

Detecting anomalous quartic gauge couplings using the isolation forest machine learning algorithm

Li Jiang, Yu-Chen Guo[✉], and Ji-Chong Yang^{✉*}

Department of Physics, Liaoning Normal University, Dalian 116029, China



(Received 28 February 2021; accepted 4 August 2021; published 23 August 2021)

The search of new physics (NP) beyond the Standard Model is one of the most important tasks of high energy physics. A common characteristic of the NP signals is that they are usually small in number and kinematically different. We use a model independent strategy to study the phenomenology of NP by directly picking out and studying the kinematically unusual events. For this purpose, the isolation forest (IF) algorithm is applied, which is found to be efficient in identifying the signal events of the anomalous quartic gauge couplings (aQGCs). The IF algorithm can also be used to constrain the coefficients of aQGCs. As a machine learning algorithm, the IF algorithm shows good prospects in future studies of NP.

DOI: [10.1103/PhysRevD.104.035021](https://doi.org/10.1103/PhysRevD.104.035021)

I. INTRODUCTION

Despite the great success of the Standard Model (SM), there are still many unanswered questions making the search for new physics (NP) beyond the SM a very important issue [1]. Except for a few cases that are known or suspected to deviate from the SM [2], in most cases the experiments are in good agreement with the SM. The search of NP is to look for a small number of anomalies in the vast amount of experimental data. Meanwhile, the kinematic features of the events induced by NP are usually different from the SM, which is the reason that event selection strategies (ESSs) are commonly applied in the phenomenological studies of NP. From the perspective of the SM effective theory (SMEFT) [3], that is because the signals are induced by new interactions different from the SM. It follows that the search of NP is to search the events which are ‘few and kinematically different’. In this paper, we make use of the above features of NP and use a model independent strategy to directly search for kinematically unusual events.

This strategy has the advantage that it can be applied generally; taking the SMEFT for example, is also a model independent way to search for the NP signals. For one generation of fermions, there are 895 baryon number conserving dimension-eight operators [4,5], and

kinematic analysis needs to be done for each operator. Compared with the SMEFT, our strategy is independent of the operators. Except for that, unlike the search of NP in a process which may turn out to be a wasted effort, the kinematically unusual events are always worth attention even if they are not from NP. They could be faults in the experiments, or they could be the rare processes allowed by the SM.

As an example, we use our strategy to study the anomalous quartic gauge couplings (aQGCs) [6], which are modifications to the SM gauge interactions intensively studied [4,7,8]. aQGCs can be contributed by many NP models such as Born-Infeld theory, composite Higgs, warped extra dimensions, two Higgs doublet models, $U(1)_{L_\mu-L_\tau}$, and axionlike particles [9,10]. Since dimension-six operators cannot contribute to aQGCs independently [4], we concentrate on the dimension-eight operators. A recent study shows that the dimension-eight operators are important in the convex geometry point of view to the SMEFT space [11]. Besides, there are cases sensitive to dimension-eight operators because the contributions from dimension-six operators are absent [9,12,13]. In the case of anomalous gauge couplings, aQGCs can lead to richer helicity combinations than dimension-six anomalous trilinear gauge couplings (aTGCs) [14]. aQGCs can originate from tree diagrams while dimension-six aTGCs are generated by loop diagrams [15]. Consequently, while the SMEFT has mainly been applied with dimension-six operators, the importance of dimension-eight operators has been pointed out in many previous studies [5,6,12].

The search of ‘few and kinematically different’ events is in fact an anomaly detection (AD), the applications of which in high energy physics (HEP) are developing

*Corresponding author.
yangjichong@lnnu.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

extensively recently [16]. AD is suitable for machine learning (ML), which has been used in various aspects of HEP [17]. When it comes to AD, there are many algorithms such as autoencoder [18–20], multivariate Gaussian mixture model [20,21], deep support vector data description [19,22], and isolation forest (IF) [19,20,23]. We use the IF algorithm because the mechanism behind IF algorithm is transparent, it merely identifies the points which are few and far away from the others. Moreover, it is expected to perform better with fewer signal events, and it is efficient to apply and easy to implement. We find that the IF algorithm works as an automatic ESS and can identify the signal events very well. Besides, IF algorithm can also be applied to constrain the parameters of NP models such as the coefficients of aQGCs, therefore it has a lot of potential in the future studies of NP.

The rest of the paper is organized as follows. In Sec. II we briefly introduce the IF algorithm. In Sec. III, the application of IF algorithm on detecting the signals of aQGCs is presented. Section IV is a summary.

II. A BRIEF INTRODUCTION OF ISOLATION FOREST

IF algorithm is an algorithm with linear complexity designed for detection of point anomalies. It makes use of the fact that the anomalies are ‘few and different’. It can be applied for multidimensional data efficiently. We briefly introduce the IF algorithm following Ref. [23].

The key step of IF algorithm is to build an ensemble of isolation trees (ITs). The IT is a binary tree structure randomly generated to isolate every single point. Denoting each point in the data set as $p^i(x_1^i, x_2^i, \dots, x_D^i)$, the construction procedure of an IT can be summarized as follows:

- (1) Put all points into a root node.
- (2) Randomly select a node which has not been partitioned yet.
- (3) Randomly select a dimension $1 \leq d \leq D$.
- (4) Randomly set a split value $\min(x_d^i) < x < \max(x_d^i)$ where i runs over all points in this node.
- (5) Generate two children nodes, put the points with $x_d < x$ into the left child, and the others into the right child.
- (6) Repeat (2) to (5) until every node is either partitioned or is filled with only one point.

In this paper, we do not set a maximum depth for the ITs. When an IT is generated, the path length from a leaf to the root node can be used to determine whether the point represented by the leaf is an anomaly. The path lengths of anomalies are generally shorter than those of normal points.

Because an IT is constructed randomly, it can be expected that the path lengths of points are not stable for a single IT. Therefore, it will be more convincing to introduce multiple ITs, together as an IF. Then the average path lengths over the ITs can be used to discriminate the anomalies from the normal points.

There are only two variables in this algorithm; the number of ITs, and the size of the data set. As will be shown later, the two variables can be made irrelevant of the problem. More details and extensions of the IF algorithm can be found in Refs. [23,24].

III. APPLICATION OF IF ALGORITHM ON THE SEARCH OF AQGCs

The IF algorithm can be applied in many different NP models. In the absence of clear signs for NP, we use the detection of aQGCs as an example.

A. aQGC signals in the process $pp \rightarrow jj\ell^+\ell^-\nu\bar{\nu}$

The vector-boson scattering (VBS) processes at the LHC are very suitable for searching for the existence of aQGCs [4,25]. They have been extensively studied by both the ATLAS group and the CMS group, and the effort will continue with future runs of the LHC. After the first evidence of VBS processes at the LHC found in 2014 [26], a number of experimental results of VBS processes have been obtained [14,27].

Recently, the evidence of exclusive or quasiexclusive $\gamma\gamma \rightarrow W^+W^-$ processes have been found [28]. As an illustration, we concentrate on this process at $\sqrt{s} = 13$ TeV. The next to leading order QCD corrections to the process $pp \rightarrow W^+W^-jj$ have been computed [29], and the K factor is found to be close to one ($K \approx 0.98$). There are some difficulties in the phenomenological studies of NP in this process because of the presence of two neutrinos in $\ell^+\ell^-\nu\bar{\nu}jj$ which makes the reconstruction of the two W bosons almost impossible. However, these difficulties just provide a good test for the IF algorithm.

The Lagrangian relevant to this process is $\mathcal{L}_{\text{aQGC}} = \sum_i (f_{M_i}/\Lambda^4) O_{M,i} + \sum_j (f_{T_j}/\Lambda^4) O_{T,j}$ with [6]

$$\begin{aligned}
 O_{M,0} &= \text{Tr}[\widehat{W}_{\mu\nu}\widehat{W}^{\mu\nu}] \times [(D^\beta\Phi)^\dagger D^\beta\Phi], & O_{T,0} &= \text{Tr}[\widehat{W}_{\mu\nu}\widehat{W}^{\mu\nu}] \times \text{Tr}[\widehat{W}_{\alpha\beta}\widehat{W}^{\alpha\beta}], \\
 O_{M,1} &= \text{Tr}[\widehat{W}_{\mu\nu}\widehat{W}^{\nu\beta}] \times [(D^\beta\Phi)^\dagger D^\mu\Phi], & O_{T,1} &= \text{Tr}[\widehat{W}_{\alpha\nu}\widehat{W}^{\mu\beta}] \times \text{Tr}[\widehat{W}_{\mu\beta}\widehat{W}^{\alpha\nu}], \\
 O_{M,2} &= [B_{\mu\nu}B^{\mu\nu}] \times [(D^\beta\Phi)^\dagger D^\beta\Phi], & O_{T,2} &= \text{Tr}[\widehat{W}_{\alpha\mu}\widehat{W}^{\mu\beta}] \times \text{Tr}[\widehat{W}_{\beta\nu}\widehat{W}^{\nu\alpha}], \\
 O_{M,3} &= [B_{\mu\nu}B^{\nu\beta}] \times [(D^\beta\Phi)^\dagger D^\mu\Phi], & O_{T,5} &= \text{Tr}[\widehat{W}_{\mu\nu}\widehat{W}^{\mu\nu}] \times B_{\alpha\beta}B^{\alpha\beta}, \\
 O_{M,4} &= [(D_\mu\Phi)^\dagger \widehat{W}_{\beta\nu}D_\nu\Phi] \times B^{\beta\nu}, & O_{T,6} &= \text{Tr}[\widehat{W}_{\alpha\nu}\widehat{W}^{\mu\beta}] \times B_{\mu\beta}B^{\alpha\nu}, \\
 O_{M,5} &= [(D_\mu\Phi)^\dagger \widehat{W}_{\beta\nu}D_\nu\Phi] \times B^{\beta\mu} + \text{H.c.}, & O_{T,7} &= \text{Tr}[\widehat{W}_{\alpha\mu}\widehat{W}^{\mu\beta}] \times B_{\beta\nu}B^{\nu\alpha}, \\
 O_{M,7} &= (D_\mu\Phi)^\dagger \widehat{W}_{\beta\nu}\widehat{W}_{\beta\mu}D_\nu\Phi. & &
 \end{aligned} \tag{1}$$

The subprocess $\gamma\gamma \rightarrow W^+W^-$ can be affected by the aQGCs via five vertices, they are $\mathcal{L}_{\gamma\gamma WW} = \sum_{i=0}^4 \alpha_i V_{\gamma\gamma WW_i}$ with

$$\begin{aligned}
 V_0 &= F_{\mu\nu}F^{\mu\nu}W^{+\alpha}W_{\alpha}^-, & V_1 &= F_{\mu\nu}F^{\mu\alpha}W^{+\nu}W_{\alpha}^-, \\
 V_2 &= F_{\mu\nu}F^{\mu\nu}W_{\alpha\beta}^+W^{-\alpha\beta}, & V_3 &= F_{\mu\nu}F^{\nu\alpha}W_{\alpha\beta}^+W^{-\beta\mu}, \\
 V_4 &= F_{\mu\nu}F^{\alpha\beta}W_{\mu\nu}^+W^{-\alpha\beta}, & &
 \end{aligned} \tag{2}$$

where $W^{\pm\mu\nu} \equiv \partial_\mu W_\nu^\pm - \partial_\nu W_\mu^\pm$. The corresponding coefficients of vertices are

$$\begin{aligned}
 \alpha_0 &= \frac{e^2 v^2}{8\Lambda^4} \left(f_{M_0} + \frac{c_W}{s_W} f_{M_4} + 2 \frac{c_W^2}{s_W^2} f_{M_2} \right), \\
 \alpha_1 &= \frac{e^2 v^2}{8\Lambda^4} \left(\frac{1}{2} f_{M_7} + 2 \frac{c_W}{s_W} f_{M_5} - f_{M_1} - 2 \frac{c_W^2}{s_W^2} f_{M_3} \right), \\
 \alpha_2 &= \frac{1}{\Lambda^4} (s_W^2 f_{T_0} + c_W^2 f_{T_5}), & \alpha_3 &= \frac{1}{\Lambda^4} (s_W^2 f_{T_2} + c_W^2 f_{T_7}), \\
 \alpha_4 &= \frac{1}{\Lambda^4} (s_W^2 f_{T_1} + c_W^2 f_{T_6}). & &
 \end{aligned} \tag{3}$$

Because each dimension-eight operator contributes to only one vertex, and because the constraints on dimension-eight operators are obtained by assuming one operator at a time in experiments, the constraints on α_i can be derived by the constraints on dimension-eight operators [8] and are listed in Table I.

TABLE I. The constraints on vertices and the corresponding limits on the dimension-eight operators at 95% CL.

Vertex	Constraint	Coefficient	Constraint
$\alpha_0(\text{TeV}^{-2})$	$[-0.013, 0.013]$	$f_{M_2}/\Lambda^4(\text{TeV}^{-4})$	$[-2.8, 2.8]$ [30]
$\alpha_1(\text{TeV}^{-2})$	$[-0.021, 0.021]$	$f_{M_5}/\Lambda^4(\text{TeV}^{-4})$	$[-8.3, 8.3]$ [30]
$\alpha_2(\text{TeV}^{-4})$	$[-0.38, 0.38]$	$f_{T_5}/\Lambda^4(\text{TeV}^{-4})$	$[-0.5, 0.5]$ [30]
$\alpha_3(\text{TeV}^{-4})$	$[-1.47, 1.69]$	$f_{T_7}/\Lambda^4(\text{TeV}^{-4})$	$[-1.91, 2.12]$ [31]
$\alpha_4(\text{TeV}^{-4})$	$[-0.95, 0.97]$	$f_{T_6}/\Lambda^4(\text{TeV}^{-4})$	$[-1.23, 1.26]$ [31]

B. Detection of the signals

In this subsection, we assume the existence of the aQGCs and investigate whether the signal events can be picked out by the IF algorithm. The dominant signal is W^+W^-jj production induced by aQGCs with leptonic decays of W^\pm bosons as shown in Fig. 1. (a). This process can also be contributed to by the triboson production induced by aQGCs shown in Fig. 1. (b). The background is the process $pp \rightarrow jj\ell^+\ell^-\nu\bar{\nu}$ in the SM; the typical diagrams are shown in Fig. 2. (a). Except for that, we also consider the $t\bar{t}$ production with b-jet mistagged as depicted in Fig. 2. (b); the b-tag efficiency is assumed to be 77% [32]. For simplicity, we neglect the triboson channel induced by aQGCs and the interference between the contributions from aQGCs and the SM which were found to be negligible [8]. In the following we consider one operator at a time, therefore the interferences between different aQGCs are also neglected.

The events are generated by using Monte Carlo (MC) simulation with MadGraph5_aMC@NLO [33], including a parton shower with PYTHIA82 [34] and a CMS-like detector simulation with DELPHES [35]. The basic cuts are set as same as the default settings of MadGraph5_aMC@NLO. The parton distribution function is NNPDF2.3 [36].

To ensure reliability, we require the particles in the final states to satisfy $N_{\ell^\pm} \geq 1, 2 \leq N_j \leq 5$ where N_{ℓ^\pm} are the numbers of (anti)leptons, N_j is the number of jets. This requirement is denoted as the $N_{\ell,j}$ cut. The cross sections of the signals and backgrounds after this cut are listed in Table II. For illustration, we concentrate on $V_{0,3}$ vertices which originate from O_{M_i} and O_{T_i} operators, respectively. Denoting $N_{\text{SM},\bar{t}\bar{t},\text{aQGC}}$ as event numbers of the SM background, $t\bar{t}$ background, and the signal, we generate the events in the ratio $N_{\text{SM}}:N_{\bar{t}\bar{t}}:N_{\text{aQGC}} = \sigma_{\text{SM}}:\sigma_{\bar{t}\bar{t}}:\sigma_{\text{aQGC}}$, where $\sigma_{\text{SM},\bar{t}\bar{t},\text{aQGC}}$ are cross sections of the SM background, $t\bar{t}$ background, and the signal, respectively. For the signals, we keep $N_{\text{aQGC}} = 50$ after the $N_{\ell,j}$ cut, and therefore the data sets consist of events with $N_{\text{SM}}:N_{\bar{t}\bar{t}}:N_{V_0} = 16846:23654:50$ and $N_{\text{SM}}:N_{\bar{t}\bar{t}}:N_{V_3} = 37390:52500:50$. Each event in the data set is assembled straightforwardly

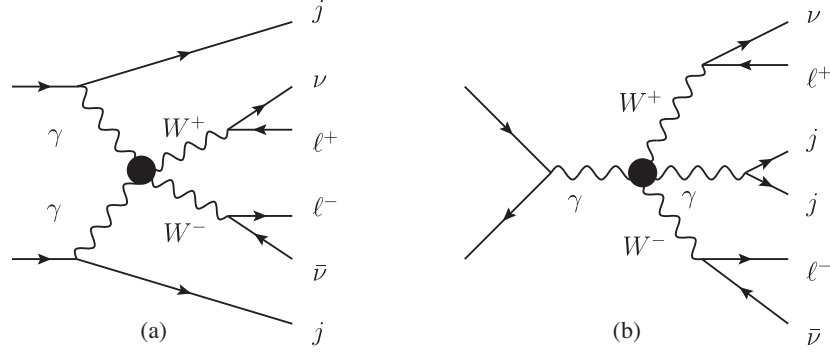


FIG. 1. Typical Feynman diagrams for the signal.

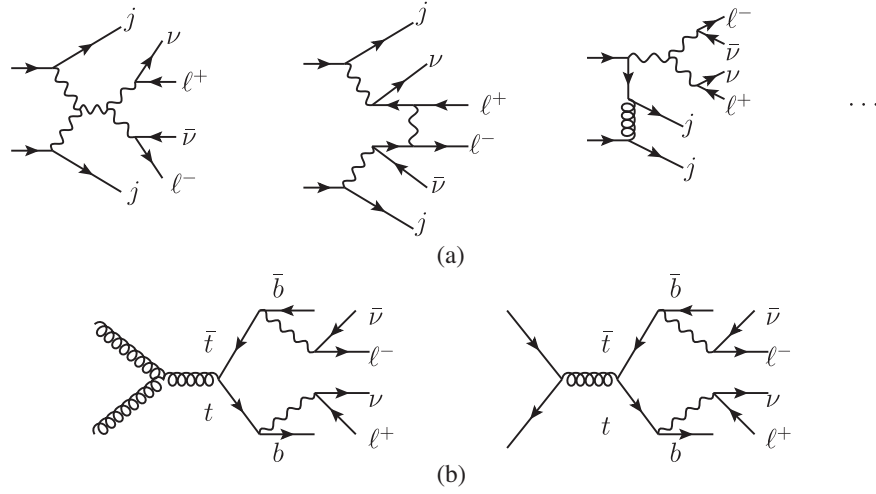


FIG. 2. Typical Feynman diagrams for backgrounds.

and consists of 18 attributes, which are components of the transverse missing momentum \cancel{p}_T , the four-momenta of the hardest two jets p_{j_1} and p_{j_2} , and the four-momenta of the hardest (anti)lepton p_{ℓ^+} and p_{ℓ^-} .

There are two parameters in the IF algorithm. One of the parameters is the number of trees, which is denoted as n . n is a model-independent parameter used to control the accuracy of the IF algorithm. Denoting the path lengths as L , we find that L converges quickly with growing n . Picking one event out of each of the SM background, $t\bar{t}$ background and V_0 signal, as shown in Fig. 3, \bar{L} becomes stable after constructing about 1000 trees. In this paper, we use $n = 2000$; the relative standard errors of L are about 1% (0.4%–1.4%) for each point.

The other parameter is the size of the data set. An anomaly score (denoted as a) which is independent of the size of the data set can be defined by normalizing the average path length (denoted as \bar{L}) with the average depth of an isolation tree $c(N)$ as $a = 2^{-\bar{L}/c(N)}$, where N is the size of the data set. [23],

$$c(N) = 2H(N-1) - 2(N-1)/N, \quad (4)$$

where $H(N)$ is the harmonic number. a is bounded in $(0,1)$; when a is larger, the corresponding event is more suspicious of anomalies.

The normalized distributions of a are shown in Fig. 4. We find that in both cases of V_0 and V_3 , a for the backgrounds are very different from those for the signals.

 TABLE II. The cross sections after $N_{\ell,j}$ cut.

	$\alpha_0 = 0.013$ (TeV ⁻²)	$\alpha_1 = 0.021$ (TeV ⁻²)	$\alpha_2 = 0.38$ (TeV ⁻⁴)	$\alpha_3 = 1.69$ (TeV ⁻⁴)	$\alpha_4 = 0.97$ (TeV ⁻⁴)	SM	$t\bar{t}$
$\sigma(\text{fb})$ after $N_{\ell,j}$ cut	0.91	0.16	0.26	0.41	0.20	306.6	430.5

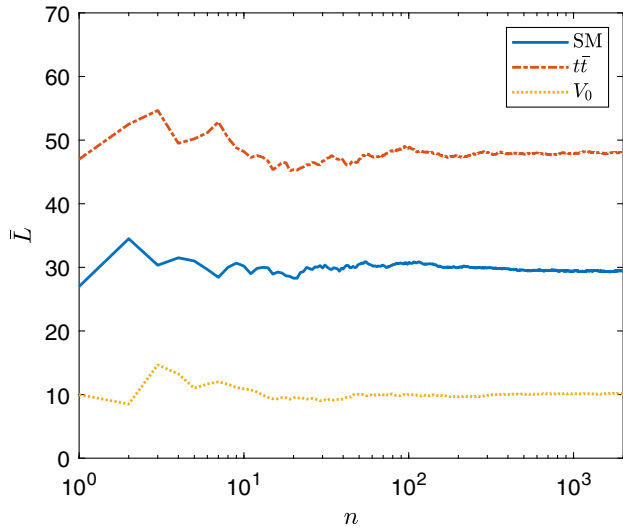


FIG. 3. \bar{L} as a function of n for V_0 data set. Three events from different sources are picked randomly as examples. \bar{L} is the average path length, n is the number of isolation trees.

One can set a minimal anomaly score, and use $a > a_{\min}$ to pick out the signal events of aQGCs. The compositions of the selected events are shown in Fig. 5. For both cases, with $a_{\min} = 0.6$, about half of the selected events are signal events. We find that the IF algorithm is powerful enough to pick out the signal events without the knowledge of the NP as long as the signal exists.

C. Using the IF algorithm as an event selection strategy

The effect of the IF algorithm is similar to an event selection strategy. Different from the traditional ESS, for the IF algorithm there is no need to study the kinematic features. The IF algorithm is like an automatic ESS which can be generally applied for a large class of NP signals.

In the search of NP, the signal significance is widely used and is defined as $\mathcal{S}_{stat} = N_s / \sqrt{N_{bg} + N_s}$ where $N_{s,bg}$ are the event numbers of signal and background. Similarly, a luminosity-independent quantity can be defined as $\hat{\mathcal{S}}_{stat} = \sigma_s / \sqrt{\sigma_{bg} + \sigma_s}$ such that $\mathcal{S}_{stat} = \sqrt{l} \hat{\mathcal{S}}_{stat}$ where l is the

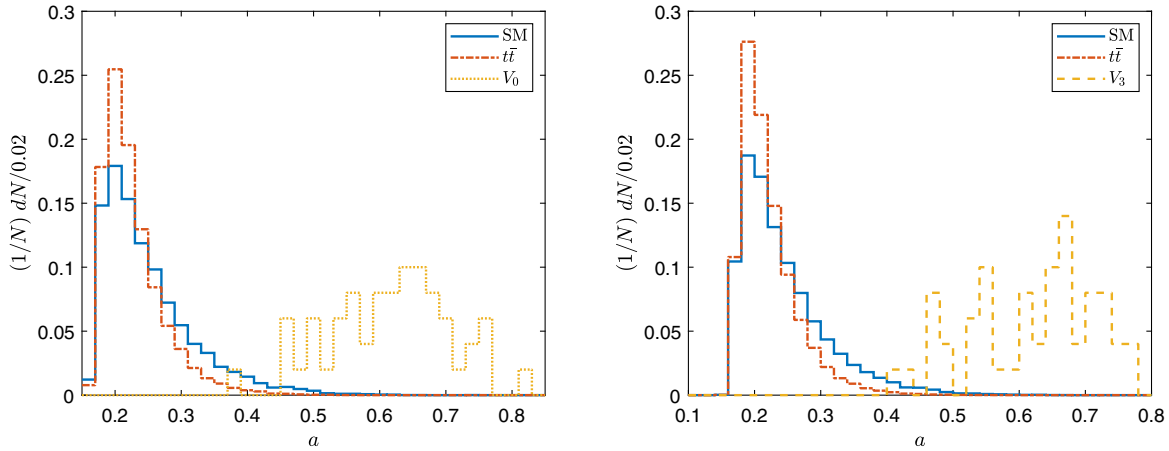


FIG. 4. Normalized distributions of a , the left panel is for V_0 and the right panel is for V_3 .

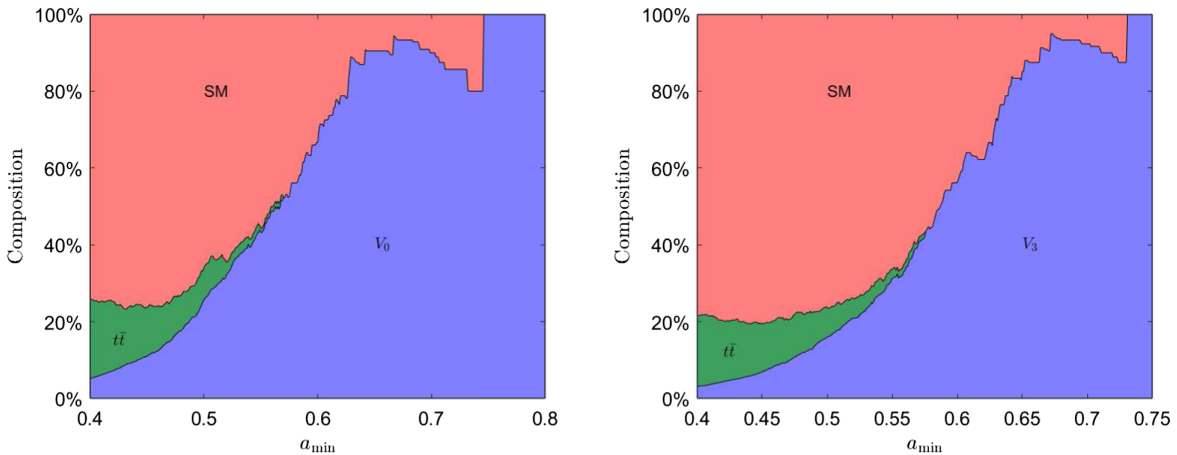


FIG. 5. Compositions of the selected events passing a cut on anomaly score $a > a_{\min}$, the left panel is for V_0 and the right panel is for V_3 .

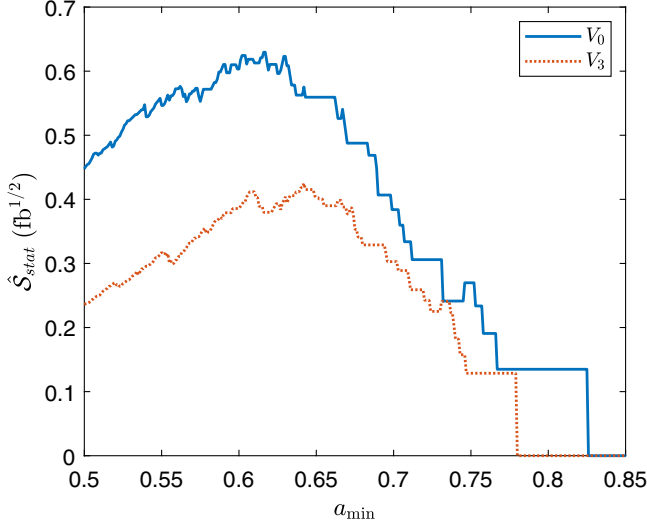
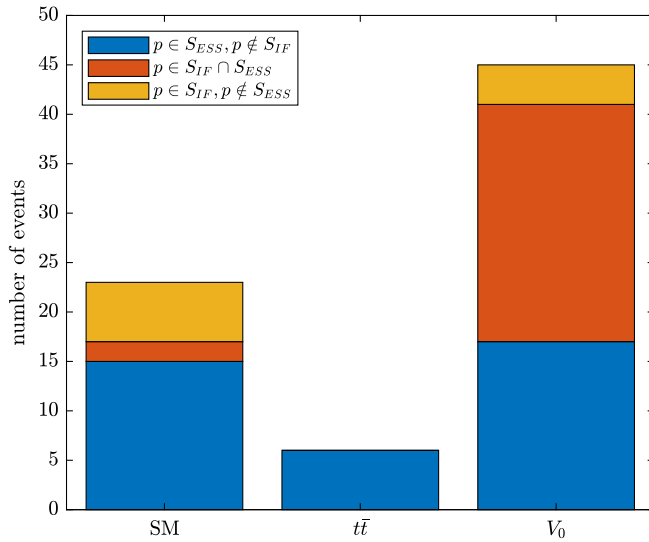


FIG. 6. $\hat{S}_{stat} = \sigma_s / \sqrt{\sigma_{bg} + \sigma_s}$ as functions of a_{min} , where $\sigma_{s,bg}$ are the cross sections of aQGC contribution and backgrounds after a cut on the anomaly score $a > a_{min}$.

luminosity. In this paper, we use \hat{S}_{stat} to qualify the ESS. By selecting events with $a > a_{min}$, \hat{S}_{stat} for $V_{0,3}$ are shown in Fig. 6. The \hat{S}_{stat} can reach $0.630 \text{ fb}^{1/2}$ at $a_{min} = 0.617$ for V_0 and $0.425 \text{ fb}^{1/2}$ at $a_{min} = 0.642$ for V_3 .

We compare the IF algorithm with the ESS designed for the aQGCs in the process $pp \rightarrow jj\ell^+\ell^-\nu\bar{\nu}$ proposed in Ref. [8], which are

$$M_{jj} > 150 \text{ GeV}, \quad \Delta y_{jj} > 1.2, \quad |\cos(\phi_{LM})| > 0.3, \\ \cos(\theta_{\ell\ell}) < 0, \quad \hat{s} > 1.5 \text{ TeV}^2, \quad M_{o1} > 600 \text{ GeV}, \quad (5)$$



where M_{jj} and Δy_{jj} are the invariant mass and difference between the rapidities of the hardest two jets, ϕ_{LM} is the angle between sum of the transverse momenta of charged leptons $\mathbf{p}_T^{\ell^+} + \mathbf{p}_T^{\ell^-}$ and $\not{\mathbf{p}}_T$, $\theta_{\ell\ell}$ is the angle between the charged leptons, and

$$M_{o1} \equiv \sqrt{(|\mathbf{p}_T^{\ell^+}| + |\mathbf{p}_T^{\ell^-}| + |\not{\mathbf{p}}_T|)^2 - |\mathbf{p}_T^{\ell^+} + \mathbf{p}_T^{\ell^-} + \not{\mathbf{p}}_T|^2}, \\ \hat{s} = ((1+|u|)E^{\ell^+} + (1+|v|)E^{\ell^-})^2 \\ - ((1+u)\mathbf{p}_z^{\ell^+} + (1+v)\mathbf{p}_z^{\ell^-})^2 - \left| \sum_{\pm} \mathbf{p}_T^{\ell^{\pm}} + \not{\mathbf{p}}_T \right|^2, \quad (6)$$

with $E^{\ell^{\pm}}$ the energies of charged leptons, and

$$u = \frac{1}{\kappa} (\not{p}_y \mathbf{p}_x^{\ell^-} - \not{p}_x \mathbf{p}_y^{\ell^-}), \quad v = -\frac{1}{\kappa} (\not{p}_y \mathbf{p}_x^{\ell^+} - \not{p}_x \mathbf{p}_y^{\ell^+}), \\ \kappa = \mathbf{p}_y^{\ell^+} \mathbf{p}_x^{\ell^-} - \mathbf{p}_x^{\ell^+} \mathbf{p}_y^{\ell^-}. \quad (7)$$

For the IF algorithm, we select events with $a > 0.617$ and $a > 0.642$ for V_0 and V_3 , respectively, the result sets are denoted as S_{IF} . The sets consisting of events selected by Eq. (5) are denoted as S_{ESS} . The numbers of events in those sets are shown in Fig. 7. As one can see, the events picked by the IF algorithm are not quite the same as the ESS in Eq. (5), especially for the backgrounds. The \hat{S}_{stat} for $V_{0,3}$ with Eq. (5) are $0.691 \text{ fb}^{1/2}$ and $0.341 \text{ fb}^{1/2}$. Compared with the results of the IF algorithm, we find that the ESS using anomaly scores shows competitive ability in discriminating signals, especially for the cases where the signal events are fewer.

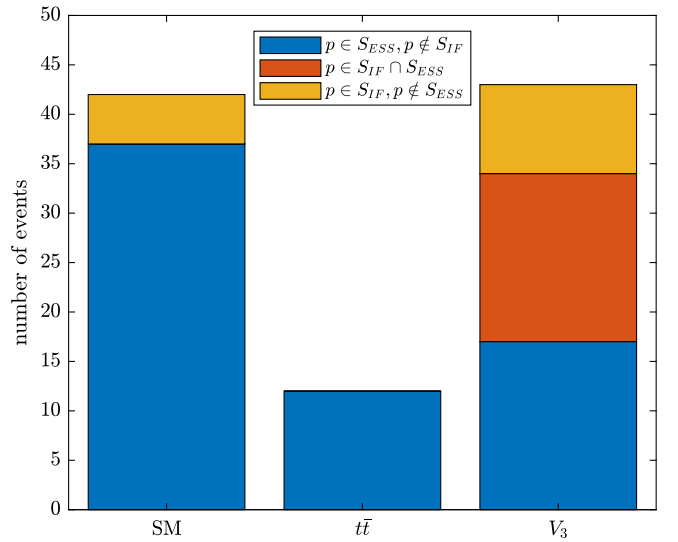


FIG. 7. Difference between the selected events by using Eq. (5) and using anomaly scores. The blue bars show the number of events selected by Eq. (5) but not by anomaly score cuts, the yellow bars show the number of events selected by anomaly score cuts but not by Eq. (5), the orange bars show the number of events selected by both methods.

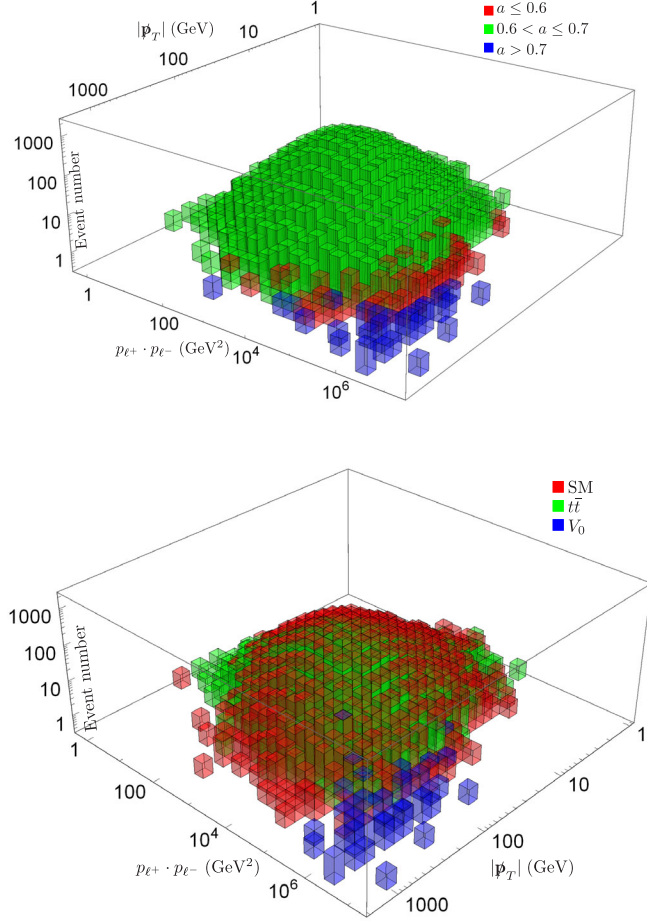


FIG. 8. Distributions of events in the $(|\not{p}_T|, p_{\ell^+} \cdot p_{\ell^-})$ space. There 31×31 bins. The widths are $[\exp(0.251(k-1)), \exp(0.251k)]$ (GeV) for $|\not{p}_T|$ dimension and $[\exp(0.558(k-1)), \exp(0.558k)]$ (GeV^2) for $p_{\ell^+} \cdot p_{\ell^-}$ dimension, where k is an integer with $0 \leq k \leq 30$.

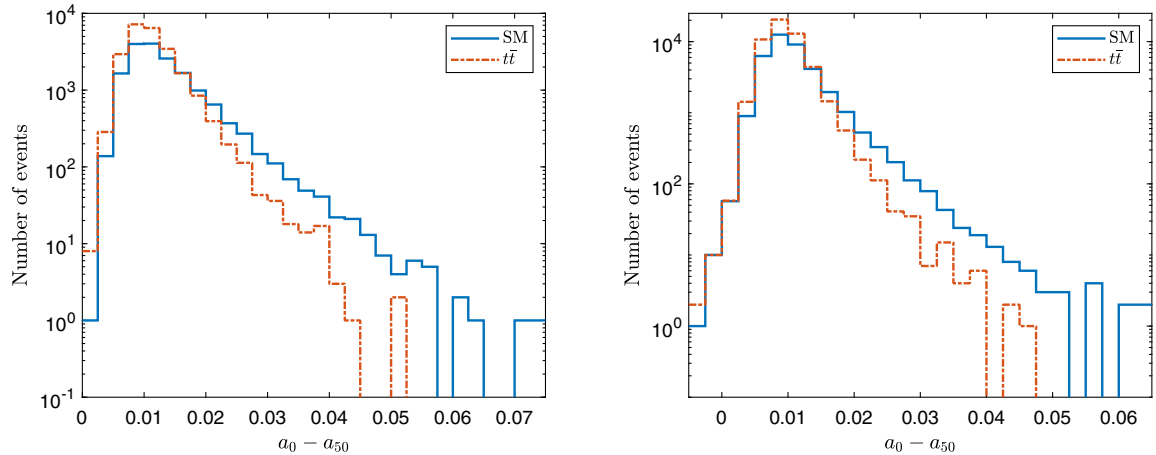


FIG. 9. The distributions of $a_0 - a_{50}$ for the backgrounds, where $a_0 - a_{50}$ are the changes of the anomaly scores from a data set without signal events to a data set with 50 signal events.

In the above we used 18 attributes which is straightforward, but not optimized. There are usually observables more sensitive to the signal, which depend on the model or operators one looks for. For example, knowing that we are searching for aQGCs, we can use attributes such as M_{o1} , $|\not{p}_T|$ and $p_{\ell^+} \cdot p_{\ell^-}$. By choosing only two attributes, $|\not{p}_T|$ and $p_{\ell^+} \cdot p_{\ell^-}$, the events can be represented by points in a 2D space, and therefore are easy to visualize. By applying the IF algorithm on these attributes, the distributions of events with different anomaly scores are shown in Fig. 8, and one can see that the events with the higher anomaly scores are indeed those events far away from the others. The distributions of the events from backgrounds and signal are also shown in Fig. 8, which indicate that the events far away from the others are indeed the signal events.

D. Setting constraints on the coefficients

A more common scenario is that signal events are not observed and one needs to set constraints on the parameters of NP models and the coefficients of operators. This can also be done with the help of the IF algorithm, because the mechanism behind the IF algorithm suggests that the anomaly scores of the backgrounds should not be sensitive to the signal events. Consequently, after constructing an IF for the MC data of the backgrounds, which is model independent, one can use anomaly scores as a cut. The expected cross section after this cut can be calculated, and can be compared with the cross section obtained by experiments under the same cut. However, when it comes to constraining the parameters of a specific model, we need the information of this model which is not model independent any more.

Again, we take $V_{0,3}$ vertices as examples to illustrate this approach. To set constraints on the coefficients, the data sets are assembled with $N_{\text{SM}}:N_{\bar{t}t}:N_{\text{aQGC}}$ with $N_{\text{SM}}, N_{\bar{t}t}$ as same as the previous section, and with $N_{\text{aQGC}} = 0, 10, 20$,

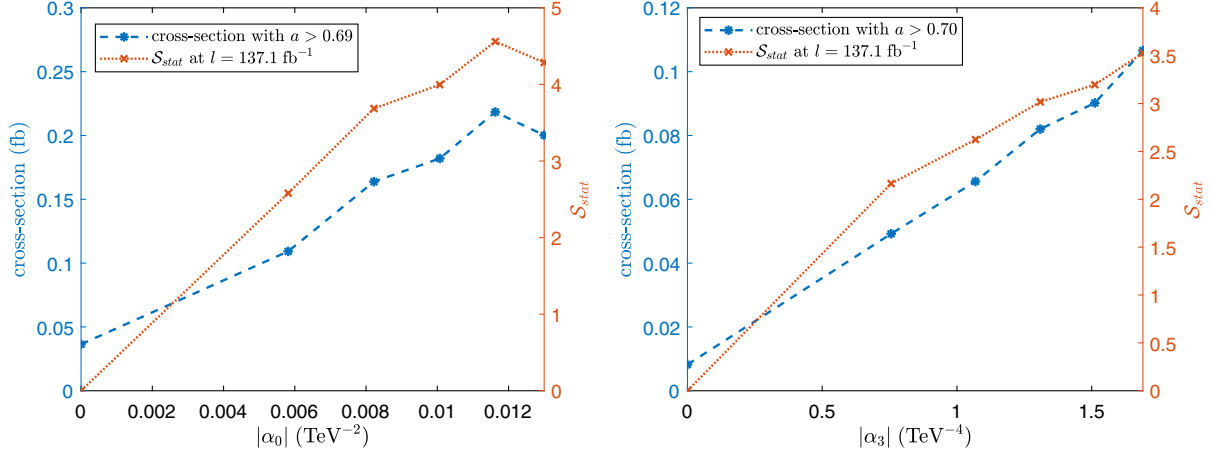


FIG. 10. The cross sections and signal significances after the anomaly score cut as functions of $\alpha_{0,3}$.

30, 40, 50. They correspond to $\alpha_0 = \sqrt{m/5} \times 0.013 \text{ TeV}^{-2}$ and $\alpha_3 = \sqrt{m/5} \times 1.69 \text{ TeV}^{-4}$ with $m = 0, 1, \dots, 5$ when neglecting the interferences.

Denoting $a_{0,50}$ as anomaly scores of events for the $N_{\text{aQGC}} = 0$ and $N_{\text{aQGC}} = 50$ data sets, respectively. The distributions of $a_0 - a_{50}$ for the backgrounds are shown in Fig. 9. We find that, the anomaly scores of the backgrounds increase a little bit without the signal events. For V_0 , $0 < a_0 - a_{50} < 0.075$ and for V_3 , $-0.005 < a_0 - a_{50} < 0.065$.

Since the anomaly scores for the backgrounds increase a little, we use $a_{\min} = 0.68$ for V_0 as a cut, and $a_{\min} = 0.70$ for V_3 . The cross sections after this cut are shown in Fig. 10. Using the cross sections, one can obtain the signal significance, which is $\mathcal{S}_{\text{stat}} = \sqrt{l} \times (\sigma(\alpha_i) - \sigma(\alpha_i = 0)) / \sqrt{\sigma(\alpha_i)}$, where $\sigma(\alpha_i)$ is the cross section at α_i after the anomaly score cut. The integrated luminosity currently at the LHC at $\sqrt{s} = 13 \text{ TeV}$ is about $l = 137.1 \text{ fb}^{-1}$ [37]. The signal significances at $l = 137.1 \text{ fb}^{-1}$ are also shown in Fig. 10. If $|\alpha_0| = 0.0058$, the signal of V_0 should be observed with $\mathcal{S}_{\text{stat}} = 2.58$, therefore if the signal is not observed with $\mathcal{S}_{\text{stat}} > 2$, the constraint is $|\alpha_0| < 0.0058 \text{ TeV}^{-2}$. Similarly, for α_3 the expected constraint is $|\alpha_3| < 1.07 \text{ TeV}^{-4}$. The result of the IF algorithm is better compared with the expected constraint at $\mathcal{S}_{\text{stat}} > 2$ in Ref. [8] which is $\alpha_0 \in [-0.0071, 0.0069] \text{ TeV}^{-2}$ and $\alpha_3 \in [-1.73, 1.30] \text{ TeV}^{-4}$. Again, for the cases that signal events are fewer, the IF algorithm performs better.

IV. SUMMARY

As more and more data are collected at the colliders, it becomes increasingly important to simplify the search of NP signals. In this paper, we investigate a model independent approach for searching for the NP signals which exploits the characteristics of the NP signals; i.e., few and kinematically different. We use an unsupervised ML algorithm, also known as the IF algorithm, to find out the kinematically unusual events directly.

The IF algorithm is transparent and easy to apply. This approach has the advantage that the suspected signals of the NP can be picked out without the knowledge of the NP models. It works as an automatic ESS which can be generally applied. We also show that the IF algorithm can be applied to constrain the parameters of NP models and the coefficients of the operators. Apart from that, the kinematically unusual events picked out are always worth studying. There are also some limitations in this approach. When anomalies appear, one needs to look deeper into them to know where they originated. Beyond that, there is room for improvement in how the data is organized. In this paper, we directly use the components of the four-momenta of the particles in the final state.

We use the dimension-eight operators contributing to the aQGCs as examples to investigate the capabilities of this approach. The process $pp \rightarrow jj\ell^+\ell^-\nu\bar{\nu}$ is chosen as an arena, which has some complexity due to the neutrinos in the final state. It can be shown that the anomaly scores of the background events are generally smaller than those of the signal events. With a minimal allowed anomaly score as a cut, the signal events can be selected efficiently. The IF algorithm shows greater ability to highlight the signal events and constrain the coefficients of the operators compared with the ESS designed for the aQGCs in this process. In addition, we also show that the IF algorithm performs better with fewer signal events. The IF algorithm or other machine learning methods can be a very promising tool in the future study of high energy physics.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grants No. 11905093 and No. 12047570, the Natural Science Foundation of the Liaoning Scientific Committee No. 2019-BS-154 and the Outstanding Research Cultivation Program of Liaoning Normal University (No. 21GDL004).

- [1] J. Ellis, *Phil. Trans. R. Soc. A* **370**, 818 (2012).
- [2] O. Cremonesi, [arXiv:1002.1437](https://arxiv.org/abs/1002.1437); A. de Gouvea, A. Friedland, P. Huber, and I. Mocioiu, [arXiv:1309.7338](https://arxiv.org/abs/1309.7338); G. W. Bennett *et al.* (Muon $g-2$ Collaboration), *Phys. Rev. D* **73**, 072003 (2006); S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *J. High Energy Phys.* **06** (2016) 092; R. Aaij *et al.* (LHCb Collaboration), *Phys. Rev. Lett.* **113**, 151601 (2014); M. Huschle *et al.* (Belle Collaboration), *Phys. Rev. D* **92**, 072014 (2015).
- [3] S. Weinberg, *Phys. Rev. Lett.* **43**, 1566 (1979); B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *J. High Energy Phys.* **10** (2010) 085; S. Willenbrock and C. Zhang, *Annu. Rev. Nucl. Part. Sci.* **64**, 83 (2014); E. Masso, *J. High Energy Phys.* **10** (2014) 128; S. Willenbrock and C. Zhang, *Annu. Rev. Nucl. Part. Sci.* **64**, 83 (2014).
- [4] C. Anders *et al.*, *Rev. Phys.* **3**, 44 (2018).
- [5] B. Henning, X. Lu, T. Melia, and H. Murayama, *J. High Energy Phys.* **08** (2017) 016; **09** (2019) 19.
- [6] O. J. P. Eboli, M. C. Gonzalez-Garcia, and J. K. Mizukoshi, *Phys. Rev. D* **74**, 073005 (2006); O. J. P. Éboli and M. C. Gonzalez-Garcia, *Phys. Rev. D* **93**, 093013 (2016).
- [7] C. Zhang and S.-Y. Zhou, *Phys. Rev. D* **100**, 095003 (2019); Q. Bi, C. Zhang, and S.-Y. Zhou, *J. High Energy Phys.* **06** (2019) 137; Y.-C. Guo, Y.-Y. Wang, J.-C. Yang, and C.-X. Yue, *Chin. Phys. C* **44**, 123105 (2020); J.-C. Yang, Y.-C. Guo, C.-X. Yue, and Q. Fu, [arXiv:2107.01123](https://arxiv.org/abs/2107.01123).
- [8] Y.-C. Guo, Y.-Y. Wang, and J.-C. Yang, *Nucl. Phys.* **B961**, 115222 (2020).
- [9] M. Born and L. Infeld, *Proc. R. Soc. A* **144**, 425 (1934); J. Ellis and S.-F. Ge, *Phys. Rev. Lett.* **121**, 041801 (2018).
- [10] D. Espriu and F. Mescia, *Phys. Rev. D* **90**, 015035 (2014); R. Delgado, A. Dobado, M. Herrero, and J. Sanz-Cillero, *J. High Energy Phys.* **07** (2014) 149; S. Fichet and G. von Gersdorff, *J. High Energy Phys.* **03** (2014) 102; T. Lee, *Phys. Rev. D* **8**, 1226 (1973); J.-C. Yang and M.-Z. Yang, *Mod. Phys. Lett. A* **31**, 1650012 (2016); X.-G. He, G. C. Joshi, H. Lew, and R. R. Volkas, *Phys. Rev. D* **44**, 2118 (1991); J.-X. Hou and C.-X. Yue, *Eur. Phys. J. C* **79**, 983 (2019); K. Mimasu and V. Sanz, *J. High Energy Phys.* **06** (2015) 173; C.-X. Yue, M.-Z. Liu, and Y.-C. Guo, *Phys. Rev. D* **100**, 015020 (2019).
- [11] C. Zhang and S.-Y. Zhou, *Phys. Rev. Lett.* **125**, 201601 (2020).
- [12] C. Degrande, *J. High Energy Phys.* **02** (2014) 101.
- [13] J. Ellis, S.-F. Ge, H.-J. He, and R.-Q. Xiao, *Chin. Phys. C* **44**, 063106 (2020); J. Ellis, H.-J. He, and R.-Q. Xiao, *Sci. China Phys. Mech. Astron.* **64**, 221062 (2021); A. Senol, H. Denizli, A. Yilmaz, I. Turk Cakir, K.Y. Oyulmaz, O. Karadeniz, and O. Cakir, *Nucl. Phys.* **B935**, 365 (2018); Q. Fu, Y.-C. Guo, and J.-C. Yang, [arXiv:2102.03623](https://arxiv.org/abs/2102.03623).
- [14] G. Perez, M. Sekulla, and D. Zeppenfeld, *Eur. Phys. J. C* **78**, 759 (2018).
- [15] C. Arzt, M. Einhorn, and J. Wudka, *Nucl. Phys.* **B433**, 41 (1995).
- [16] G. Kasieczka *et al.*, [arXiv:2101.08320](https://arxiv.org/abs/2101.08320).
- [17] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, *Nature (London)* **560**, 41 (2018); P. Baldi, P. Sadowski, and D. Whiteson, *Nat. Commun.* **5**, 4308 (2014); J. Ren, L. Wu, J. M. Yang, and J. Zhao, *Nucl. Phys.* **B943**, 114613 (2019); M. Abdughani, J. Ren, L. Wu, and J. M. Yang, *J. High Energy Phys.* **08** (2019) 055; R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, *Phys. Rev. Lett.* **124**, 010508 (2020); J. Ren, L. Wu, and J. M. Yang, *Phys. Lett. B* **802**, 135198 (2020); A. De Simone and T. Jacques, *Eur. Phys. J. C* **79**, 289 (2019); M. A. Md Ali, N. Badrud'din, H. Abdullah, and F. Kemi, *Int. J. Mod. Phys. A* **35**, 2050131 (2020); E. Fol, R. Tomás, J. Coello de Portugal, and G. Franchetti, *Phys. Rev. Accel. Beams* **23**, 102805 (2020); R. T. D'Agnolo and A. Wulzer, *Phys. Rev. D* **99**, 015014 (2019).
- [18] M. Farina, Y. Nakai, and D. Shih, *Phys. Rev. D* **101**, 075021 (2020); O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu, and J.-R. Vlimant, *J. High Energy Phys.* **05** (2019) 036.
- [19] M. Crispim Romão, N. F. Castro, and R. Pedro, *Eur. Phys. J. C* **81**, 27 (2021).
- [20] M. van Beekveld *et al.*, [arXiv:2010.07940](https://arxiv.org/abs/2010.07940).
- [21] M. Kuusela, T. Vatanen, E. Malmi, T. Raiko, T. Aaltonen, and Y. Nagai, *J. Phys. Conf. Ser.* **368**, 012032 (2012).
- [22] L. Ruff *et al.*, in *Proceedings of the 35th International Conference on Machine Learning*, edited by J. Dy and A. Krause, *Proceedings of Machine Learning Research* Vol. 80 (PMLR, Stockholm, Sweden, 2018), pp. 4393–4402.
- [23] F. T. Liu, K. M. Ting, and Z. Zhou, in *2008 Eighth IEEE International Conference on Data Mining (IEEE, 2008)*, pp. 413–422, <https://www.doi.org/10.1109/ICDM.2008.17>.
- [24] P. Karczmarek, A. Kiersztyn, W. Pedrycz, and E. Al, *Knowledge-Based Systems* **195**, 105659 (2020).
- [25] D. R. Green, P. Meade, and M.-A. Pleier, *Rev. Mod. Phys.* **89**, 035008 (2017); A. Alboteanu, W. Kilian, and J. Reuter, *J. High Energy Phys.* **11** (2008) 010; W. Kilian, T. Ohl, J. Reuter, and M. Sekulla, *Phys. Rev. D* **91**, 096007 (2015); S. Tizchang and S. M. Etesami, *J. High Energy Phys.* **07** (2020) 191.
- [26] G. Aad *et al.* (ATLAS Collaboration), *Phys. Rev. Lett.* **113**, 141803 (2014).
- [27] M. Aaboud *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **07** (2017) 107; V. Khachatryan *et al.* (CMS Collaboration), *Phys. Lett. B* **770**, 380 (2017); V. Khachatryan *et al.* (CMS Collaboration), *J. High Energy Phys.* **06** (2017) 106; A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Lett. B* **774**, 682 (2017); A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Lett. B* **789**, 19 (2019); M. Aaboud *et al.* (ATLAS Collaboration), *Phys. Lett. B* **793**, 469 (2019); A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Lett. B* **795**, 281 (2019); A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Rev. Lett.* **120**, 081801 (2018); A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Lett. B* **798**, 134985 (2019).
- [28] V. Khachatryan *et al.* (CMS Collaboration), *J. High Energy Phys.* **08** (2016) 119.
- [29] M. Rauch, [arXiv:1610.08420](https://arxiv.org/abs/1610.08420).
- [30] CMS, A. M. Sirunyan *et al.*, *Phys. Lett. B* **811**, 135988 (2020).
- [31] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **06** (2020) 076.
- [32] G. Aad *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **79**, 970 (2019).

- [33] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *J. High Energy Phys.* **07** (2014) 079; N. D. Christensen and C. Duhr, *Comput. Phys. Commun.* **180**, 1614 (2009).
- [34] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *Comput. Phys. Commun.* **191**, 159 (2015).
- [35] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELHI 3 Collaboration), *J. High Energy Phys.* **02** (2014) 057.
- [36] R. D. Ball, V. Bertone, S. Carrazza, L. D. Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo (NNPDF Collaboration), *Nucl. Phys.* **B877**, 290 (2013).
- [37] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **03** (2020) 051.