Rotation-equivariant graph neural network for learning hadronic SMEFT effects

Suman Chatterjee[1,*], Sergio Sánchez Cruz[2,§], Robert Schöfbeck[1,†] and Dennis Schwarz[1,‡]

[1]*Institute for High Energy Physics, Austrian Academy of Sciences,*
Nikolsdorfergasse 18, A-1050 Vienna, Austria
[2]*Department of Physics, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland*

We introduce a graph neural network architecture designed to extract novel phenomena in the standard model effective field theory (SMEFT) context from LHC collision data. The proposed infrared- and collinear-safe architecture is sensitive to the angular orientation of radiation patterns in jets from hadronic decays of highly energetic massive particles. Equivariance with respect to rotations around the jet axis allows for extracting the information on the angular orientation decoupled from the jet substructure. We demonstrate the robustness of the approach and its potential for future probes of the SMEFT at the LHC through toy studies and with realistic event simulations of the associated production of a W and a Z boson in the semileptonic decay channel.

I. INTRODUCTION

The Large Hadron Collider (LHC) is a veritable gold mine of data, among whose most complex signatures are collimated sprays of particles (jets) originating from the hadronic decays of boosted massive particles. This richness presents a challenge and a fertile ground for exploring fundamental physics in hadronic final states.

In recent years, machine learning algorithms have been increasingly adapted to more complex data representations, with the ensuing rise of the input feature dimension tamed by imposed symmetries of the underlying physical problem (see Refs. [1–3] for an overview). In particular, message-passing graph neural networks (gNNs) [4–6] learn from relationships among particles that are interpreted as point clouds with manifest permutation invariance [7–10]. Such developments have markedly improved the tagging performance for the various objects that are reconstructed as highly energetic jets at collider experiments [11]. Incorporating invariance or equivariance for symmetry groups specific to certain tagging challenges refines the algorithms' inductive bias and is an active area of research [12–23]. In particular, infrared and collinear (IRC) safety was integrated into gNNs by means of energy-weighted message passing [24], generalizing energy flow networks [25], and providing robustness to uncertainties in the modeling of the splitting or merging of soft and collinear particles. A rich technological toolkit thus extends the experiments' grasp on lower-level data representations that are particularly relevant for the hadronic final states considered in this work.

In the theoretical domain, a parallel interest in more low-level representations of hypothetical new phenomena is fueled by the absence of any compelling signal of new resonant phenomena. The standard model effective field theory (SMEFT) [26–28] has emerged as the preferred language for describing hypothetical phenomena below an assumed energy scale, which separates the energy scale of the LHC from resonant phenomena at much higher scales. The SMEFT is a powerful framework extending the standard model (SM) Lagrangian by field monomials whose prefactors, the Wilson coefficients, are the parameters of interest in experimental measurements. These give the experimentalist a low-level representation of a wide range of hypothetical high-scale models.

The SMEFT organizing principle is the operators' mass dimension, starting at six for the new physics scenarios relevant at the LHC [28]. Because the lowest-order amplitude modifications are linear in the Wilson coefficients, the cross-section deviations are accurately described by quadratic polynomials within the SMEFT's range of validity. This simple analytic structure, in turn, is exploited in various simulation-based inference techniques [29–39].

*suman.chatterjee@oeaw.ac.at
†robert.schoefbeck@oeaw.ac.at
‡dennis.schwarz@oeaw.ac.at
§sergio.sanchez.cruz@cern.ch
Present address: The European Organization for Nuclear Research (CERN).

The linear term in the polynomial describes the SM-SMEFT interference. It is the only contribution not subject to further contributions from SMEFT operators with a higher mass dimension, and, therefore, the unambiguous harbinger of dimension-6 SMEFT effects. In a wide range of final states involving decays of massive vector bosons, however, a naive experimental analysis may unintentionally remove the sensitivity to the linear contribution. In these cases, a dedicated angular analysis can sometimes "resurrect" the interference terms [40–44]. In such cases, the orientation of the decay planes of the W or Z boson provides crucial sensitivity because it can resolve the amplitudes' helicity configuration which is altered in the SMEFT [40]. Analyses of leptonic final states of diboson processes (VV with V = W, Z, γ, or H) already profit from the boost in experimental sensitivity [45–47] and served as a motivating use case for the development of machine-learned optimal observables [36–39].

The main idea of this paper is to extend this simulation-based inference approach to hadronic final states and extract SMEFT sensitivity via a gNN that is equivariant with respect to azimuthal rotations around the axis of a highly energetic jet, originating from the hadronic decay of a boosted massive particle. The variable-length set of the jet's constituents is the input to the gNN. Its output is fed into dense layers that can also accept other features of the event. The algorithm is tailored towards the linear SM-SMEFT interference and exploits the spatial angular orientation of the radiation patterns with respect to the event remainder. Its main purpose is to learn the score vector associated with the SMEFT Wilson coefficients. It thus provides machine-learned optimal observables for small deviations from the SM. A prototypical situation is sketched in Fig. 1.

The rest of the paper is organized as follows. In Sec. II we describe the subject of our study, the associated production of a W and a Z boson WZ decaying semi-leptonically, and the datasets' structure. In Sec. III, we formulate the algorithm. We elucidate its main features in simple toy studies presented in Sec. IV. The application to simulated WZ events is provided in Sec. V, and we give conclusions and an outlook in Sec. VI.
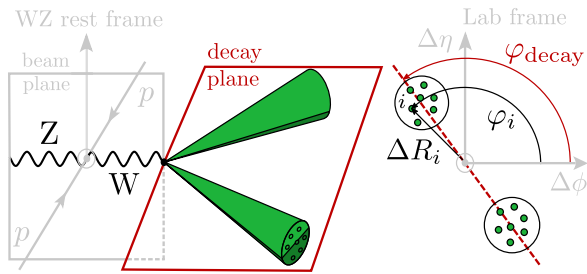

FIG. 1. Sketch of the pp → WZ process, the decay plane, the decay plane angle φ_{decay} of the W boson decay. The ΔR_i and φ_i coordinates of a constituent i from a highly energetic jet are defined with respect to the beam plane and the jet axis.

II. THE SEMILEPTONIC WZ FINAL STATE

Despite the relatively small number of expected events in the high-p_T regime, we can nevertheless probe SMEFT operators as they induce energy-growing corrections to the SM amplitudes [38,39]. Among the diboson final states providing SMEFT sensitivity to LHC data analyses [45–47], we chose the comparably simple case of pp → WZ as a motivating example. We consider the semileptonic decay channel, where Z → $\overline{\ell}\ell$ and W → $\bar{q}$q, and restrict to the highly energetic phase space $p_T(\text{W}) > 300$ GeV, where the W boson is reconstructed as a jet by the anti-k_T [48] algorithm in the FASTJET implementation [49] with a radius parameter of $R = 0.8$ (AK8) of approximately massless jet constituents. We focus on the interference contribution to the differential cross section from the operators

$$\mathcal{O}_{\text{W}} = \varepsilon^{ijk} W_\mu^{i\nu} W_\nu^{j\rho} W_\rho^{i\mu} \quad \text{and}$$
$$\mathcal{O}_{\tilde{\text{W}}} = \varepsilon^{ijk} \tilde{W}_\mu^{i\nu} W_\nu^{j\rho} W_\rho^{i\mu} \tag{1}$$

whose Wilson coefficients we denote by C_{W} and $C_{\tilde{\text{W}}}$. The operators $\mathcal{O}_{\text{W}}$ and $\mathcal{O}_{\tilde{\text{W}}}$ induce *CP*-even and *CP*-odd modifications of the gauge boson self-interactions, while another operator, $\mathcal{O}_{\text{Hq}}^{(3)} = (H^\dagger i \overset{\leftrightarrow}{D}_\mu^i H)(\bar{q}_L \sigma^i \gamma^\mu q_L)$, modifies the light quark–gauge boson vectorlike interaction. At the linear level, these operators exhibit energy growth with the center-of-mass energy of the diboson system, denoted by s. The operator $\mathcal{O}_{\text{Hq}}^{(3)}$ dominantly contributes to the helicity configuration where both bosons are longitudinally polarized and where SM contribution is constant in s at high energy. Therefore, the SM-SMEFT interference term introduces energy growth but does not modify the distributions of the azimuthal decay-plane angles. Therefore, we do not consider it further.

The operators $\mathcal{O}_{\text{W}}$ and $\mathcal{O}_{\tilde{\text{W}}}$, in contrast, contribute energy-growing amplitudes for transverse gauge boson polarization with the same helicities which are small in the SM. Consequently, the interference contribution vanishes in the tail if it is not resurrected by a dedicated angular analysis [38–40]. In the leptonic decay channels, the azimuthal orientation of the decay plane φ_{decay}, as shown in Fig. 1, drives the sensitivity. A dedicated multivariate analysis of the fully leptonic decay mode exploits this fact by extracting SMEFT sensitivity from global event kinematics, i.e., per-event features that are reconstructed from lepton kinematics and in particular angular observables in the diboson rest frame [38,39].

In our case, we attempt to similarly access this sensitivity by extracting it with a graph neural network. Because of the special role of the azimuthal decay plane angle, we construct it equivariantly with respect to rotations of the boosted jet's constituents around the jet axis. The relevant symmetry group is SO(2), irrespective of whether we choose our reference frame as the lab frame or, as done

in Fig. 1, as the rest frame of the WZ system. Our events are instances in a dataset

$$\mathcal{D} = \{\boldsymbol{x}_{\text{global},j}, \{\boldsymbol{x}_{\text{p}}\}_{i=1}^{N_p(j)}\}_{j=1}^{N_{\text{events}}}, \qquad (2)$$

where $\boldsymbol{x}_{\text{global}}$ denotes optional global event features, pertaining to the kinematics of the WZ candidate event, the Z boson, the transverse missing energy, etc. In each event, there is also a list of N_p approximately massless constituent particles of an AK8 jet [48]. These constituents could, for example, be provided by a particle-flow algorithm [50]. Each particle's feature vector $\boldsymbol{x}_{\text{p}}$ contains the four-momentum p_i^μ and, in principle, a number of features which could, e.g., represent the quality of the particle tracks association with the primary collision vertex, its probability of originating from a pileup vertex [51], or the charge of the particle. We emphasize that the dataset in Eq. (2) is hierarchical in the sense that the reference frame for the p_i^μ depends on the event kinematic that also enters via $\boldsymbol{x}_{\text{global}}$. There are two natural choices for these reference frames. In the lab frame, the constituents' four-momentum is given by the transverse momentum $p_{\text{T},i}$ of the particle and the azimuthal angular difference to the jet axis $\Delta\phi_i$, as well as the difference in pseudorapidity $\Delta\eta_i$. Alternatively, we can obtain a polar and an azimuthal angle after a Lorentz boost into the diboson rest frame. In this case, we can align the z axis of the spherical coordinate system with the jet axis and measure the azimuthal angle, denoted by φ_i, with respect to the beam plane as depicted in Fig. 1. In both cases, we denote Euclidean distances in the two-dimensional angular coordinates by ΔR. Either way, the dataset encodes information on the per-event decay plane that is disguised in the radiation patterns mapped to the variable-length constituent vector. The algorithm described in the next section is designed to extract its SMEFT sensitivity.

III. THE ALGORITHM

The constituents $\{\boldsymbol{x}_p\}_{i=1}^{N_p}$ of the highly energetic jet are individually reconstructed particles that can be viewed as a point cloud whose elements are feature vectors, composed of observables like four-momenta, charge, and other constituent properties.

General message-passing gNNs construct graphs on the point cloud and iteratively change the representations of the graphs' nodes by updating the feature vectors $\mathbf{h}_i$ by an aggregate of messages from the nodes in a neighborhood $\mathcal{N}(i)$ and, possibly, from the edges connecting the nodes [4–6]. In the most general case, a node update function $\psi^{(l)}$ and a message-passing function $\phi^{(l)}$ are highly expressive learnable functions, typically implemented as multilayer perceptrons (MLPs). At each iteration $l+1$, these determine the feature vector of a node i from the messages $^i\mathbf{m}_j^{(l)} = f_{\text{m}}^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \mathbf{e}_{ij})$ that are obtained

from the node features $\mathbf{h}_i^{(l)}$, the neighbors' node features $\mathbf{h}_j^{(l)}$, and, possibly, from edge features $\mathbf{e}_{ij}^{(l)}$ via the general node-update formula $\mathbf{h}_i^{(l+1)} = f_{\text{r}}^{(l)}(\mathbf{h}_i^{(l)}, \square_{j\in\mathcal{N}(i)}{}^i\mathbf{m}_j^{(l)})$. The aggregation function, denoted by $\square$, is permutation invariant and accumulates the messages from a suitably defined neighborhood $\mathcal{N}(i)$ of particle i. After a number of L iterations, the nodes are read out by accumulating the features $\mathbf{h}^{(L)}$ with an aggregation function that typically spans all nodes in the graph. In the simplest setting, this output is fed into a final global readout MLP whose parameters, together with the parameters of the message-passing and per-iteration readout MLPs, $f_{\text{m}}^{(l)}$ and $f_{\text{r}}^{(l)}$, are adjusted to minimize a problem-specific loss function.

First, we simplify the problem by demanding IRC safety via energy-weighted message passing, following Ref. [24]. This requirement imposes a number of restrictions. The input features for iteration $l=0$ are only the four-vectors p_i of the particles. The message-passing function, specifically, is given by $^i\mathbf{m}_j^{(l)} = \omega_j^{(\mathcal{N}(i))} f_{\text{m}}^{(l)}(\hat{p}_i, \hat{p}_j)$, where $\hat{p}$ denotes the direction of the three-momentum of the particles. The energy weighting is implemented via the relative hardness

$$\omega_j^{\mathcal{N}} = \frac{p_{\text{T},j}}{\sum_{k\in\mathcal{N}} p_{\text{T},k}}, \qquad (3)$$

where instead of the transverse particle momentum p_T, we could alternatively use another measure of the hardness of the particle, e.g., the energy or the three-momentum measured in a well-defined rest frame. Furthermore, the definition of the neighborhood cannot depend on the occurrence of infrared or collinear splits of the particles, excluding, e.g., the otherwise common kNN algorithm. Instead, a particle j is defined to be in the neighborhood $\mathcal{N}(i)$ of a particle i according to the Euclidean distance in the (pseudo)rapidity-azimuth plane, $\Delta R_{ij} = \sqrt{\Delta\phi(\hat{p}_i, \hat{p}_j)^2 + \Delta\eta(\hat{p}_i, \hat{p}_j)^2} \le \Delta R$. The threshold ΔR is a hyperparameter of the network. Choosing a specific per-event reference frame for the N_p massless constituents of a given W jet candidate, the gNN inputs for $l=0$ are represented by a feature list $\{p_{T,i}, \varphi_i, \Delta R_i\}_{i=1}^{N_p}$ as shown in Fig. 1.

Our algorithm is also SO(2) equivariant. In general, a function $\gamma: X \to Y$ is equivariant with respect to group transformations g of a group G, if there are two representations T_g and S_g, acting on the spaces X and Y, respectively, and the function satisfies $S_g(\gamma(x)) = \gamma(T_g(x))$. Because SO(2) is a one-dimensional group, we single out one angular coordinate, denoted by h_φ, and demand that the new representation $(h_\varphi, \mathbf{h}) \in [-\pi, \pi] \times \mathbb{R}^M$ transforms equivariantly under the SO(2) group action as

$$S_{\Delta\varphi}(h_\varphi, \mathbf{h}) = (h_\varphi + \Delta\varphi, \mathbf{h}). \qquad (4)$$

This group action applies a common shift $\Delta\varphi$ to the azimuthal coordinates h_φ and to the azimuthal input features φ_i. The M remaining features $\mathbf{h}$ of this data representation transform invariantly. $\mathbf{h}$ will encode information related to the substructure of the jet, which is also invariant with respect to these rotations, while h_φ can only contain information associated to quantities that transform equivariantly with the angle, effectively decoupling the two aspects.

Because the group manifold of SO(2) is a circle, we represent this coordinate in the complex plane as $e^{ih_{\varphi,i}}$ such that the message-passing relations

$$\mathbf{h}_i^{(l+1)} = \sum_{j \in N(i)} \omega_j^{(N(i))} \boldsymbol{f}_{\mathrm{h}}^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, h_{\varphi,i}^{(l)} - h_{\varphi,j}^{(l)})$$

$$e^{ih_{\varphi,i}^{(l+1)}} = e^{ih_{\varphi,i}^{(l)} + i \sum_{j \in N(i)} \omega_j^{(N(i))} f_\Phi^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, h_{\varphi,i}^{(l)} - h_{\varphi,j}^{(l)})} \quad (5)$$

are manifestly equivariant under the transformation (4) if we restrict the inputs to the message-passing neural networks, $f_\Phi^{(l)}$ and $\boldsymbol{f}_{\mathrm{h}}^{(l)}$, to group invariants. The distance function for $N(i)$ for $l > 1$ can be generalized to $\mathbf{h}_i^2 + \mathbf{h}_j^2 - 2\mathbf{h}_i \cdot \mathbf{h}_j \cos(h_{\varphi,i} - h_{\varphi,j})$. Besides the features $\mathbf{h}$, invariant under S by construction, the only other invariants are differences of φ coordinates, such that Eq. (5) is general. Other dependencies are not allowed when IRC safety should be ensured, such that $x_{\mathrm{p},i}$ reduces to p_i^μ, and concretely for the inputs to the first iteration,

$$\mathrm{h}_{\varphi,i}^{(0)} = \varphi_i, \qquad \mathbf{h}_i^{(0)} = \Delta R_i, \quad (6)$$

where ΔR_i is measured with respect to the jet axis. Because the inputs to the neural networks f_Φ and $\boldsymbol{f}_{\mathrm{h}}$ are the same, weights can be shared. In practice, a single neural network with $M^{(l)} + 1$ outputs, simultaneously providing f_Φ and $\boldsymbol{f}_{\mathrm{h}}$, has proven efficient. After L iterations, the global pooling is carried out with the energy weighting in Eq. (3), but summing over all the constituents, i.e., $\mathcal{N} = \{\mathbf{x}_{\mathrm{p}}\}_{i=1}^{N_{\mathrm{P}}}$ [24]. Together with the global features $\mathbf{x}_{\mathrm{global}}$, the pooled IRC-safe and SO(2) equivariant outputs are fed into a final MLP. A sketch of the whole construction is provided in Fig. 2.

Technically, the algorithm is implemented in PYTORCH [52] using the PYTORCH-GEOMETRIC (PyG) package [53]. Best performances were obtained with leaky ReLU activation functions with a slope parameter of 0.3. For training, we used the ADAM PYTORCH optimizer with a learning rate of 10^{-3}. Other choices for the activation functions (e.g., ReLU or sigmoid) and for the optimizer led to comparable performances, albeit with a slightly longer training time. This suggests that the algorithm performance is relatively robust. For the studies in the remainder of the paper, we found that a single iteration ($L = 1$) with a message-passing neural network with two hidden layers of node
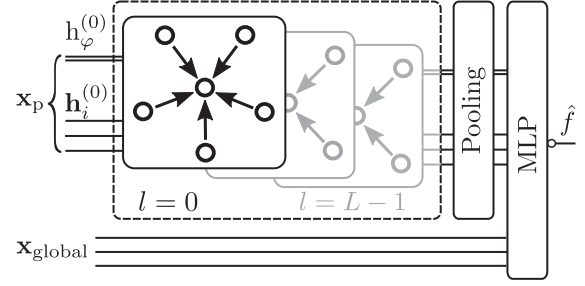

FIG. 2. Sketch of the network configuration as explained in the text. The equivariant feature is represented by a double line and the algorithms output by $\hat{f}$.

size 20 and a readout MLP with the same configuration performs sufficiently well. This is, therefore, our baseline configuration. Higher numbers of iterations train slower, with similar performance for the cases discussed below.

Finally, we note the possibility of breaking IRC safety by including extra information via the input features and still keeping the energy weighting in Eq. (3). For example, the constituents charge could help resolve the parton-level ambiguity of the light quarks from the W decay. While jet charge measurements were done by the ATLAS and CMS collaborations [54,55], the benefit of non-IRC-safe information, including vertex-association quality or the probability of originating from pileup vertices, cannot be confidently assessed without access to a comprehensive model of systematic uncertainties.

IV. TWO TOY SCENARIOS

We illustrate the algorithms' basic properties using a toy dataset that aims to reflect a simple two-prong structure in a variable-length feature vector. We first let $\bar{N}_{\mathrm{part}} = 50$ be the mean number of constituents in a mock-up jet. For each training instance, we sample two Poisson random numbers $N_{1,2}$ with means $\bar{N}_{1,2} = \bar{N}_{\mathrm{part}}/2$ such that each prong has, on average, the same number of constituents. Subsequently, we draw N_1 and N_2 two-dimensional coordinates $x_i = (x_1, x_2)_i$ from Normal distributions $\mathrm{N}(\boldsymbol{\mu_1}, \boldsymbol{\Sigma})$ and $\mathrm{N}(\boldsymbol{\mu_2}, \boldsymbol{\Sigma})$, respectively. We choose $\boldsymbol{\Sigma} = \mathrm{Diag}(0.3, 0.3)$. For the constituents' momenta $p_{\mathrm{T},i}$, we draw random numbers from a log-normal distribution $\ln\mathrm{N}(1, 0.2)$. We scale the $p_{\mathrm{T},i}$ by the likelihood of the constituent's 2D coordinate in the Normal distribution it was obtained from, which is $\mathrm{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_{1,2}, \boldsymbol{\Sigma})$. Because we sampled the $\mathbf{x}$ from Normal distributions, each prong will likely have high-momentum particles close to the location parameters $\boldsymbol{\mu}_{1,2}$. The sum of all constituent momenta, taken over both prongs, is finally scaled to 100 such that the mockup jet's total momentum is not random. During training, there is no information on the constituents' origin; each instance in the training set $\mathcal{D}$ is given solely by a total of $N_j = N_{j,1} + N_{j,2}$ constituents $\{p_{\mathrm{T},i}, x_i\}_{i=1}^{N_j}$ and a binary truth label y_j. There are no global
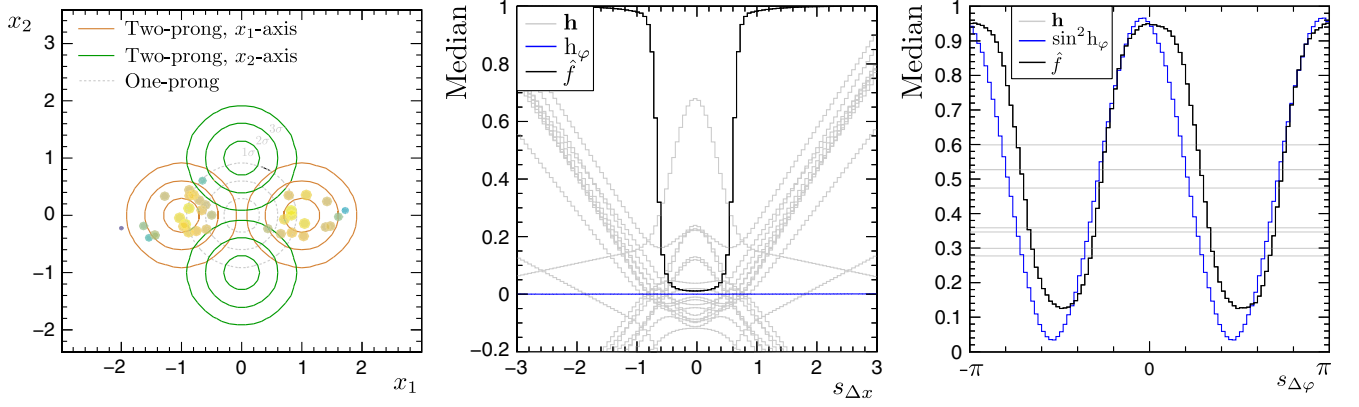
FIG. 3.   Left: contours of the probability density functions used in the toy example, along with an illustrative single event. Middle: median of the MLP output score (black) in scenario (A) as well as the medians of the 20 gNN outputs that serve as inputs to the redout MLP (gray) in arbitrary units. The equivariant feature h_γ is shown in blue. Right: median of the MLP output score (black) in scenario (B) as well as the medians of the first seven pooled gNN outputs that serve as inputs to the redout MLP (gray) in arbitrary units. The median of $\sin^2 h_\varphi$ is shown in blue.

event features in this study. The loss function, for simplicity, is the mean-squared-error

$$L = \sum_{\{p,x\}_j, y_j \in \mathcal{D}} (\hat{f}(\{\mathbf{x}_i, p_i\}_{i=1}^{N_j}) - y_j)^2. \qquad (7)$$

We have also tested the binary cross-entropy loss function and found no change in the performance.

In a scenario (A) we attempt a simple classification task between a two-prong signal category $\boldsymbol{\mu}_{1,2}^{\mathrm{sig}} = (\pm 1, 0)$ and a background category $\boldsymbol{\mu}_{1,2}^{\mathrm{bkg}} = (0, 0)$, the latter equivalent to a single prong with a mean of $\bar{N}_{\mathrm{part}}$ constituents. Contour lines of the probability density function (PDF) at 1σ, 2σ, and 3σ levels alongside the constituents of a single illustrative signal event are shown in Fig. 3 (left). Training this classifier on 10^5 pseudoevents for both categories leads to a background efficiency of approximately 1% for a signal efficiency of 99%. It is more interesting to study the behavior of the gNN output features before they enter the final MLP for different types of input. For this purpose, we exploit the simplicity of our toy dataset: a sample of signal events is, by construction, indistinguishable from a background sample when the constituents are shifted by an amount $\mathbf{x}_i \to \mathbf{x}_i + \Delta\mathbf{x}_i$ with $\Delta\mathbf{x}_i = -\boldsymbol{\mu}_{1,2}$, depending on whether constituent i came from prong 1 or prong 2. For illustration, we generate 10^4 independent test events and shift the constituents by $\Delta\mathbf{x}_i = -\boldsymbol{\mu}_1 + (s_{\Delta x}, 0)$ for prong 1 and $\Delta\mathbf{x}_i = -\boldsymbol{\mu}_2 - (s_{\Delta x}, 0)$. In this construction, the test sample is background-like for $s_{\Delta x} = 0$ and signal-like for $s_{\Delta x} = 1$. In Fig. 3 (middle) we show the median of the output score and the medians of the pooled gNN outputs before they enter the readout MLP. It is evident that the readout MLP learns to distinguish signal events from background events using the 20-dimensional internal representation $\mathbf{h}$. The insignificance of the

equivariant feature is expected, as there is no relevant azimuthal dependence of the prong structure in this classification task.

The situation is very different in scenario (B), where we train a classifier to separate between a signal with $\boldsymbol{\mu}_{1,2}^{\mathrm{sig}} = (\pm 1, 0)$ from a background category with $\boldsymbol{\mu}_{1,2}^{\mathrm{bkg}} = (0, \pm 1)$. This time, the signal has a prong structure along the x_1 axis while the background has a prong structure along the x_2 axis. In Fig. 3 (right) we show similar results as before, but this time for a signal sample that is rotated around the origin by an angle $s_{\Delta\varphi}$. Therefore, $s_{\Delta\varphi} = 0$ or $s_{\Delta\varphi} = \pm\pi$ correspond to a signal sample, while $s_{\Delta\varphi} = \pm\frac{\pi}{2}$ corresponds to a background sample. We first note from Fig. 3 (right) that the internal features do not vary at all with the angle of rotation $s_{\Delta\varphi}$. This manifests the invariance of $\mathbf{h}$ under the operation $S_{\Delta\varphi}$ in Eq. (5). Second, we note that there is a reflection ambiguity in the toy setup that mimics a similar reflection ambiguity of the light quarks in the W boson decay. The probability density function is symmetric under the exchange of x_1 and x_2; however, the value of h_φ changes by π under this operation. Therefore, we investigate a function of the equivariant feature which is invariant under this reflection. The simplest such function is $\sin^2 h_\phi$. The median of this value in the test dataset is shown in blue color in Fig. 3 (right) as a function of $s_{\Delta\varphi}$. It transforms with the rotation of inputs such that the events in the test sample are identified as signal events for rotations that are close to even multiples of $\pi/2$ and as background when the rotation is close to odd multiples of $\pi/2$. The modulations of $\sin^2 h_\varphi$ and the classifier output are not in phase, but that only reflects the networks' freedom in the internal representation.

In summary, the trained algorithm correctly reflects the basic properties of the classification task in both scenarios.

V. LEARNING FROM THE WZ PROCESS

We now showcase the performance of the algorithm in a more realistic setup using a sample of simulated WZ events in a semileptonic decay chain. The pp $\to$ W($\to$ q$\bar{\text{q}}$)Z($\to$ $\ell\bar{\ell}$) events are generated with MADGRAPH5_AMC@NLO v2.65 event generator [56] in the leading-order configuration with one extra parton in the hard-scatter process. The simulation sample uses the NNPDF3.1 PDF set [57]. The renormalization and factorization scales are not kept fixed, and their values are the default in MADGRAPH5, namely the transverse mass of the $2 \to 2$ system resulting from k_{T} clustering. A requirement of $H_{\text{T}} > 300$ GeV selects events with a sufficiently boosted W boson candidate, where H_{T} is the scalar sum of transverse final-state parton momenta. We use the SMEFTSIMv3.0 [58] model with single SMEFT operator insertions to simulate an event sample at $C_{\text{W}} = C_{\tilde{\text{W}}} = 1$ in order to ensure the kinematic phase space is well populated under the SM and SMEFT scenarios. The PYTHIAv8.24 [59] package is used to simulate the parton shower and hadronization. The matching between matrix element calculation in MADGRAPH5 and PYTHIA parton shower model is performed following the MLM [56] prescription. The detector response is emulated using DELPHESv3.5.0 [60] with the ATLAS card, such that the setup is equivalent to the top quark reference dataset described in Ref. [61].

Since the SMEFT effects induce a modulation in the distribution of φ_{decay}, we first use the algorithm to regress in this variable from the jet's constituents. Analogously to Sec. IV, the observed lab frame features of a training event j are given by $\mathbf{x}_j = \{p_{\text{T},i}, \varphi_i, \Delta R_i\}_{i=1}^{N_j}$, where the φ_i and ΔR_i are measured with respect to the jet's axis and N_j denotes the number of jet constituents. The parton-level $\varphi_{j,\text{decay}}$ is the regression target. We consider only events with a reconstructed AK8 jet with $p_{\text{T}} > 500$ GeV and we do not use any global event information. We train the algorithm on 80% of the WZ dataset by minimizing the loss function

$$L = \sum_{\mathbf{x}_j \in \mathcal{D}_{\text{sim}}} \sin^2(\hat{f}(\mathbf{x}_j) - \varphi_{j,\text{decay}}). \quad (8)$$

This choice, in particular the sine function, implements invariance to the reflection ambiguity, described in Sec. IV, at the level of the loss function: the data does not provide information to distinguish constituents originating from up- from down-type quarks and therefore will be invariant under transformations $\varphi_{j,\text{decay}} \to \pi - \varphi_{j,\text{decay}}$. The sine removes the ambiguity and this inductive bias improves the speed of convergence. It is, however, not needed in principle. We have tested that the network also converges with a standard mean-squared error loss or, e.g., with a piecewise linear function with the same symmetry properties as the sine function. Figure 4 shows the two-dimensional distribution of the true and the regressed φ_{decay}
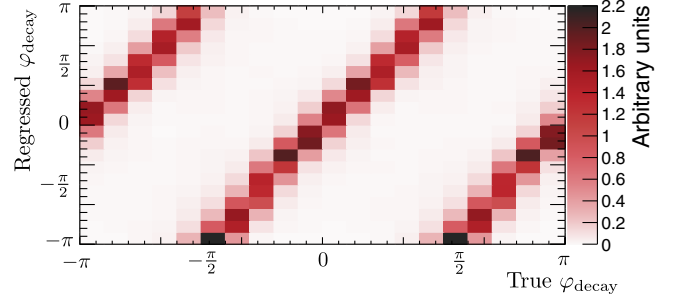

FIG. 4. Two-dimensional distribution of the true and regressed angle φ_{decay} in the test dataset.

in the remaining 20% WZ dataset, indicating a sensible behavior of the algorithm.

We now turn to extracting SMEFT sensitivity, considering events with a Z boson candidate constructed from its decay products, and a reconstructed jet, both with $p_{\text{T}} > 300$ GeV. For predicting SMEFT effects, we use a variant of the morphing technique [33] to obtain weighted predictions in the parameter space spanned by the Wilson coefficients. To this end, MADGRAPH5 with the MADWEIGHT [62] module is used to compute a per-event weight for a sufficiently large number of SMEFT parameter points $\mathbf{C} = (C_{\text{W}}, C_{\tilde{\text{W}}})$. Because second-order polynomials accurately describe the SMEFT matrix elements with a single operator insertion, a small number of such evaluations is sufficient to determine the coefficients of this polynomial, denoted by $\omega_j(\mathbf{C})$ for an event with a label j. We choose an overall normalization of the event weights such that at the SM parameter point $\mathbf{C} = 0$ we have $\sum_j \omega_j(0) = \mathcal{L}\sigma^{\text{WZ}}(\text{SM})$. It follows that the SMEFT prediction for any Poisson yield $\lambda(\mathbf{C})$ in a phase space volume $\Delta\mathbf{x}$, defined in terms of the observed features $\mathbf{x}$, is given by the sum of the polynomial per-event weights

$$\lambda(\mathbf{C}) = \mathcal{L} \int_{\Delta\mathbf{x}} d\mathbf{x} \frac{d\sigma(\mathbf{C})}{d\mathbf{x}} = \mathcal{L}\sigma(\mathbf{C}) \int_{\Delta\mathbf{x}} d\mathbf{x}\, p(\mathbf{x}|\mathbf{C})$$

$$= \mathcal{L}\sigma(\mathbf{C}) \int_{\Delta\mathbf{x}} d\mathbf{x} \int d\mathbf{z}\, p(\mathbf{x}, \mathbf{z}|\mathbf{C}) \approx \sum_{\mathbf{x}_j \in \Delta\mathbf{x}} \omega_j(\mathbf{C}). \quad (9)$$

The second expression gives the yield in terms of the differential cross section, the third relates the cross section to the detector-level PDF $p(\mathbf{x}|C)$, the fourth introduces the joint PDF of the observation $\mathbf{x}$ and the latent features $\mathbf{z}$, and the last term, finally, is the Monte-Carlo approximation from a simulated sample $\mathcal{D}_{\text{sim}} = \{\omega_j(\mathbf{C}), \mathbf{x}_j, \mathbf{z}_j\}_{j=1}^{N_{\text{sim}}}$. Because Monte-Carlo simulation is a sampling of the joint $(\mathbf{x}, \mathbf{z})$ space [31], the ratio of two polynomial event weights is expressed in terms of the ratio of the joint likelihood,

$$\frac{\omega_j(\mathbf{C})}{\omega_j(0)} = \frac{\sigma(\mathbf{C})}{\sigma(\text{SM})} \frac{p(\mathbf{x}_j, \mathbf{z}_j|\mathbf{C})}{p(\mathbf{x}_j, \mathbf{z}_j|\text{SM})}. \quad (10)$$

Here, the first factor on the rhs accounts for the dependence of the total cross section on the Wilson coefficients. The simulation-based inference technique we use for learning the SM-SMEFT interference is based on Refs. [29–34] and similar to the SALLY method [30]. It capitalizes on the fact that the joint likelihood ratio is tractable, i.e., simulation allows evaluating Eq. (10) as a function of $\mathbf{C}$.

Because our targeted effect is linear in $\mathbf{C}$, we are interested in the model's behavior near the SM where the score vector

$$\mathbf{t}(\mathbf{x}) = \nabla_\mathbf{C} \log p(\mathbf{x}|\mathbf{C})|_{\mathbf{C}=\mathrm{SM}} = \frac{\nabla_\mathbf{C} p(\mathbf{x}|\mathbf{C})}{p(\mathbf{x}|\mathrm{SM})}|_{\mathbf{C}=\mathrm{SM}} \quad (11)$$

is a sufficient statistic [31]. It provides a locally optimal observable, i.e., extracts the maximum amount of information from the training data [31]. At first sight, it appears problematic that our quantity of interest in Eq. (11) is a ratio of detector-level likelihoods, with separate implicit latent-space integrations in the numerator and denominator, while in Eq. (10), the quantity available in simulation, there appears a ratio of latent-space PDFs without that integration. Following Refs. [29,30], we can nevertheless learn the detector-level score from expanding the log derivative of Eq. (10) to linear order,

$$\mathbf{t}(\mathbf{x}_j, \mathbf{z}_j) = \nabla_\mathbf{C} \log \frac{\omega_j(\mathbf{C})}{\sigma(\mathbf{C})}\bigg|_{\mathbf{C}=\mathrm{SM}}. \quad (12)$$

This quantity is tractable for simulated events. It is then straightforward to show [29] that the analytic loss function

$$L = \int d\mathbf{x}\, d\mathbf{z}\, p(\mathbf{x}, \mathbf{z}|\mathrm{SM})(\mathbf{t}(\mathbf{x}, \mathbf{z}) - \hat{f}(\mathbf{x}))^2 \quad (13)$$

for an infinitely expressive $\hat{f}$, depending on the observable $\mathbf{x}$ but not on the latent $\mathbf{z}$, provides a minimum $f^*(\mathbf{x}) = \mathbf{t}(\mathbf{x})$ as defined in Eq. (11). Our empirical loss function is, therefore, the empirical version of Eq. (13), given by

$$L = \sum_{\{\omega_j(\mathbf{C}), \mathbf{x}_j, \mathbf{z}_j\} \in \mathcal{D}_{\mathrm{sim}}} \omega_j(\mathrm{SM})(\mathbf{t}(\mathbf{x}_j, \mathbf{z}_j) - \hat{f}(\mathbf{x}_j))^2, \quad (14)$$

where we make use of Eq. (12). All of its ingredients are available in the simulation. Finally, we can remove $\sigma(\mathbf{C})$ in Eq. (12) if we use $\hat{f}$ only as a discriminator in a hypothesis test. The total cross section has no impact on its discriminative power as this term only provides a constant shift to the predicted value, common to all events.

Minimizing this loss function over 3×10^5 training events provides a surrogate of the detector-level two-component score vector of C_W and $C_{\tilde{\mathrm{W}}}$. The distance parameter ΔR in the gNN is set to the value of 0.4. The same network configuration as in Sec. IV is used, with little observed dependence on the MLP configuration. The detector-level transverse momenta of the leptonic and hadronic bosons show mild dependence on C_W at the linear level and are, therefore, used as global features for the C_W component, while we do not use global features for the $C_{\tilde{\mathrm{W}}}$ component.

We show the result of the training in Fig. 5 (left) on a statistically independent test dataset. The detector-level score exhibits the expected sinusoidal modulations as a function of the true decay plane angle. The gNN very accurately reproduces this dependence, showing that the gNN can indeed recover SMEFT effects in the angular radiation patterns.

In Fig. 5 (middle, right) we perform one-dimensional likelihood scans of our surrogate model normalized to 2000 expected events. Without incorporating systematic
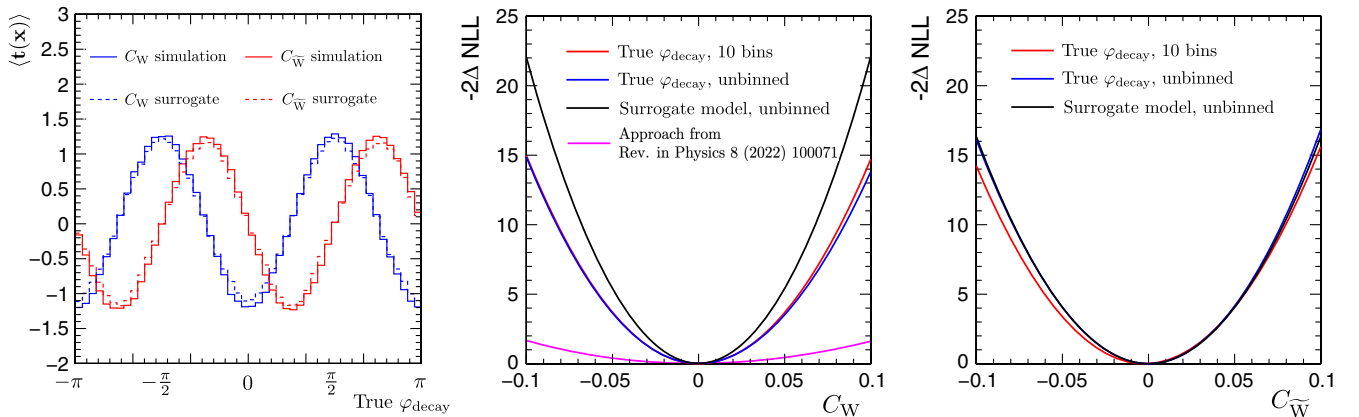

FIG. 5.   Left: mean score as a function of φ_{decay}, obtained with the true model and the regressed surrogate model. Center and right: likelihood scan as a function of C_W (center) and $C_{\tilde{\mathrm{W}}}$ (right) using different likelihood models, assuming 2000 observed events. The surrogate model described in this paper is compared with a MLP that takes as an only input the φ_{decay} and the likelihood function of a counting experiment considering ten bins in φ_{decay}. We also compare to a strategy based on the soft-drop mass M_{SD} [63] and N-subjettiness ratio τ_{21} [64].

uncertainties, this procedure will lead to optimistic results. Nevertheless, comparing the surrogate model's performance to the truth level is enlightening. The negative log-likelihood (NLL) ratio is shown for three different test statistics. First, the true decay-plane angle φ_{decay} shows similar performance in a binned Poisson and as an unbinned test statistic, suggesting that a binned analysis in the regressed φ_{decay} may have good sensitivity. Second, the unbinned surrogate model performs well for both, the C_{W} and $C_{\tilde{\text{W}}}$ Wilson coefficients. Despite the complexity of the hadronic final state, the gNN thus manages to extract the leading linear SMEFT dependence from the boosted jet's constituents. In the case of C_{W}, the surrogate model also profits from the SMEFT dependence of the global event features, increasing the sensitivity. This showcases the ability of the algorithm to combine information from the radiation patterns and observables sensitive to energy growth. For the $C_{\tilde{\text{W}}}$ coefficient, the energy growth is much smaller, and φ_{decay} dominates the sensitivity.

We also compare to a strategy suggested in Ref. [43] that is based on the soft-drop mass observable M_{SD} [63] and the $\mathcal{N}$-subjettiness ratio τ_{21} [64], both computed from the constituents of the jet. In the training for this regressor, we also include all the global event features from the nominal training, but our graph neural network is not used. With this approach, we obtain a negligible sensitivity for the $C_{\tilde{\text{W}}}$ coefficient, which is not shown, and a small sensitivity to the C_{W} coefficient compared to our gNN approach. This can be understood from the almost exclusive effect of $C_{\tilde{\text{W}}}$ and C_{W} on the orientation of the decay plane, which is only captured by the gNN and, therefore, the gNN shows good sensitivity gains in both cases.

In combination, the results show that our gNN can extract the linear SMEFT dependence from the decay plane angle of a hadronically decaying W boson, unlocking the large hadronic branching fractions for future analyses in search of new physics.

VI. CONCLUSIONS AND OUTLOOK

In this paper, we have constructed an IRC-safe and rotation-equivariant gNN, whose main input is the variable-length list of the particle constituents of a jet produced in the hadronic decay of a boosted massive particle. The algorithm is able to extract information on the angular orientation of the radiation patterns present in the jet, decoupled from other aspects of the jet substructure thanks to the equivariance of the network.

We have investigated the main features in simple toy studies and applied the algorithm to linear SMEFT effects in semileptonic final states of the WZ process at the LHC. It learns a surrogate of the score vector, thus providing an optimal observable for small deviations from the SM. We have shown that the score is well reproduced as a function of the true decay plane angle, thus fully extracting its SMEFT sensitivity. Moreover, we incorporate information from additional observables encoding effects from energy growth, boosting the sensitivity towards the theoretical optimum.

The algorithm allows the large hadronic branching fractions to be utilized in future SMEFT analyses. It is also suitably general to be useful in a variety of applications, potentially including three-prong decays of boosted top quarks or semileptonic final states of vector boson scattering at the LHC.

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91,** 045002 (2019).

[2] M. Erdmann, J. Glombitza, G. Kasieczka, and U. Klemradt, *Deep Learning for Physics Research* (World Scientific, Singapore, 2021).

[3] A. Tanaka, A. Tomiya, and K. Hashimoto, Epilogue, in *Deep Learning and Physics* (Springer Singapore, Singapore, 2021), p. 195.

[4] T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, in *International Conference on Learning Representations* (2017); arXiv:1609.02907.

[5] J. Gilmer *et al.*, Neural message passing for quantum chemistry, in *Proceedings of the 34th International Conference on Machine Learning* (2017); arXiv:1704.01212.

[6] Y. Wang *et al.*, Dynamic graph CNN for learning on point clouds, arXiv:1801.07829.

[7] M. Erdmann, E. Geiser, Y. Rath, and M. Rieger, Lorentz boost networks: Autonomous physics-inspired feature engineering, J. Instrum. **14,** P06006 (2019).

[8] H. Qu and L. Gouskos, ParticleNet: Jet tagging via particle clouds, Phys. Rev. D **101,** 056019 (2020).

[9] P. Onyisi, D. Shen, and J. Thaler, Comparing point cloud strategies for collider event classification, Phys. Rev. D **108,** 012001 (2023).

[10] G. DeZoort, P. W. Battaglia, C. Biscarat, and J.-R. Vlimant, Graph neural networks at the Large Hadron Collider, Nat. Rev. Phys. **5**, 281 (2023).

[11] A. Butter *et al.*, The machine learning landscape of top taggers, SciPost Phys. **7**, 014 (2019).

[12] V. G. Satorras, E. Hoogeboom, and M. Welling, E(n) equivariant graph neural networks, in *Proceedings of the 38th International Conference on Machine Learning* (2021); arXiv:2102.09844.

[13] S. Villar *et al.*, Scalars are universal: Equivariant machine learning, structured like classical physics, Adv. Neural Inf. Process. Syst. **34**, 28848 (2021), https://proceedings.neurips.cc/paper_files/paper/2021/hash/f1b0775946bc0329b35b-823b86eeb5f5-Abstract.html.

[14] M. J. Dolan and A. Ore, Equivariant energy flow networks for jet tagging, Phys. Rev. D **103**, 074022 (2021).

[15] M. Favoni, A. Ipp, D. I. Müller, and D. Schuh, Lattice gauge equivariant convolutional neural networks, Phys. Rev. Lett. **128**, 032003 (2022).

[16] S. Bulusu *et al.*, Equivariance and generalization in neural networks, EPJ Web Conf. **258**, 09001 (2022).

[17] S. Gong, Q. Meng, J. Zhang, H. Qu, C. Li, S. Qian, W. Du, Z.-M. Ma, and T.-Y. Liu, An efficient Lorentz equivariant graph neural network for jet tagging, J. High Energy Phys. 07 (2022) 030.

[18] A. Bogatskiy *et al.*, Symmetry group equivariant architectures for physics, in *Snowmass 2021* (2022); arXiv:2203.06153.

[19] M. Favoni, A. Ipp, and D. I. Müller, Applications of lattice gauge equivariant neural networks, EPJ Web Conf. **274**, 09001 (2022).

[20] A. Bogatskiy, T. Hoffman, D. W. Miller, and J. T. Offermann, PELICAN: Permutation equivariant and Lorentz invariant or covariant aggregator network for particle physics, in *Machine Learning and the Physical Sciences Workshop, NeurIPS 2022* (2022); arXiv:2211.00454.

[21] E. Buhmann, G. Kasieczka, and J. Thaler, EPiC-GAN: Equivariant point cloud generation for particle jets, SciPost Phys. **15**, 130 (2023).

[22] D. Murnane, S. Thais, and A. Thete, Equivariant graph neural networks for charged particle tracking, in *21th International Workshop on Advanced Computing and Analysis Techniques in Physics Research: AI Meets Reality* (2023); arXiv:2304.05293.

[23] A. Bogatskiy *et al.*, Explainable equivariant neural networks for particle physics: PELICAN, arXiv:2307.16506.

[24] P. Konar, V. S. Ngairangbam, and M. Spannowsky, Energy-weighted message passing: An infra-red and collinear safe graph neural network algorithm, J. High Energy Phys. 02 (2022) 060.

[25] P. T. Komiske, E. M. Metodiev, and J. Thaler, Energy flow networks: Deep sets for particle jets, J. High Energy Phys. 01 (2019) 121.

[26] W. Buchmuller and D. Wyler, Effective Lagrangian analysis of new interactions and flavor conservation, Nucl. Phys. **B268**, 621 (1986).

[27] G. F. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi, The strongly-interacting light Higgs, J. High Energy Phys. 06 (2007) 045.

[28] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, Dimension-six terms in the standard model Lagrangian, J. High Energy Phys. 10 (2010) 085.

[29] K. Cranmer, J. Pavez, and G. Louppe, Approximating likelihood ratios with calibrated discriminative classifiers, arXiv:1506.02169.

[30] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, Constraining effective field theories with machine learning, Phys. Rev. Lett. **121**, 111801 (2018).

[31] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, A guide to constraining effective field theories with machine learning, Phys. Rev. D **98**, 052004 (2018).

[32] J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, Mining gold from implicit models to improve likelihood-free inference, Proc. Natl. Acad. Sci. U.S.A. **117**, 5242 (2020).

[33] J. Brehmer, F. Kling, I. Espejo, and K. Cranmer, MadMiner: Machine learning-based inference for particle physics, Comput. Software Big Sci. **4**, 3 (2020).

[34] J. Brehmer, S. Dawson, S. Homiller, F. Kling, and T. Plehn, Benchmarking simplified template cross sections in WH production, J. High Energy Phys. 11 (2019) 034.

[35] A. Butter, T. Plehn, N. Soybelman, and J. Brehmer, Back to the formula—LHC edition, SciPost Phys. **16**, 037 (2024).

[36] S. Chatterjee, S. Rohshap, R. Schöfbeck, and D. Schwarz, Learning the EFT likelihood with tree boosting, arXiv:2205.12976.

[37] R. Gomez Ambrosio, J. t. Hoeve, M. Madigan, J. Rojo, and V. Sanz, Unbinned multivariate observables for global SMEFT analyses from machine learning, J. High Energy Phys. 03 (2023) 033.

[38] S. Chen, A. Glioti, G. Panico, and A. Wulzer, Parametrized classifiers for optimal EFT sensitivity, J. High Energy Phys. 05 (2021) 247.

[39] S. Chen, A. Glioti, G. Panico, and A. Wulzer, Boosting likelihood learning with event reweighting, arXiv:2308.05704.

[40] G. Panico, F. Riva, and A. Wulzer, Diboson interference resurrection, Phys. Lett. B **776**, 473 (2018).

[41] S. Banerjee, R. S. Gupta, J. Y. Reiness, S. Seth, and M. Spannowsky, Towards the ultimate differential SMEFT analysis, J. High Energy Phys. 09 (2020) 170.

[42] S. Banerjee, R. S. Gupta, O. Ochoa-Valeriano, M. Spannowsky, and E. Venturini, A fully differential SMEFT analysis of the golden channel using the method of moments, J. High Energy Phys. 06 (2021) 031.

[43] D. Buarque Franzosi *et al.*, Vector boson scattering processes: Status and prospects, Rev. Phys. **8**, 100071 (2022).

[44] C. Degrande and J. Touchèque, A reduced basis for *CP* violation in SMEFT at colliders and its application to diboson production, J. High Energy Phys. 04 (2022) 032.

[45] CMS Collaboration, Measurement of the inclusive and differential WZ production cross sections, polarization angles, and triple gauge couplings in pp collisions at $\sqrt{s} = 13$ TeV, J. High Energy Phys. 07 (2022) 032.

[46] CMS Collaboration, Measurement of $W^{\pm}\gamma$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV and effective field theory constraints, Phys. Rev. D **105**, 052003 (2022).

[47] ATLAS Collaboration, Combined effective field theory interpretation of differential cross-sections measurements

of WW, WZ, 4l, and Z-plus-two-jets production using ATLAS data, Technical Report No. ATL-PHYS-PUB-2021-022, CERN, Geneva, 2021, All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-022.

[48] M. Cacciari, G. P. Salam, and G. Soyez, The anti-k_T jet clustering algorithm, J. High Energy Phys. 04 (2008) 063.

[49] M. Cacciari, G. P. Salam, and G. Soyez, FastJet user manual, Eur. Phys. J. C **72**, 1896 (2012).

[50] CMS Collaboration, Particle-flow reconstruction and global event description with the CMS detector, J. Instrum. **12**, P10003 (2017).

[51] D. Bertolini, P. Harris, M. Low, and N. Tran, Pileup per particle identification, J. High Energy Phys. 10 (2014) 059.

[52] A. Paszke *et al.*, PYTORCH: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32* (Curran Associates, Inc., Vancouver, Canada, 2019), p. 8024, http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[53] M. Fey and J. E. Lenssen, Fast graph representation learning with PYTORCH geometric, in *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019); arXiv:1903.02428.

[54] ATLAS Collaboration, Measurement of jet charge in dijet events from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector, Phys. Rev. D **93**, 052003 (2016).

[55] CMS Collaboration, Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV, J. Instrum. **12**, P02014 (2017).

[56] 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.

[57] NNPDF Collaboration, Parton distributions from high-precision collider data, Eur. Phys. J. C **77**, 663 (2017).

[58] I. Brivio, SMEFTSIM3.0—A practical guide, J. High Energy Phys. 04 (2021) 073.

[59] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, An introduction to PYTHIA8.2, Comput. Phys. Commun. **191**, 159 (2015).

[60] DELPHES 3 Collaboration, DELPHES 3, a modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.

[61] G. Kasieczka, T. Plehn, J. Thompson, and M. Russel, Top quark tagging reference dataset (2019), http://dx.doi.org/10.5281/zenodo.2603256.

[62] P. Artoisenet, V. Lemaitre, F. Maltoni, and O. Mattelaer, Automation of the matrix element reweighting method, J. High Energy Phys. 12 (2010) 068.

[63] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, Soft Drop, J. High Energy Phys. 05 (2014) 146.

[64] J. Thaler and K. Van Tilburg, Identifying boosted objects with N-subjettiness, J. High Energy Phys. 03 (2011) 015.