

# Polarization measurement for the dileptonic channel of $W^+W^-$ scattering using generative adversarial network

Jinmian Li<sup>✉\*</sup>, Cong Zhang,<sup>†</sup> and Rao Zhang<sup>✉‡</sup>

College of Physics, Sichuan University, Chengdu 610065, China

 (Received 27 September 2021; accepted 7 December 2021; published 5 January 2022)

Measuring the polarization fractions of the  $W^+W^-$  scattering reveals the interactions of the Higgs boson as well as new neutral states that are related to the standard model electroweak symmetry breaking. The dileptonic channel has a relatively lower background rate, but the kinematics of its final states cannot be fully reconstructed due to the presence of two neutrinos. We propose neural networks to establish maps between the distributions of measurable quantities and the distributions of the lepton angles in  $W$  boson rest frames. New physics contributions and collision energy can largely affect the kinematic properties of the  $W^+W^-$  scattering beside the lepton angles. To make the network applicable to processes with varying kinematics, the loss function is modified in two different ways. We show that the networks are promising in reproducing the lepton angle distributions, and the precision of the fitted polarization fractions obtained from network predictions is comparable to that obtained with the truth lepton angle. Although the best-fit values of the polarization fractions do not change much after including the background uncertainty, the precision is substantially reduced.

DOI: [10.1103/PhysRevD.105.016005](https://doi.org/10.1103/PhysRevD.105.016005)

## I. INTRODUCTION

Vector boson scattering (VBS) [1–4] represents a sensitive probe to any new physics that is interacting with the electroweak sector of the Standard Model (SM). If the Higgs sector is extended or the couplings between the Higgs boson and the gauge bosons deviate from the SM predictions, the scattering amplitudes for the longitudinal mode of the VBS will increase with center-of-mass energy and violate unitarity.

At hadron colliders, the VBS processes result in final states with two gauge bosons and a pair of forward-backward jets. VBS channels have been observed at the LHC Run 2, including the dileptonic same-sign  $W^\pm W^\pm$  [5,6], fully leptonic  $ZZ$  [7,8], fully leptonic  $WZ$  [9,10], and semileptonic  $WV/ZV$  with the  $V$  decaying hadronically [11,12]. Investigating the polarization modes of the VBS processes is an important step afterward. The polarization of the vector bosons can be measured by their decay products. The interferences among different polarization channels disappear when the azimuthal angles of the decay

products are integrated over. Although the selection cuts in analyses induce a certain amount of interference, it is still possible to extract polarization fractions by fitting the data with simulated templates. There have been many studies of polarization measurement for the  $W^+W^-$  channel [13,14], the fully leptonic  $W^\pm W^\pm$  channel [15], the fully leptonic  $WZ/ZZ$  channel [16], as well as the  $WW/ZZ$  from the SM Higgs decay [17] and generic processes with a boosted hadronically decaying  $W$  boson [18]. The CMS Collaboration studied the prospects for measuring the longitudinal modes of  $W^\pm W^\pm$  and  $WZ$  channels at the future HL-LHC [19,20]. Some recent studies take advantage of deep learning techniques. Taking the final-state momenta as input, the network is able to either regress the lepton angle in the gauge boson rest frame [21,22] or classify events from different polarizations [23,24].

In this work, we study the polarization measurement for the dileptonic  $W^+W^-$  channel, as it has a large production cross section at the LHC and is relevant to the neutral scalar bosons. Although resolving the polarization of the hadronically decaying  $W$  boson is possible [18], it suffers from large uncertainties and backgrounds. Focusing on the  $WW$  scattering with fully leptonic decay in the SM, the fractions of the  $W$  boson polarization can be determined from distributions of many kinematic variables [15,25], e.g., the transverse momenta of leptons, the invariant mass of two leptons, and so on. Studies in Refs. [23,24] use the neural network to discriminate different polarization modes of VBS processes, and the output of the network can be used to extract the fraction of each polarization mode. However, those methods

\*jqli@scu.edu.cn  
<sup>†</sup>zhangcong.phy@gmail.com  
<sup>‡</sup>zhangrao@stu.scu.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.

utilize information that also depends on other properties of the process besides the polarization. Thus, they cannot be applied directly to VBS processes with significant beyond the SM (BSM) contributions. The most exclusive variable that characterizes the vector boson polarization is the angle between the charged lepton in the gauge boson rest frame and the gauge boson direction of motion (denoted by  $\theta_\ell^*$  hereafter). Because of the presence of two neutrinos in the final state, the lepton angles cannot be fully reconstructed. There have been attempts to use a neural network to regress two lepton angles in the gauge boson rest frames [21,22] for the same-sign  $W^\pm W^\pm$  scattering.

We further develop the machine learning methods by adopting the transformer network [26] and the generative adversarial network. The transformer network is known to be quite successful in extracting features of polarizations for VBS processes [27]. The generative adversarial network is used to regress the distribution of lepton angle  $\theta_\ell^*$ . As a result, the network can be used to measure the polarization fraction of the  $W^+W^-$  scattering in a wide class of models. For illustration, we apply the network to a simplified model with an effective operator and the two-Higgs-doublet model (2HDM). In particular, there is an extra neutral Higgs boson in the 2HDM which induces resonant  $W^+W^-jj$  production, so the kinematic properties of the  $W^+W^-jj$  final state in the 2HDM are quite different from the SM ones. We show that our network works well in regressing the lepton angles for both BSM scenarios. The polarization fraction can be extracted from fitting the distribution of the predicted lepton angle to a linear combination of pure longitudinal/transverse templates. However, to reduce the background events, some preselection cuts need to be applied before constructing the templates. Those cuts will affect the shapes of the templates. We find that the shapes of the templates (according to our preselection) are similar in different models but exhibit some dependences on collision energy. This means different sets of templates should be reconstructed at different collision energy.

This paper is organized as follows. The analysis framework is explained in Sec. II, including the setup of the network, event preparation, and fitting procedure. In Secs. III and IV, we study the performance of the network applied on different models and different collision energies. The effects of backgrounds are discussed in Sec. V. We summarize our work and conclude in Sec. VI.

## II. ANALYSIS FRAMEWORK

### A. Definitions of loss functions and the network

The following issues need to be addressed in network construction:

- (i) Because of missing information for two neutrinos in the final state, it is not possible to fully determine the two lepton angles in the rest frames of the  $W$  bosons for each event. The lepton decay angles of events with the

same values of observables (including the momenta of leptons and jets, as well as the missing transverse momentum) form a distribution. Our network is built to establish a map between the distributions of measurable quantities and the distributions of the lepton angles based on a large number of events.

- (ii) Since we expect that the VBS process is affected by unknown new physics, the network should be able to extract the  $W$  boson polarization for processes that have kinematic properties quite different from the SM. This means that the features extracted by the network should be only related to  $\theta_\ell^*$  and decorrelated from other process-dependent variables.
- (iii) Extracting the polarization fraction requires fitting to the given templates. The shapes of the templates are affected by preselection cuts; thus, they will be different at different collision energies of the hadron collider. On the other hand, the network used to extract polarization information needs to provide features that do not change with the collision energy.

Because of the first issue, we cannot use the mean-squared-error loss function, which can only reproduce the average value of the lepton angle distribution (for given values of observables) and lead to the deviation between the truth-level distribution and the predicted distribution [21,22]. Events with the same measurable momenta of final states while having different  $\theta_{\ell^\pm}^*$  are grouped into subsets denoted by  $e^i$ . The measurable momenta for the subset  $e^i$  are denoted by  $\mathbf{p}^i$  (same for all events in  $e^i$ ), and the set of  $\theta_{\ell^\pm}^*$  for events in  $e^i$  is denoted by  $t^i$ . The goal of the network is to establish a map that maximizes the probability of  $P(t^i|\mathbf{p}^i)$  while minimizing the probability  $P(t^j|\mathbf{p}^i)$  for  $j \neq i$ , where  $i$  and  $j$  run over all subsets. The loss function of the conditional generative adversarial network (CGAN) [28] meets the needs

$$L_{\text{CGAN}} = \min_G \max_D (\mathbb{E}[\log D(t^i|\mathbf{p}^i)] + \mathbb{E}_{z \in \mathcal{N}(0,1)}[\log(1 - D(G(z, \mathbf{p}^i)|\mathbf{p}^i))]), \quad (2.1)$$

where  $z$  is sampled from a Gaussian distribution and  $\mathbb{E}$  denotes the average over all events. The distribution of  $(\theta_{\ell^+}^*, \theta_{\ell^-}^*)$  in subset  $t^i$  is replaced by a two-dimensional Gaussian distribution  $t^i$  [which centers on the mean of  $(\theta_{\ell^+}^*, \theta_{\ell^-}^*)$  with standard derivative 0.01] for simplicity. The discriminative network ( $D$ ) evaluates the consistency between the  $\mathbf{p}^i$  and a lepton angle distribution. The generative network ( $G$ ) aims to reproduce the  $t^i$  distribution with the input of  $z$  and  $\mathbf{p}^i$ . The GAN enables us to obtain the lepton angle by sampling instead of taking the average, and it transforms the random distribution  $z$  into meaningful distributions based on the information obtained from the training samples.

To address the second problem, we adopt the mutual information (MI) variable to measure the nonlinear correlation between features and the target variables. For any two sets of variables  $X$  and  $Y$ , the MI is defined as

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} \mathbb{P}_{(x,y)} \log \left( \frac{\mathbb{P}_{(x,y)}}{\mathbb{P}_x \mathbb{P}_y} \right), \quad (2.2)$$

where  $\mathbb{P}_{(x,y)}$  is the joint probability density function, and  $\mathbb{P}_x$  and  $\mathbb{P}_y$  are the marginal probability density functions. The  $I(X; Y)$  is larger if  $X$  and  $Y$  share similar information, while  $I(X; Y) = 0$  if  $X$  and  $Y$  are independent of each other. However, MI is difficult to calculate in practice. We use the following approximation to estimate the MI [29] instead:

$$I(X; Y) = \sup_{\omega} [\mathbb{E}_{\mathbb{P}_{XY}} [T_{\omega}] - \log(\mathbb{E}_{\mathbb{P}_X \otimes \mathbb{P}_Y} [e^{T_{\omega}}])], \quad (2.3)$$

where  $T_{\omega}$  is an arbitrary function described by a neural network in which the weights  $\omega$  are trained to provide the least upper limit for the  $I(X; Y)$ . In our study, the loss function is written such that the MI between the two leptons angles  $\theta_{\ell^{\pm}}^*$  and the features (the transformer output, which has dimension 64, will be discussed later) is maximized, while the MI between the  $W$  boson pair momentum [including invariant mass  $m(WW)$ , energy  $E(WW)$ , rapidity  $y(WW)$ , and azimuth  $\phi(WW)$ ] and the features are minimized to reduce the dependence of network performance on the  $W$  boson pair production mechanism. So the  $L_{\text{MI}} = I(F; WW) - I(F; \Theta^*)$  is added to the category loss of the transformer network. The  $F$ ,  $WW$ , and  $\Theta^*$  denote the sets of feature variables,  $W$  boson pair momentum, and two leptons angles, respectively. The  $T_{\omega}$  networks in  $I(F; WW)$  and  $I(F; \Theta^*)$  are fully connected neural networks, which consist of four layers with  $[272 = 4 \times (64 + 4), 272, 272, 1]$  and  $[264 = 4 \times (64 + 2), 264, 264, 1]$  numbers of neurons, respectively, and the rectified linear unit (ReLU) function acts on all layers except the last layer of the two fully connected neural networks.

The transverse momenta of particles in the final state are approximately linearly related to the initial collision energy. Thus, we further add the Pearson correlation coefficient (PCC) to the loss function to reduce the dependences of the feature variables on collision energy. The PCC is defined as

$$\rho_{XY} = \left| \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}} \right|, \quad (2.4)$$

where  $i$  runs over all events, and  $\bar{X}$  and  $\bar{Y}$  denote the average of the variables. In our case, the features from the transformer output are taken as  $X$ , and the variable  $Y$  indicates the transverse momenta of the  $W$  bosons, leptons, as well as forward-backward jets.<sup>1</sup> Thus,  $64 \times 8$   $\rho_{XY}$ 's can be calculated. The average  $\bar{\rho}_{XY}$  is added to the loss of the transformer network.

<sup>1</sup>We also include the pseudorapidities of the  $W$  bosons, although they are not linearly correlated with collision energy.

Having defined the loss function, we can construct a network based on the transformer network and the CGAN. The transformer network [26] with a multihead self-attention mechanism provides a variety of different attentions and improves the learning ability; thus, it can be used to effectively extract the internal connections of the features. In Ref. [27], it is found to be efficient in extracting polarization information for the  $W^+W^-$  scattering in both semileptonic and dileptonic channels. We adopt the same transformer network as used for the dileptonic channel in this work. The low-level inputs (momenta of final-state particles) are transformed into a 64-dimensional feature variable which is supposed to contain the full polarization information. In this study, the loss function of the transformer network is modified according to the discussions above, using  $L_{\text{MI}}$  and  $\bar{\rho}_{XY}$ . The CGAN uses the feature as a condition and reproduces the two-dimensional lepton angle distribution. The generator takes the input of the condition (64 dimensions) and a 64-dimensional Gaussian distribution and aims to regenerate the lepton angle distribution. The discriminator takes the input of the condition and  $\Theta$  (or  $\Theta^*$ )<sup>2</sup> and determines whether the input  $\Theta$  (or  $\Theta^*$ ) is consistent with the condition.

More details of the data processing and architecture of the network are depicted in Fig. 1. The upper-left plot illustrates the processing pipeline of the modified transformer network. Note that the variables in sets of  $\Theta^*$ ,  $WW$ , and PCC, which can only be calculated on the Monte Carlo events, are only used for training. During the inference stage, only the inputs of the measurable momenta of final states are required. The lower plots show the architectures of the generative network and the discriminative network. In both networks, the condition is processed by a dense network with eight layers. The outputs of the dense networks are reused multiple times in the network as indicated by labels 1 and 2. The ResBlock is proposed to address the degradation problem [30] in training deep networks. We illustrate its decomposition in the upper-right plot. The StyleBlock combines the condition with its input by convolution operation (for more detail, see Ref. [31]). The parameters NF, SF, and S in one-dimensional convolution Conv1D[NF, SF, S] and StyleBlock[NF, SF, S] are the number of filters, filter size, and stride, respectively. The MergeBlock multiplies the condition with lepton angle information in matrix form. The LeakyReLU is an activation function, which is consistent with the ReLU for  $x \geq 0$  and equals  $0.1x$  for  $x < 0$ . The symbol  $\oplus$  means the sum of corresponding elements;  $\textcircled{C}$  means concatenating by channel; upsampling  $\times 2$  (downsampling/2) means using linear interpolation to expand (reduce) the dimension of input by twice (half).

<sup>2</sup>The predicted lepton angles'  $\theta_{\ell^{\pm}}^*$  distribution is denoted by  $\Theta$  and the truth lepton angles'  $\theta_{\ell^{\pm}}^*$  distribution is denoted by  $\Theta^*$ .

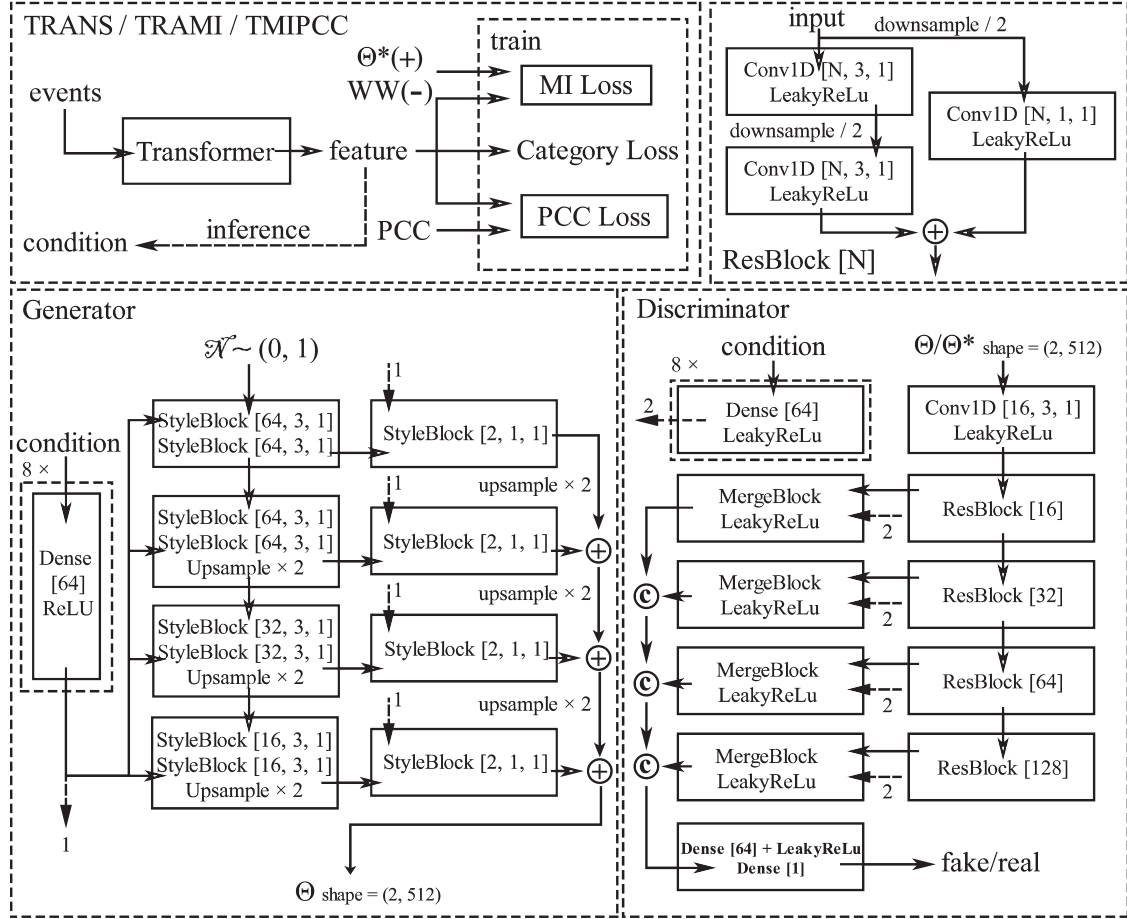


FIG. 1. Workflow of the network.

## B. Event simulation and network training

New physics models are implemented in FEYNRULES [32] (in our case, we consider the effective field theory [33] and the 2HDM). Events at the LHC are simulated within the MG5\_AMC@NLO framework [34], including those with fixed helicities of gauge bosons in the final state [25].<sup>3</sup> MADSPIN [35] is turned on to preserve the polarization information in the decay products of the gauge bosons. PYTHIA8 [36] is used for the parton shower, hadronization, and decays of hadrons. The final-state jets are reconstructed by FASTJET [37] using the anti- $k_T$  algorithm with cone size parameter  $R = 0.4$ . The detector effects are simulated by DELPHES3 [38] with the ATLAS configuration card, where  $b$ -tagging efficiency is set to 70%, and the mistagging rates for the charm- and light-flavor jets are 0.15 and 0.008, respectively [39].

The  $W^+W^-$  scattering is simulated at order of  $\mathcal{O}(\alpha_{EW}^4)$  in the SM model. There are also  $W^+W^-jj$  productions at

<sup>3</sup>The definitions of the polarizations are reference-frame dependent. We take the initial parton center of the mass frame as the reference frame in this work.

<sup>4</sup>The  $\alpha_{EW}$  and  $\alpha_s$  denote the electroweak coupling constant and strong coupling constant, respectively.

$\mathcal{O}(\alpha_{EW}^2\alpha_s^2)$  with much higher rates, but they do not correspond to VBS. They will be treated as the background for the  $W^+W^-$  scattering because the interference contributions at  $\mathcal{O}(\alpha_{EW}^3\alpha_s)$  are found to be small [14,40,41]. As for simulating the processes in BSM, the new physics coupling ( $\alpha_{NP}$ ) is assumed to be close to the electroweak coupling. The processes at the order of  $\mathcal{O}(\alpha_{EW}^a\alpha_{NP}^b)$  with  $a + b = 4$  are considered.

In order to separate signal and background events in the dileptonic channel, the following preselections are applied:

- (i) exactly two opposite-sign leptons with  $p_T(\ell) > 20$  GeV,  $|\eta(\ell)| < 2.5$ ;
- (ii) at least two jets with  $p_T(j) > 20$  GeV,  $|\eta(j)| < 4.5$ ;
- (iii) the two jets with leading  $p_T$  should give large invariant mass ( $m_{jj} > 500$  GeV) and have large pseudorapidity separation ( $|\Delta\eta|_{jj} > 3.6$ );
- (iv) no  $b$ -tagged jet in the final state.

The preselected events are used for training and testing the network. The network input consists of momenta ( $p_x, p_y, p_z, E$ ) of two leptons, forward and backward jets, the vectorial sum of all detected particles, and the vectorial sum of jets that are not assigned as forward-backward jets. The transformer network that is used to

extract the features of different polarizations is trained on the events of the SM  $W^+W^-$  scattering (with given final-state polarizations) at the 13 TeV LHC.<sup>5</sup> Moreover, as

discussed in the previous subsection, the following variables are calculated for each Monte Carlo event (used at the training stage):

$$\begin{aligned} &\text{MI variables, } \theta_{\ell^+}^*, \theta_{\ell^-}^*, m(WW), E(WW), y(WW), \phi(WW); \\ &\text{PCC variables, } p_T(W^\pm), p_T(\ell^\pm), p_T(j_{\text{fb}}), \eta(W^\pm). \end{aligned}$$

To show the performance gain of adding Eqs. (2.3) and (2.4) to the loss function, three versions of networks are trained:

- (i) network with normal transformer loss function, denoted by TRANS;
- (ii) network with  $L_{\text{MI}}$  being added to the loss function, denoted by TRAMI;
- (iii) network with both  $L_{\text{MI}}$  and  $\bar{\rho}_{XY}$  being added to the loss function, denoted by TMIPCC.

The CGAN is trained independently of the transformer (with modified loss). It takes the input condition provided by the well-trained TRANS network, TRAMI network, and TMIPCC network, respectively. Events of the  $W^+W^-$  scattering in both the SM and BSM at several collision energies are used for training the CGAN. The BSM scenarios include the effective field theory (EFT) and the 2HDM with several choices of benchmark parameters, as will be discussed later.

In Table I, we present the classification accuracies of the transformer networks and correlation information for the well-trained full networks (transformer + CGAN). The classification accuracy can reach 44% for the TRANS network and decrease a bit in TRAMI and TMIPCC. However, the great enhancement (reduction) of  $\text{MI}(F; \Theta^*)$  [ $\text{MI}(F; WW)$ ] in TRAMI indicates that the features have been changed dramatically. The  $\bar{\rho}_{XY}$  is effectively reduced in TMIPCC, although there is also a mild reduction in TRAMI. As for the correlation between the truth lepton angles and the predicted ones, we find it is much increased in TRAMI, although adding the  $\bar{\rho}_{XY}$  reduces the value by a small amount. We note that the absolute value of MI is not useful, only the relative size has physical meaning.

For demonstration, we show the distributions of the lepton angles  $\cos \theta_{\ell^\pm}^*$  predicted by the TRAMI network for different polarization modes of the SM  $W^+W^-$  scattering at 13 TeV in Fig. 2. The truth-level  $\cos \theta_{\ell^\pm}^*$  distributions are also presented for comparison. Ideally, the truth-level distributions for the transverse and longitudinal polarized  $W$  are  $(1 \pm \cos \theta_\ell^*)^2$  and  $\sin^2 \theta_\ell^*$ , respectively. In practice, those shapes are distorted by the preselection cuts, especially around  $\cos \theta_\ell^* \sim \pm 1$ . We can conclude that the

TRAMI network can reproduce the lepton angle distributions well, although its performance of the  $W_L^+W_T^-/W_T^+W_L^-$  processes is slightly worse than that of the  $W_L^+W_L^-/W_T^+W_T^-$  processes, but the situation may change for different networks. In TMIPCC, the performance of the  $W_L^+W_T^-/W_T^+W_L^-$  processes is improved, while that of the  $W_L^+W_L^-$  process becomes worse. We provide the trained networks in the GitHub repository [42].

### C. Templates and fitting procedure

To obtain the polarization fractions, we need to fit the predicted two-dimensional  $\cos \theta_{\ell^+}^* - \cos \theta_{\ell^-}^*$  distribution to the predefined templates for the  $W_L^+W_L^-$ ,  $W_L^+W_T^-$ ,  $W_T^+W_L^-$ , and  $W_T^+W_T^-$  polarizations, respectively. The templates are obtained by applying the network to the events of the SM  $W^+W^-$  scattering with fixed final-state polarization. However, due to the presence of preselections, the templates exhibit some dependences on collision energy. We will need to use different sets of templates for different collision energies. In Fig. 3, we plot the two-dimensional templates obtained from both the TRAMI network predictions and the truth-level lepton angles. The network predictions act as nice proxies for the truth lepton angles.

Having established the template for each polarization state ( $T^i$ ,  $i = \text{LL, TL, LT, TT}$ ), given a two-dimensional  $\cos \theta_{\ell^+}^* - \cos \theta_{\ell^-}^*$  distribution  $O$ , we can perform the binned  $\chi^2$  fit to estimate the fraction ( $f_i$ ) of each polarization mode. Because of limited statistics, the two-dimensional lepton angle distribution is divided into  $10 \times 10$  bins. The particle swarm optimization [43] is adopted to minimize the  $\chi^2$  with constraints:  $\sum_i f_i = 1$  and  $f_i \in [0, 1]$ .

TABLE I. Network performances after training. Accuracy is defined as the number of correctly predicted events (true positives + true negatives) divided by the total number of events.  $\text{MI}(F; \Theta^*)$ ,  $\text{MI}(F; WW)$ ,  $\text{MI}(\Theta; \Theta^*)$  are the mutual information calculated over all training samples.  $F$  denotes the 64-dimensional features;  $WW$  denotes the variables of  $W$  pair momentum  $m(WW)$ ,  $E(WW)$ ,  $y(WW)$ ,  $\phi(WW)$ .

|        | Accuracy | $\text{MI}(F; \Theta^*)$ | $\text{MI}(F; WW)$ | $\bar{\rho}_{XY}$ | $\text{MI}(\Theta; \Theta^*)$ |
|--------|----------|--------------------------|--------------------|-------------------|-------------------------------|
| TRANS  | 0.44095  | 0.07139                  | 1.26769            | 0.23673           | 0.318141                      |
| TRAMI  | 0.42746  | 0.84366                  | 0.04764            | 0.19595           | 0.548189                      |
| TMIPCC | 0.42559  | 0.84437                  | 0.0473             | 0.08593           | 0.486755                      |

<sup>5</sup>This leads to different performances of a network at 13 and 100 TeV when the features are correlated with the collision energy.

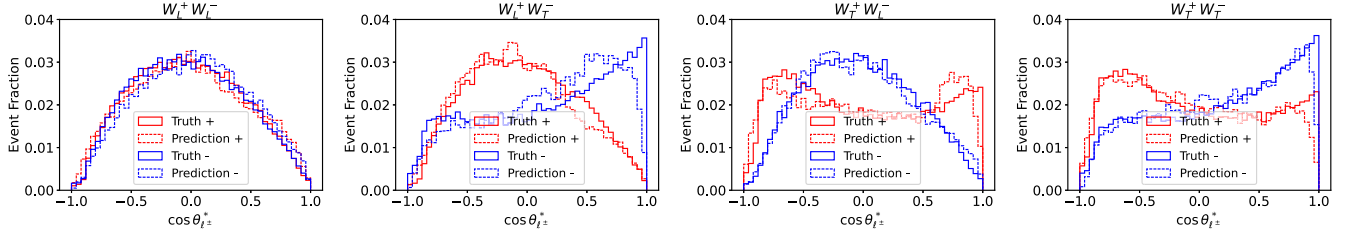


FIG. 2. Comparison of truth-level  $\cos \theta_{e^{\pm}}^*$  distributions and the TRAMI network output distributions for different polarization modes of the SM  $W^+W^-$  scattering at 13 TeV.

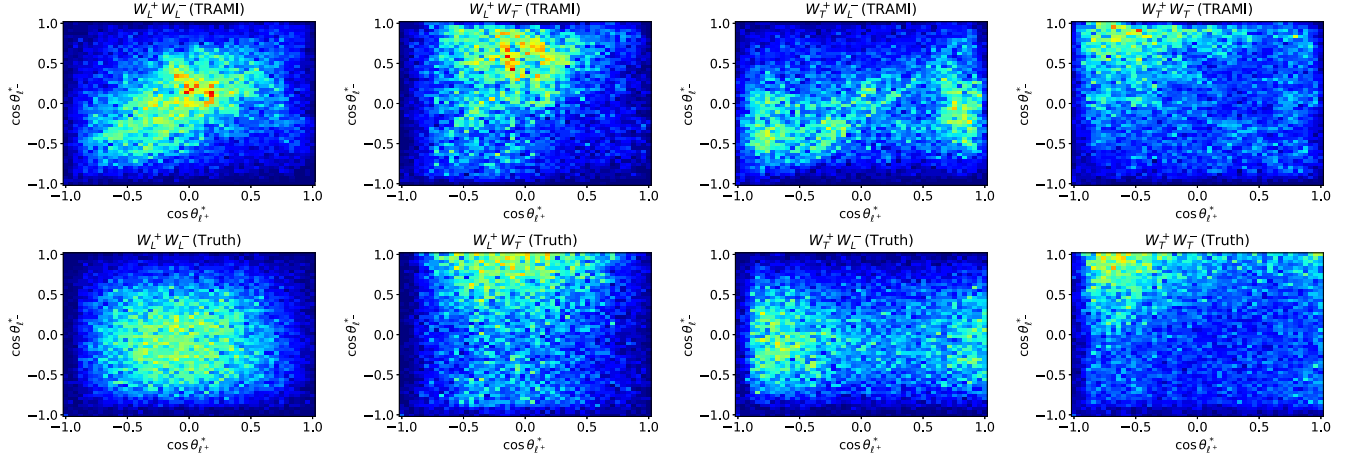


FIG. 3. The two-dimensional distributions of  $\cos \theta_{e^+}^*$  versus  $\cos \theta_{e^-}^*$  for different polarization modes of the SM  $W^+W^-$  scattering at 13 TeV. Upper panels: lepton angles  $\cos \theta_{e^{\pm}}^*$  calculated by the TRAMI network. Lower panels: truth-level lepton angles  $\cos \theta_{e^{\pm}}^*$ .

### III. TEST ON THE DIFFERENT MODELS AT 13 TeV

#### A. The $W^+W^-$ polarization in the SM and EFT

We first apply our network to the SM  $W^+W^-$  scattering at the 13 TeV LHC. The one-dimensional lepton angle distributions and the fitted polarization fractions obtained from three networks are shown in Fig. 4. The left panels show the comparison of the truth-level  $\cos \theta_{e^{\pm}}^*$  and the network output distributions, where we have projected the two-dimensional  $\cos \theta_{e^+}^* - \cos \theta_{e^-}^*$  distributions into each component for visibility. In the middle and right panels, the  $\Delta\chi^2 = 1$  contours on the  $f_{LT} - f_{TL}$  plane and  $f_{LL} - f_{TT}$  plane for integrated luminosities  $30 \text{ ab}^{-1}$  (may not realistic) and  $3 \text{ ab}^{-1}$  are shown. We can find that the networks reproduce the distributions of the truth-level lepton angle well.<sup>6</sup> In particular, the TRAMI network, which makes the features focus on the  $\cos \theta_{e^{\pm}}^*$  and decorrelate with the momentum of the  $W$  boson pair, has almost the same

reconstruction precisions as the truth  $\cos \theta_{e^{\pm}}^*$ ; i.e., the sizes of the contours are similar. As for the TMIPCC network, while the precision of the  $f_{LT}$  and  $f_{TL}$  fractions is similar to that obtained with the truth  $\cos \theta_{e^{\pm}}^*$ , both the  $f_{LL}$  and  $f_{TT}$  precision is worse, partly because of the difficulty in training the network with a more complex loss function. Overall, giving the cross section (after preselection cuts) of SM  $W^+W^-$  scattering as  $4.36 \text{ fb}$ , each fraction of polarization can be resolved with deviation of  $\sim 0.2$  at the LHC for an integrated luminosity of  $3 \text{ ab}^{-1}$ .

The  $W^+W^-$  scattering could be affected by any new physics that is related to the electroweak symmetry breaking of the SM. A general framework to describe the new physics effects is the EFT. To study the network performances in new physics models, following the strategy as discussed in Ref. [27], we consider the following operator [44,45]:

$$\mathcal{O}_H = \frac{\bar{c}_H}{2v^2} \partial^\mu [\Phi^\dagger \Phi] \partial_\mu [\Phi^\dagger \Phi] \Rightarrow \frac{\bar{c}_H}{2} \partial^\mu h \partial_\mu h, \quad (3.1)$$

where the  $\Phi$  field is the Higgs doublet and  $h$  denotes the SM Higgs boson field. This operator leads to the following changes to the Higgs couplings:

<sup>6</sup>Note that they are not so accurate in reproducing the lepton angle for a single event, as discussed in Sec. II A. For example, the root-mean-square errors (RMSEs) for the  $\cos \theta_{e^+}^*$  ( $\cos \theta_{e^-}^*$ ) of TRANS, TRAMI, and TMIPCC are 0.548 (0.546), 0.474 (0.470), and 0.515 (0.483), respectively.

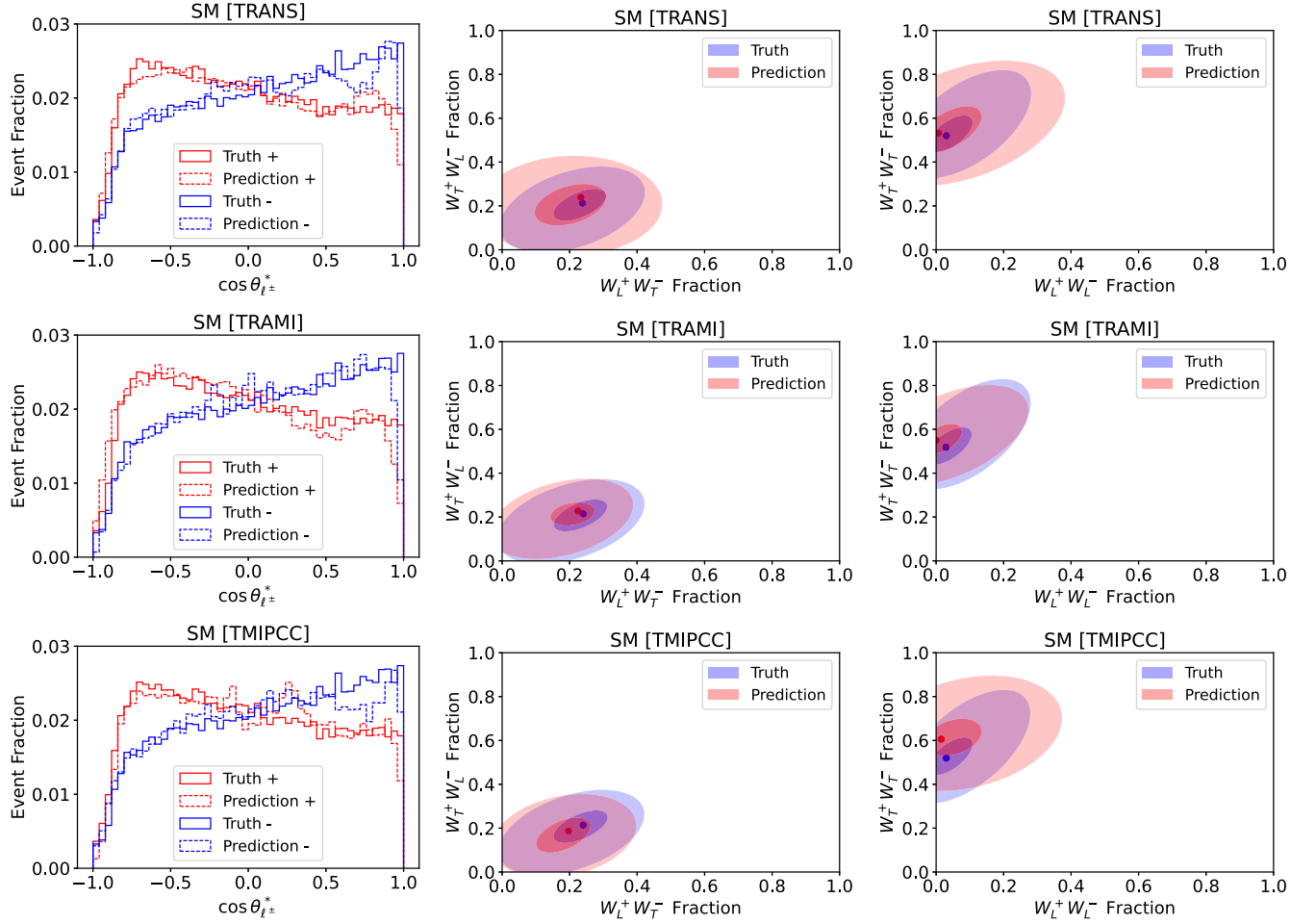


FIG. 4. The projected lepton angle ( $\cos \theta_{\ell^\pm}^*$ ) distributions and the  $\Delta\chi^2$  contours on the polarization fraction planes for the SM  $W^+W^-$  scattering at the 13 TeV LHC. Different shades from inside out correspond to  $\Delta\chi^2 = 1$  calculated on datasets with integrated luminosities of 30 and 3  $\text{ab}^{-1}$ . Plots from top to bottom are obtained with the TRANS, TRAMI, and TMIPCC networks, respectively. The corresponding results for truth-level  $\cos \theta_{\ell^\pm}^*$  are also presented for comparison.

$$\begin{aligned} \mathcal{L}_H \supset & \frac{gm_W}{c_W^2} \left[ 1 - \frac{1}{2} \bar{c}_H \right] Z_\mu Z^\mu h + gm_W \left[ 1 - \frac{1}{2} \bar{c}_H \right] W_\mu^\dagger W^\mu h \\ & + \left[ \frac{y_f}{\sqrt{2}} \left[ 1 - \frac{1}{2} \bar{c}_H \right] \bar{f} P_R f h + \text{H.c.} \right]. \end{aligned} \quad (3.2)$$

Although the updated global fit requires  $\bar{c}_H \lesssim 0.4$  [46], we apply our network to the case with  $\bar{c}_H = -1$  to illustrate the network performance when the new physics contributions are sizable. The results are given in Fig. 5. Compared to the SM case, the fraction of the longitudinal  $W$  boson is greatly enhanced due to the incomplete cancellation in  $W_L W_L \rightarrow W_L W_L$  scattering. However, the overall kinematic properties and the total production cross section (which is 4.82 fb after preselection for  $\bar{c}_H = -1$ ) of the  $W^+W^-$  scattering in the EFT with nonzero  $\bar{c}_H$  are similar to the SM ones, so the performances of the three networks on the EFT are similar to that on the SM as shown in Fig. 4. The one-dimensional lepton

angle distributions predicted by all three networks match the truth lepton angle distributions well. Among the three networks, the TRAMI performs the best—the discrimination power of which is quite close to the truth  $\cos \theta_{\ell^\pm}^*$ . On the other hand, in the TMIPCC network, only the precision of  $f_{LT}$  and  $f_{TL}$  is comparable to those obtained with the truth  $\cos \theta_{\ell^\pm}^*$ .

## B. The $W^+W^-$ polarization in the 2HDM and from resonant production

Many new physics models predict light states mediating the  $W^+W^-$  scattering, the effects of which cannot be fully described in the EFT. We consider the type-II 2HDM [47,48] as a benchmark model, as it is featured by both the Higgs coupling modification and the existence of another scalar mediator (besides the SM Higgs) in  $W^+W^-$  scattering. There are six parameters: masses of scalar bosons ( $m_{H_1}, m_{H_2}, m_A$ , and  $m_{H^\pm}$ ), the mixing angle between two

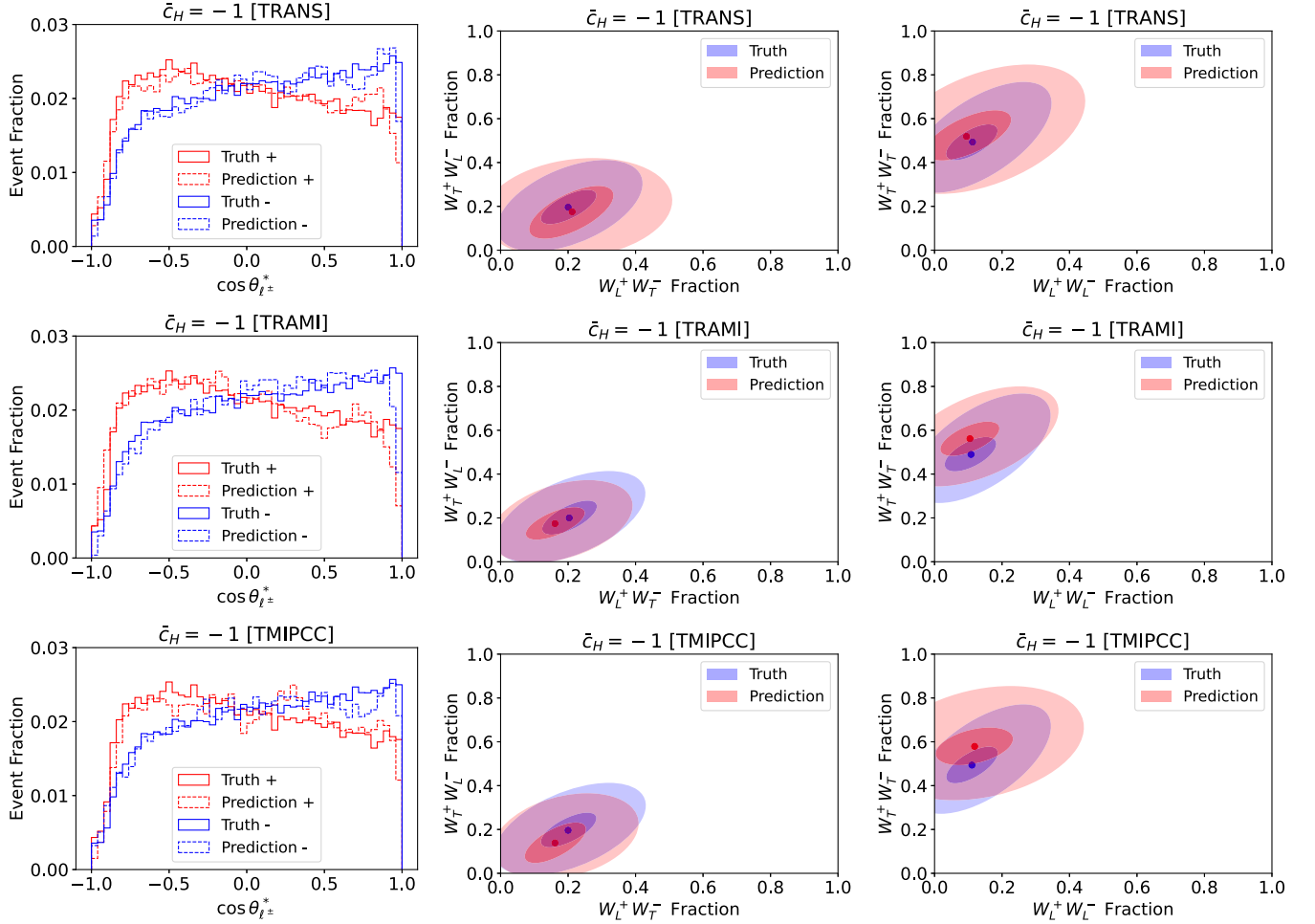


FIG. 5. For the  $W^+W^-$  scattering in EFT with  $\bar{c}_H = -1$  at the 13 TeV LHC. Meanings of the plots are the same as Fig. 4.

$CP$ -even scalars  $\alpha$ , and the ratio between two vacuum expectation values  $\tan\beta$ . The  $m_{H_1}$  has been measured to be around 125 GeV (we only consider the case of  $H_1$  being the SM-like Higgs boson). The  $m_A$  and  $m_{H^\pm}$  are not relevant in the  $W^+W^-jj$  production, assuming they are much larger than  $m_{H_2}$ . The couplings of  $CP$ -even scalars to the  $W$  boson are given by

$$\mathcal{L} \supset \frac{2m_W^2}{v} \sin(\alpha - \beta) H_1 W_\mu^+ W^{\mu-} + \frac{2m_W^2}{v} \cos(\alpha - \beta) H_2 W_\mu^+ W^{\mu-}. \quad (3.3)$$

The values of  $\tan\beta$  alone are not related to the  $HWW$  coupling, although it can influence the  $WW$  scattering indirectly via changing the decay width of  $H_2$ . We fix  $\tan\beta = 5$  without loss of generality and only need to deal with two free parameters:  $m_{H_2}$  and  $\sin(\alpha - \beta)$ .

In Fig. 6, the projected one-dimensional lepton angle ( $\cos\theta_{l^\pm}^*$ ) distributions and the  $\Delta\chi^2$  contours on the polarization fraction planes for the  $W^+W^-$  scattering in 2HDM

with  $m_{H_2} = 300$  GeV and  $\sin(\alpha) = 0.7$  at the 13 TeV LHC are shown. Because of the resonant contribution from the  $H_2$ , the cross section of the  $W^+W^-$  scattering is increased to 8.362 fb. As in the SM and EFT, the lepton angle distributions can be reproduced well by all three networks. However, the precision of the polarization fractions obtained from the TRANS network are not as good as those obtained from the truth lepton angles. This is because the features in the TRANS network (which is trained only with the SM events) contain the SM kinematic information (in particular, the invariant mass of the  $W$  boson pair). The differences between the kinematic properties of the 2HDM and the SM degrade the performance. The situation is much improved for the TRAMI network, in which the information of  $W$  boson pair momentum is decorrelated from the features.

To illustrate the performance in a more extreme case, we apply those networks to the process of  $W^+W^-$  scattering solely through a heavy resonance. Both the polarization pattern and the kinematic features are dramatically different from the SM ones. Assuming the mediator is a scalar



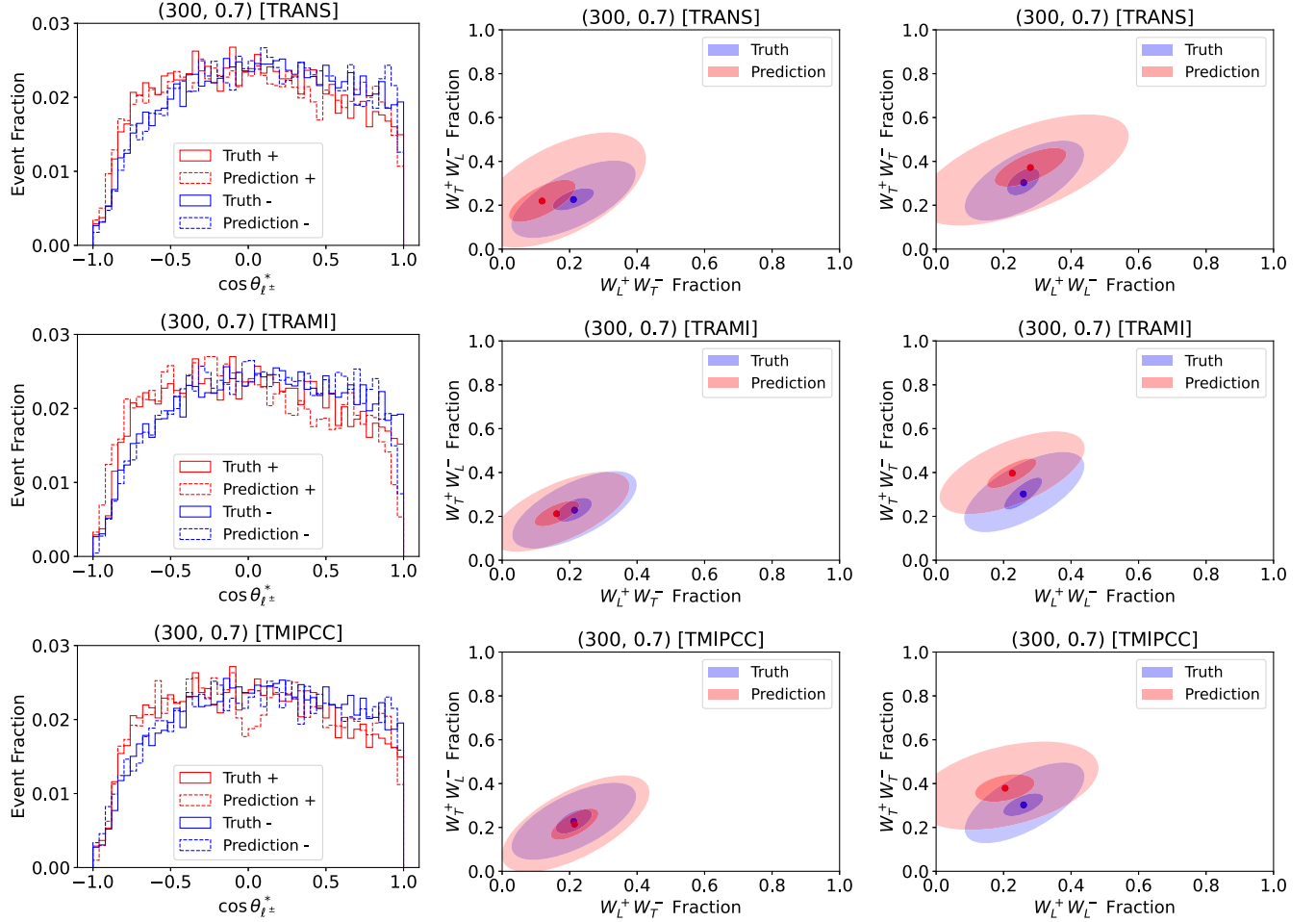


FIG. 6. For the  $W^+W^-$  scattering in the 2HDM with  $m_{H_2} = 300$  GeV and  $\sin(\alpha) = 0.7$  at the 13 TeV LHC. Meanings of the plots are the same as Fig. 4.

boson, only the polarization modes  $W_L W_L$  and  $W_T W_T$  are allowed. The ratio  $f_{LL}/f_{TT} \sim 67$  for the scalar mass is around 400 GeV and is increasing fast for a heavier scalar. In Fig. 7, we show the comparison of the truth-level  $\cos \theta_{\ell^\pm}^*$  and the network output distributions for the  $W^+W^-$  scattering through a heavy scalar resonance at the 13 TeV LHC. As expected, the TRAMI network performs much better than the TRANS network, since the TRAMI is not supposed to be sensitive to the  $WW$  production mechanism. The reconstructed lepton angle distributions from the TRAMI network remain close to the truth ones for resonance mass less than  $\sim 1$  TeV. The heavier the resonance is, the larger the deviation between the network predictions and truth values.

Up to this point, we have not found it necessary to add the PCC to the loss function, as the performance of the TMIPCC network is not comparable to that of the TRAMI network in all cases. This is mainly because the training samples for the transformer network are generated at 13 TeV, and the decorrelation of collision energy will not be necessary if we are extracting the  $W$  polarization

fractions at the same collision energy. In this case, the TRAMI network performs the best and is recommended to use. However, if we want to apply the same network to processes at different collision energy, the subtraction of collision energy dependence becomes essential. In the next section, we further study the network performance at different collision energy by taking the  $W^+W^-$  scattering in the 2HDM at 100 TeV as an example.

#### IV. THE $W^+W^-$ POLARIZATION IN 100 TeV $p$ - $p$ COLLISION

As we discussed above, although the distributions of  $\theta_{\ell^\pm}^*$  are supposed to be only related to the  $W$  boson polarization, the effects of preselection cuts, which distort the lepton angle distribution, depend on the collision energy. The same preselection cuts as proposed in Sec. II B for 13 TeV are also adopted here.

Figure 8 shows the two-dimensional templates on the  $\cos \theta_{\ell^+}^* - \cos \theta_{\ell^-}^*$  plane for the TMIPCC network prediction and the truth lepton angles at 100 TeV. Compared with Fig. 3,

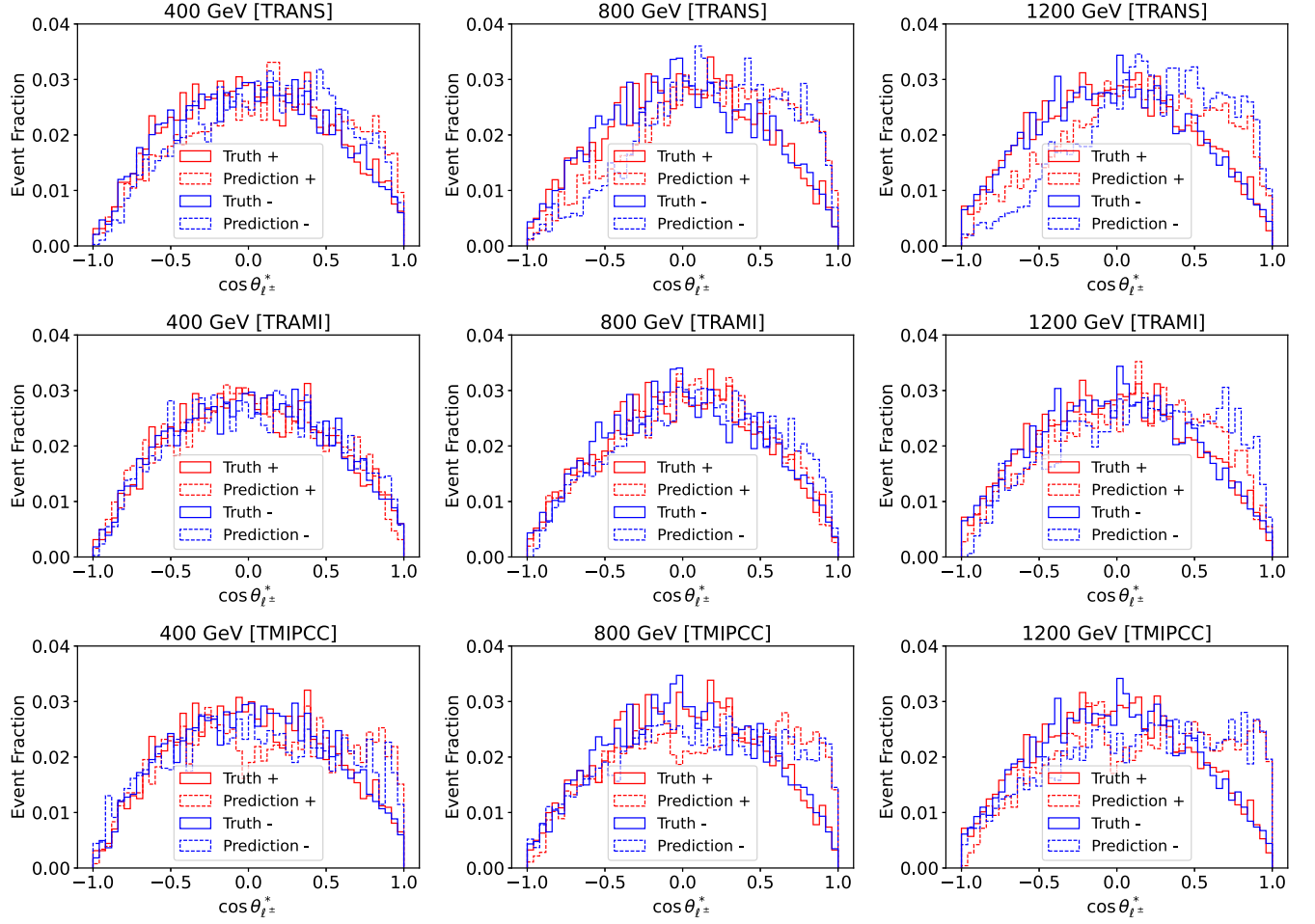


FIG. 7. A comparison of truth-level  $\cos\theta_{l^\pm}^*$  and the network output distributions for the resonant  $W^+W^-$  scattering at the 13 TeV LHC with resonance mass 400, 800, and 1200 GeV (from left to right). Plots from top to bottom are obtained with the TRANS, TRAMI, and TMIPCC networks, respectively.

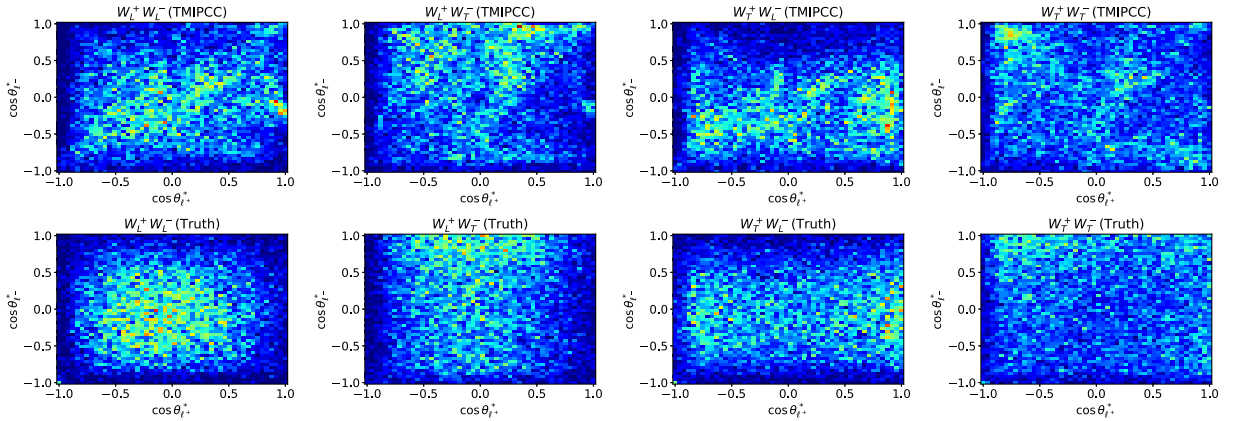


FIG. 8. Two-dimensional distributions of  $\cos\theta_{l^+}^* - \cos\theta_{l^-}^*$  for different polarization modes of the SM  $W^+W^-$  scattering at 100 TeV. Upper panels: TMIPCC network prediction. Lower panels: truth  $\cos\theta_{l^\pm}^*$ .

we can find that the distributions for the truth lepton angle vary with collision energy. In particular, the effects of the preselection cuts are milder at the 100 TeV collision, leading

to slightly sharper lepton angle distributions at the truth level. On the other hand, the templates from the network predictions become less precise for the 100 TeV case even after applying

both MI and PCC in the loss function. This is attributed to the fact that the transformer networks are only trained on events at the 13 TeV LHC. It is possible that one can optimize the results at 100 TeV by using 100 TeV event samples to train the transformer networks.

The projected one-dimensional  $\cos\theta_{\ell^\pm}^*$  distributions, as well as the fitted polarization fractions from the network predictions and truth lepton angles, are presented in Fig. 9. The  $W^+W^-$  scattering in the 2HDM with  $m_{H_2} = 300$  GeV and  $\sin(\alpha) = 0.7$  at 100 TeV has been taken as an example. After preselection, the production cross section for the process is 148.74 fb. Unlike the 13 TeV case, the reduced performance of the TRANS network is visible in the  $\cos\theta_{\ell^\pm}^*$  distribution this time. As for the polarization fraction, the TRAMI network can no longer work well, and decorrelating the collision energy dependence is essential. We can see that the TMIPCC network outperforms TRAMI in the 100 TeV case, although there is still a certain amount of deviation between the predicted ones and the truth ones.

## V. SUBTRACTING THE BACKGROUNDS: SM AS A CASE STUDY

So far, we have only considered the application of the networks to the  $W^+W^-$  scattering processes. In practice, there will be events of non-VBS processes that pass the preselections, behaving as backgrounds in our analysis. As a result, we can only obtain the superposed distribution of  $\cos\theta_{\ell^+}^* - \cos\theta_{\ell^-}^*$ , from which the contributions from the background processes need to be subtracted out before applying the fit to the templates. However, due to the uncertainties in the backgrounds simulation, the background subtraction cannot be perfect. This will lead to reduced precision in extracting the polarization fractions.

Since we are considering the dileptonic channel of the  $W^+W^-$  scattering, the dominant background processes are the dileptonic  $t\bar{t}$  and  $tW$  processes, mixed electroweak-QCD  $W^+W^-jj$  production, as well as the  $WZjj$  production [both at orders of  $\mathcal{O}(\alpha_{EW}^4)$  and  $\mathcal{O}(\alpha_{EW}^2\alpha_s^2)$ ] with gauge bosons decaying leptonically. The production cross

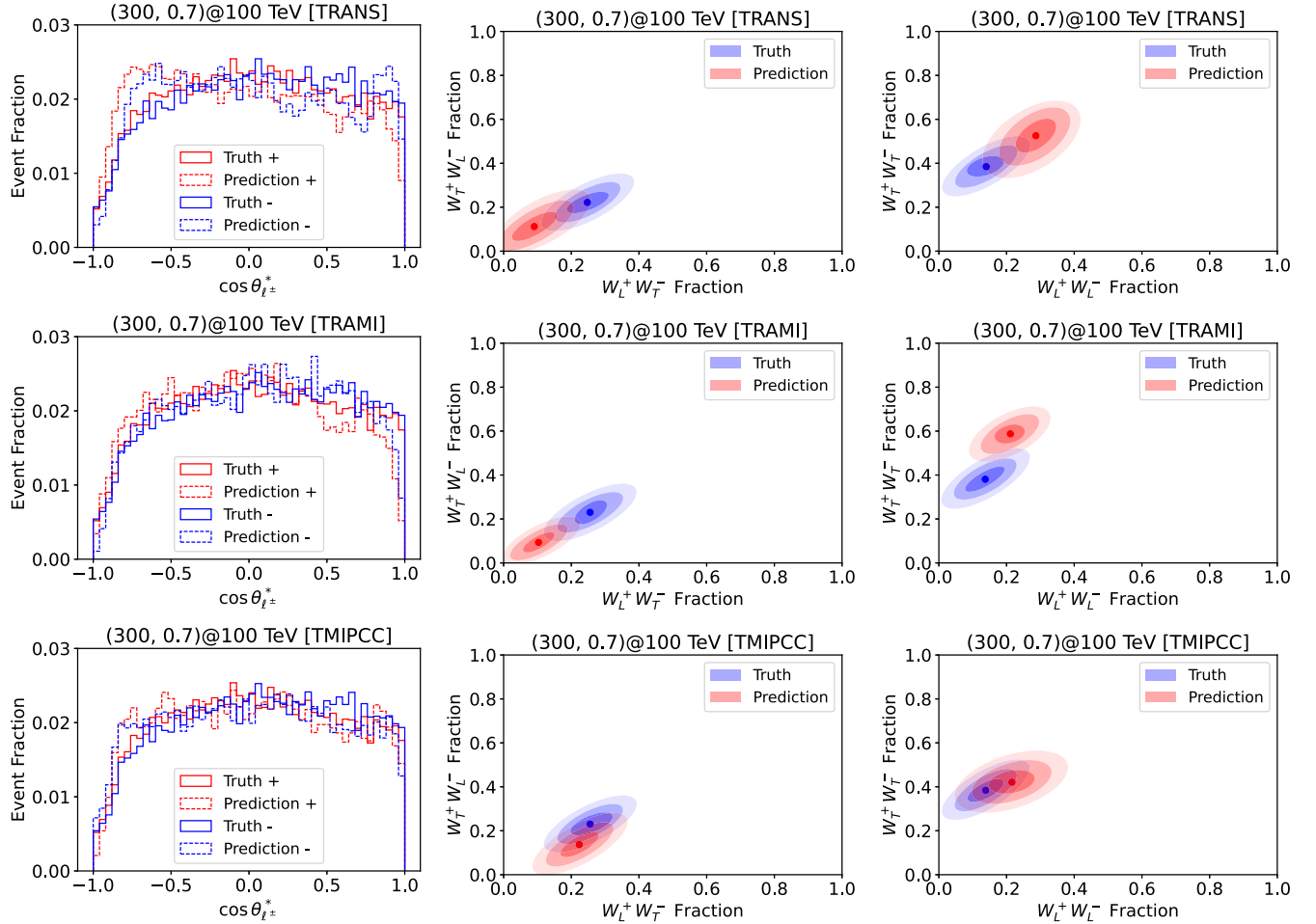


FIG. 9. For the  $W^+W^-$  scattering in the 2HDM with  $m_{H_2} = 300$  GeV and  $\sin(\alpha) = 0.7$  at 100 TeV. Meanings of the plots are the same as Fig. 4, except that the different shades of the  $\Delta\chi^2$  contours from inside out correspond to  $\Delta\chi^2 = 1$  calculated on datasets with integrated luminosities of  $3 \text{ ab}^{-1}$ ,  $1 \text{ ab}^{-1}$ , and  $500 \text{ fb}^{-1}$ , respectively.

TABLE II. The production cross sections of background processes before and after preselections at the 13 TeV LHC. The superscripts EW and QCD denote the processes at order of  $\mathcal{O}(\alpha_{\text{EW}}^4)$  and  $\mathcal{O}(\alpha_{\text{EW}}^2\alpha_s^2)$ , respectively. The subscript  $\ell$  denotes the leptonic decay of that particle.

|                            | $t\bar{t}_\ell$ | $tW_\ell/t_\ell W$ | $W_\ell W_{jj}^{\text{QCD}}$ | $W_\ell Z_{jj}^{\text{QCD}}$ | $W_\ell Z_{jj}^{\text{EW}}$ |
|----------------------------|-----------------|--------------------|------------------------------|------------------------------|-----------------------------|
| $\sigma^{\text{fid}}$ (pb) | 210.3           | 15.9               | 4.68                         | 2.20                         | 0.487                       |
| $\sigma^{\ell\ell}$ (fb)   | 139.8           | 11.6               | 14.7                         | 4.49                         | 3.68                        |

sections at 13 TeV for the simulated background events before ( $\sigma^{\text{fid}}$ ) and after ( $\sigma^{\ell\ell}$ ) the preselection cuts are listed in Table II. For diboson processes, the transverse momenta of final-state jets are required to be greater than 20 GeV. We will use the measured inclusive cross sections at the LHC for the  $t\bar{t}$  [49] and  $tW$  [50] processes, and use the leading order cross sections which are calculated by MG5\_AMC@NLO for the diboson processes. We note that background events are simulated with at least one lepton in the final state because there could be a misidentified fake lepton due to detector effects.

With background contamination, we adopt the results from the TRAMI network to extract the  $W^+W^-$  polarization fractions for the SM production at the 13 TeV LHC. The results are shown in Fig. 10 with varying uncertainties in background subtraction. We have assumed uncorrelated systematic uncertainties for the event numbers in lepton angle bins ( $10 \times 10$  on the  $\cos\theta_{\ell^+}^* - \cos\theta_{\ell^-}^*$  plane). The size of the systematic uncertainty in each bin is indicated in the legend. The left panel shows the projected lepton angle ( $\cos\theta_{\ell^\pm}^*$ ) distributions given by the summed templates with the best-fitted fractions, as well as that obtained at the truth level. Since the results with three levels of background uncertainties have similar best-fit values, the lepton angle distributions are similar for all three cases. However, the total background cross section after preselection is around two orders of magnitude larger than the signal cross section. The uncertainties of the fitted fractions are very sensitive to the background uncertainties; i.e., the size of

the  $\Delta\chi^2 = 1$  contour is substantially enlarged for increasing background uncertainty. The precision of the extracted fractions is promising only if the background uncertainty in subtraction can be controlled at the 0.1% level. Note that this uncertainty can be much smaller than that of the total cross section. More refined cuts are necessary for large systematic uncertainty of the background. In this case, the template for each polarization should be adjusted accordingly, and the  $\chi^2$  fit should be done on the (network-predicted) lepton angle distribution after cuts. Moreover, with more stringent cuts, a larger number of background events need to be simulated in order to guarantee relatively small statistical uncertainties in our analysis. The main point of the paper is to reproduce the lepton angle distribution, so we decide to leave those more involved analyses for future work.

## VI. CONCLUSION

We propose networks composed of a transformer network and CGAN to predict the distributions of the angles between the charged leptons in the gauge boson rest frames and the gauge boson directions of motion for the dileptonic channel of  $W^+W^-$  scattering so that the polarization fractions of the  $W^+W^-$  final state can be obtained from fitting the predicted lepton angle distribution to the given templates.

There could be unknown new physics contributing to the  $W^+W^-$  scattering, which may lead to dramatically different kinematic properties for final states. To ensure that the network is able to predict the lepton angle distribution precisely, irrespective of the  $W^+W^-$  production mechanism, the loss function of the transformer network is modified with MI and PCC as defined in Eqs. (2.3) and (2.4) so that the features produced by the transformer network contain the lepton angle information as much as possible while decorrelating with other kinematic variables. For comparison, three different versions of networks are trained, denoted by TRANS, TRAMI, and TMIPCC.

To illustrate the performances of the networks, we apply them to the events of  $W^+W^-$  scattering with dileptonic

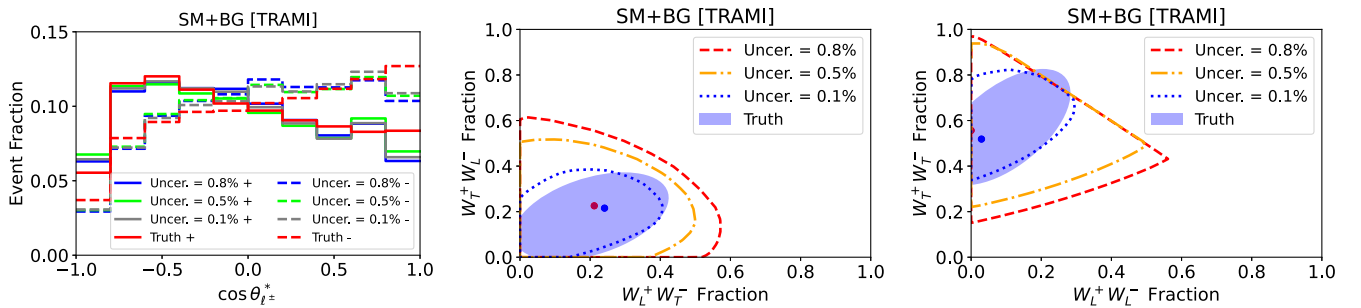


FIG. 10. The projected lepton angle ( $\cos\theta_{\ell^\pm}^*$ ) distributions and fitted  $\Delta\chi^2$  contours for the SM  $W^+W^-$  scattering at the 13 TeV LHC with integrated luminosity of  $3 \text{ ab}^{-1}$ . The background contributions are subtracted with uncertainties as indicated in the legends. The results of the TRAMI network are shown.

TABLE III. Best-fit values and  $1\sigma$  uncertainties for the fraction of longitudinal polarized  $W_L^+ W_L^-$ , as shown in Figs. 4–6 and 9. With different integrated luminosity, the best-fit values are kept the same while the uncertainties are different.

|        | SM      |                                       | EFT $\bar{c}_H = -1$ |                                       | 2HDM (300, 0.7) |                                       | 2HDM (300, 0.7)@100 TeV |   |
|--------|---------|---------------------------------------|----------------------|---------------------------------------|-----------------|---------------------------------------|-------------------------|---|
|        | Center  | Uncertainty<br>3(30) $\text{ab}^{-1}$ | Center               | Uncertainty<br>3(30) $\text{ab}^{-1}$ | Center          | Uncertainty<br>3(30) $\text{ab}^{-1}$ | Center                  | Uncertainty<br>0.5(1)[3] $\text{ab}^{-1}$ |
| Truth  | 0.03041 | 0.19(0.06)                            | 0.11254              | 0.21(0.07)                            | 0.25959         | 0.17(0.04)                            | 0.14001                 | 0.13(0.09)[0.05]                          |
| TRANS  | 0.00682 | 0.30(0.09)                            | 0.09464              | 0.29(0.12)                            | 0.27940         | 0.31(0.10)                            | 0.28665                 | 0.14(0.10)[0.06]                          |
| TRAMI  | 0.00040 | 0.23(0.05)                            | 0.10504              | 0.23(0.08)                            | 0.22584         | 0.21(0.07)                            | 0.21123                 | 0.12(0.08)[0.04]                          |
| TMIPCC | 0.01529 | 0.28(0.08)                            | 0.11879              | 0.30(0.11)                            | 0.20456         | 0.27(0.08)                            | 0.21595                 | 0.16(0.11)[0.06]                          |

decay in the SM, in the EFT with nonzero  $\bar{c}_H$ , as well as in the 2HDM with chosen benchmark points. The results are summarized in Table III.

The TRAMI network performs best at 13 TeV for all models, as the features of it have been trained to focus on the lepton angle while not being sensitive to the  $W$  boson pair production mechanism. The fitting precision of the polarization fraction based on the TRAMI predictions is quite similar to that obtained from using the truth lepton angle, except for the  $f_{TT}$  in the 2HDM with  $m_{H_2} = 300$  GeV and  $\sin(\alpha - \beta) = 0.7$ . There is a certain amount of deviation, mainly due to the remaining information of the kinematic variables in the features of the TRAMI network. The  $1\sigma$  ranges for the fitted fractions are around 0.2–0.3 for an integrated luminosity of  $3 \text{ ab}^{-1}$ . When applying to the events at 100 TeV, the reduced performances of the TRANS network and TRAMI network become visible in the projected one-dimensional lepton angle  $\cos \theta_{\ell^\pm}^*$  distributions. The situation is much improved for the TMIPCC network in which the decorrelation with collision energy is conducted, although there are still mild

deviations between the polarization fractions obtained from the TMIPCC network and the truth lepton angle. Benefited from the increased production rate at higher collision energy, the  $1\sigma$  ranges for the fitted fractions can reach  $\sim 0.1$  (0.05) for an integrated luminosity of  $1 \text{ ab}^{-1}$  ( $3 \text{ ab}^{-1}$ ).

In practice, the opposite-sign dileptonic channel of  $W^+ W^-$  scattering suffers from backgrounds of dileptonically decaying  $t\bar{t}$ ,  $tW$ , mixed electroweak QCD  $W^+ W^- jj$ , as well as  $WZjj$  productions. Considering the uncertainty in background subtraction, the fitting precision of polarization fractions is substantially reduced, mainly due to the relatively small signal-to-background ratio (after applying the preselections).

## ACKNOWLEDGMENTS

This work was supported in part by the Fundamental Research Funds for the Central Universities by the National Natural Science Foundation of China under Grant No. 11905149.

- 
- [1] M. Rauch, Vector-boson fusion and vector-boson scattering, [arXiv:1610.08420](#).
- [2] D. R. Green, P. Meade, and M.-A. Pleier, Multiboson interactions at the LHC, *Rev. Mod. Phys.* **89**, 035008 (2017).
- [3] D. Buarque *et al.*, Vector boson scattering processes: Status and prospects, [arXiv:2106.01393](#).
- [4] R. Covarelli, M. Pellen, and M. Zaro, Vector-boson scattering at the LHC: Unraveling the electroweak sector, *Int. J. Mod. Phys. A* **36**, 2130009 (2021).
- [5] M. Aaboud *et al.* (ATLAS Collaboration), Observation of Electroweak Production of a Same-Sign  $W$  Boson Pair in Association with Two Jets in  $pp$  Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector, *Phys. Rev. Lett.* **123**, 161801 (2019).
- [6] A. M. Sirunyan *et al.* (CMS Collaboration), Observation of Electroweak Production of Same-Sign  $W$  Boson Pairs in the Two Jet and Two Same-Sign Lepton Final State in Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV, *Phys. Rev. Lett.* **120**, 081801 (2018).
- [7] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of vector boson scattering and constraints on anomalous quartic couplings from events with four leptons and two jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV, *Phys. Lett. B* **774**, 682 (2017).
- [8] G. Aad *et al.* (ATLAS Collaboration), Observation of electroweak production of two jets and a Z-boson pair with the ATLAS detector at the LHC, [arXiv:2004.10612](#).
- [9] M. Aaboud *et al.* (ATLAS Collaboration), Observation of electroweak  $W^\pm Z$  boson pair production in association with two jets in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector, *Phys. Lett. B* **793**, 469 (2019).
- [10] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of electroweak  $WZ$  boson production and search for new

- physics in  $WZ + \text{two jets}$  events in  $pp$  collisions at  $\sqrt{s} = 13$  TeV, *Phys. Lett. B* **795**, 281 (2019).
- [11] G. Aad *et al.* (ATLAS Collaboration), Search for the electroweak diboson production in association with a high-mass dijet system in semileptonic final states in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector, *Phys. Rev. D* **100**, 032007 (2019).
- [12] A. M. Sirunyan *et al.* (CMS Collaboration), Search for anomalous electroweak production of vector boson pairs in association with two jets in proton-proton collisions at 13 TeV, *Phys. Lett. B* **798**, 134985 (2019).
- [13] T. Han, D. Krohn, L.-T. Wang, and W. Zhu, New physics signals in longitudinal gauge boson scattering at the LHC, *J. High Energy Phys.* **03** (2010) 082.
- [14] A. Ballestrero, E. Maina, and G. Pelliccioli,  $W$  boson polarization in vector boson scattering at the LHC, *J. High Energy Phys.* **03** (2018) 170.
- [15] A. Ballestrero, E. Maina, and G. Pelliccioli, Different polarization definitions in same-sign  $WW$  scattering at the LHC, *Phys. Lett. B* **811**, 135856 (2020).
- [16] A. Ballestrero, E. Maina, and G. Pelliccioli, Polarized vector boson scattering in the fully leptonic  $WZ$  and  $ZZ$  channels at the LHC, *J. High Energy Phys.* **09** (2019) 087.
- [17] E. Maina, Vector boson polarizations in the decay of the Standard Model Higgs, *Phys. Lett. B* **818**, 136360 (2021).
- [18] S. De, V. Rentala, and W. Shepherd, Measuring the polarization of boosted, hadronic  $W$  bosons with jet substructure observables, [arXiv:2008.04318](https://arxiv.org/abs/2008.04318).
- [19] CMS Collaboration, Prospects for the study of vector boson scattering in same sign  $WW$  and  $WZ$  interactions at the HL-LHC with the upgraded CMS detector, CERN Technical Report No. CMS-PAS-SMP-14-008, 2016.
- [20] CMS Collaboration, Study of  $W^\pm W^\pm$  production via vector boson scattering at the HL-LHC with the upgraded CMS detector, CERN Technical Report No. CMS-PAS-FTR-18-005, 2018.
- [21] J. Searcy, L. Huang, M.-A. Pleier, and J. Zhu, Determination of the  $WW$  polarization fractions in  $pp \rightarrow W^\pm W^\pm jj$  using a deep machine learning technique, *Phys. Rev. D* **93**, 094033 (2016).
- [22] M. Grossi, J. Novak, D. Rebutti, and B. Kersevan, Comparing traditional and deep-learning techniques of kinematic reconstruction for polarisation discrimination in vector boson scattering, *Eur. Phys. J. C* **80**, 1144 (2020).
- [23] J. Lee, N. Chanon, A. Levin, J. Li, M. Lu, Q. Li, and Y. Mao, Polarization fraction measurement in same-sign  $WW$  scattering using deep learning, *Phys. Rev. D* **99**, 033004 (2019).
- [24] J. Lee, N. Chanon, A. Levin, J. Li, M. Lu, Q. Li, and Y. Mao, Polarization fraction measurement in  $ZZ$  scattering using deep learning, *Phys. Rev. D* **100**, 116010 (2019).
- [25] D. Buarque Franzosi, O. Mattelaer, R. Ruiz, and S. Shil, Automated predictions from polarized matrix elements, *J. High Energy Phys.* **04** (2020) 082.
- [26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, and I. Polosukhin, Attention is all you need, in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Red Hook, NY, USA* (Curran Associates Inc., New York City, 2017), pp. 6000–6010.
- [27] J. Li, S. Yang, and R. Zhang, Detecting anomaly in vector boson scattering, *Chin. Phys. C* **45**, 073104 (2021).
- [28] M. Mirza and S. Osindero, Conditional generative adversarial nets, [arXiv:1411.1784](https://arxiv.org/abs/1411.1784).
- [29] I. Belghazi, S. Rajeswar, A. Baratin, R. D. Hjelm, and A. C. Courville, MINE: Mutual information neural estimation, [arXiv:1801.04062](https://arxiv.org/abs/1801.04062).
- [30] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition* (IEEE Computer Society, Los Alamitos, 2016).
- [31] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, Analyzing and improving the image quality of STYLEGAN, [arXiv:1912.04958](https://arxiv.org/abs/1912.04958).
- [32] 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**, 2250 (2014).
- [33] A. Alloul, B. Fuks, and V. Sanz, Phenomenology of the Higgs effective Lagrangian via FEYNRULES, *J. High Energy Phys.* **04** (2014) 110.
- [34] 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, *J. High Energy Phys.* **07** (2014) 079.
- [35] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations, *J. High Energy Phys.* **03** (2013) 015.
- [36] T. Sjostrand, S. Mrenna, and P. Z. Skands, A brief introduction to PYTHIA 8.1, *Comput. Phys. Commun.* **178**, 852 (2008).
- [37] M. Cacciari, G. P. Salam, and G. Soyez, FASTJET user manual, *Eur. Phys. J. C* **72**, 1896 (2012).
- [38] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast simulation of a generic collider experiment, *J. High Energy Phys.* **02** (2014) 057.
- [39] ATLAS Collaboration, Optimisation of the ATLAS  $b$ -tagging performance for the 2016 LHC Run, Report No. ATL-PHYS-PUB-2016-012, 2016.
- [40] B. Biedermann, A. Denner, and M. Pellen, Complete NLO corrections to  $W^+W^+$  scattering and its irreducible background at the LHC, *J. High Energy Phys.* **10** (2017) 124.
- [41] F. Campanario, M. Kerner, D. Ninh, and I. Rosario, Diphoton production in vector-boson scattering at the LHC at next-to-leading order QCD, *J. High Energy Phys.* **06** (2020) 072.
- [42] GitHub repository, <https://github.com/scu-heplab/wlwl-polarization>.
- [43] G. Venter and S. S. Jaroslaw, Particle swarm optimization, *AIAA J.* **41**, 1583 (2003).
- [44] G. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi, The strongly-interacting light Higgs, *J. High Energy Phys.* **06** (2007) 045.
- [45] R. Contino, M. Ghezzi, C. Grojean, M. Muhlleitner, and M. Spira, Effective Lagrangian for a light Higgs-like scalar, *J. High Energy Phys.* **07** (2013) 035.

- [46] S. Dawson, S. Homiller, and S. D. Lane, Putting SMEFT fits to work, *Phys. Rev. D* **102**, 055012 (2020).
- [47] M. Aoki, S. Kanemura, K. Tsumura, and K. Yagyu, Models of Yukawa interaction in the two Higgs doublet model, and their collider phenomenology, *Phys. Rev. D* **80**, 015017 (2009).
- [48] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J.P. Silva, Theory and phenomenology of two-Higgs-doublet models, *Phys. Rep.* **516**, 1 (2012).
- [49] CMS Collaboration, Measurement of the  $t\bar{t}$  production cross section at 13 TeV in the all-jets final state, Report No. CMS-PAS-TOP-16-013, 2016.
- [50] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of the production cross section for single top quarks in association with  $W$  bosons in proton-proton collisions at  $\sqrt{s} = 13$  TeV, *J. High Energy Phys.* **10** (2018) 117.