernreuther and Z.-G. Si, *Top quark spin correlations and polarization at the LHC: standard model predictions and effects of anomalous top chromo moments*, *Phys. Lett.* **B725** (2013) 115–122, [[arXiv:1305.2066](#)]. [Erratum: *Phys. Lett.* **B744**,413(2015)].
- [81] J. Shelton, *Polarized tops from new physics: signals and observables*, *Phys. Rev.* **D79** (2009) 014032, [[arXiv:0811.0569](#)].

- [82] R. M. Godbole, L. Hartgring, I. Niessen, and C. D. White, *Top polarisation studies in  $H^-t$  and  $Wt$  production*, *JHEP* **01** (2012) 011, [[arXiv:1111.0759](#)].
- [83] S. D. Rindani and P. Sharma, *Probing anomalous  $tbW$  couplings in single-top production using top polarization at the Large Hadron Collider*, *JHEP* **11** (2011) 082, [[arXiv:1107.2597](#)].
- [84] A. Prasath V, R. M. Godbole, and S. D. Rindani, *Longitudinal top polarisation measurement and anomalous  $Wtb$  coupling*, *Eur. Phys. J.* **C75** (2015), no. 9 402, [[arXiv:1405.1264](#)].
- [85] R. M. Godbole, G. Mendiratta, and S. Rindani, *Looking for bSM physics using top-quark polarization and decay-lepton kinematic asymmetries*, *Phys. Rev.* **D92** (2015), no. 9 094013, [[arXiv:1506.07486](#)].
- [86] A. Jueid, *Probing anomalous  $Wtb$  couplings at the LHC in single  $t$ -channel top quark production*, *Phys. Rev. D* **98** (2018), no. 5 053006, [[arXiv:1805.07763](#)].
- [87] A. Arhrib, A. Jueid, and S. Moretti, *Top quark polarization as a probe of charged Higgs bosons*, *Phys. Rev. D* **98** (2018), no. 11 115006, [[arXiv:1807.11306](#)].
- [88] R. Godbole, M. Guchait, C. K. Khosa, J. Lahiri, S. Sharma, and A. H. Vijay, *Boosted Top quark polarization*, *Phys. Rev.* **D100** (2019), no. 5 056010, [[arXiv:1902.08096](#)].
- [89] A. Arhrib, A. Jueid, and S. Moretti, *Searching for Heavy Charged Higgs Bosons through Top Quark Polarization*, *Int. J. Mod. Phys. A* **35** (2020), no. 15n16 2041011, [[arXiv:1903.11489](#)].
- [90] S. Chatterjee, R. Godbole, and T. S. Roy, *Jets with electrons from boosted top quarks*, *JHEP* **01** (2020) 170, [[arXiv:1909.11041](#)].
- [91] **ATLAS** Collaboration, M. Aaboud et al., *Measurements of top quark spin observables in  $t\bar{t}$  events using dilepton final states in  $\sqrt{s} = 8$  TeV  $pp$  collisions with the ATLAS detector*, *JHEP* **03** (2017) 113, [[arXiv:1612.07004](#)].
- [92] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the top quark polarization and  $t\bar{t}$  spin correlations using dilepton final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. D* **100** (2019), no. 7 072002, [[arXiv:1907.03729](#)].
- [93] **TMVA** Collaboration, A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, [physics/0703039](#).
- [94] T. N. Kipf and M. Welling, *Semi-supervised classification with graph convolutional networks*, in *International Conference on Learning Representations*, 2017.
- [95] W. Esmail, A. Hammad, and S. Moretti, *Sharpening the  $A \rightarrow Z^{(*)}h$  signature of the Type-II 2HDM at the LHC through advanced Machine Learning*, *JHEP* **11** (2023) 020, [[arXiv:2305.13781](#)].
- [96] M. Fey and J. E. Lenssen, *Fast graph representation learning with pytorch geometric*, *arXiv preprint arXiv:1903.02428* (2019).

- [97] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, *Comput. Phys. Commun.* **185** (2014) 2250–2300, [[arXiv:1310.1921](#)].
- [98] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter, *UFO - The Universal FeynRules Output*, *Comput. Phys. Commun.* **183** (2012) 1201–1214, [[arXiv:1108.2040](#)].
- [99] 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](#)].
- [100] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8, [[arXiv:2203.11601](#)].
- [101] J. Y. Araz, B. Fuks, and G. Polykratis, *Simplified fast detector simulation in MADANALYSIS 5*, *Eur. Phys. J. C* **81** (2021), no. 4 329, [[arXiv:2006.09387](#)].
- [102] E. Conte, B. Fuks, and G. Serret, *MadAnalysis 5, A User-Friendly Framework for Collider Phenomenology*, *Comput. Phys. Commun.* **184** (2013) 222–256, [[arXiv:1206.1599](#)].
- [103] E. Conte, B. Dumont, B. Fuks, and C. Wymant, *Designing and recasting LHC analyses with MadAnalysis 5*, *Eur. Phys. J. C* **74** (2014), no. 10 3103, [[arXiv:1405.3982](#)].
- [104] E. Conte and B. Fuks, *Confronting new physics theories to LHC data with MADANALYSIS 5*, *Int. J. Mod. Phys. A* **33** (2018), no. 28 1830027, [[arXiv:1808.00480](#)].
- [105] J. Y. Araz, M. Frank, and B. Fuks, *Reinterpreting the results of the LHC with MadAnalysis 5: uncertainties and higher-luminosity estimates*, *Eur. Phys. J. C* **80** (2020), no. 6 531, [[arXiv:1910.11418](#)].
- [106] M. Frank, B. Fuks, A. Jueid, S. Moretti, and O. Ozdal, *A novel search strategy for right-handed charged gauge bosons at the Large Hadron Collider*, *JHEP* **02** (2024) 150, [[arXiv:2312.08521](#)].
- [107] CMS Collaboration, A. Tumasyan et al., *Search for a right-handed W boson and a heavy neutrino in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **04** (2022) 047, [[arXiv:2112.03949](#)].
- [108] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063, [[arXiv:0802.1189](#)].
- [109] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896, [[arXiv:1111.6097](#)].
- [110] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., *Pytorch: An imperative style, high-performance deep learning library*, *Advances in neural information processing systems* **32** (2019).
- [111] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., *Scikit-learn: Machine learning in python, the Journal of machine Learning research* **12** (2011) 2825–2830.

- [112] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554, [[arXiv:1007.1727](#)]. [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)].
- [113] **LHC Dark Matter Working Group** Collaboration, T. Abe et al., *LHC Dark Matter Working Group: Next-generation spin-0 dark matter models*, *Phys. Dark Univ.* **27** (2020) 100351, [[arXiv:1810.09420](#)].
- [114] S. Antusch, E. Cazzato, O. Fischer, A. Hammad, and K. Wang, *Lepton Flavor Violating Dilepton Dijet Signatures from Sterile Neutrinos at Proton Colliders*, *JHEP* **10** (2018) 067, [[arXiv:1805.11400](#)].
- [115] S. Antusch, A. Hammad, and A. Rashed, *Probing  $Z'$  mediated charged lepton flavor violation with taus at the LHeC*, *Phys. Lett. B* **810** (2020) 135796, [[arXiv:2003.11091](#)].
- [116] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, *Parameterized neural networks for high-energy physics*, *Eur. Phys. J. C* **76** (2016), no. 5 235, [[arXiv:1601.07913](#)].