Improved constraints on effective top quark interactions using edge convolution networks

Oliver Atkinson, Akanksha Bhardwaj, Stephen Brown, Christoph Englert, David J. Miller and Panagiotis Stylianou

School of Physics & Astronomy, University of Glasgow,
Glasgow G12 8QQ, U.K.

E-mail: o.atkinson.1@research.gla.ac.uk,
akanksha.bhardwaj@glasgow.ac.uk, s.brown.7@research.gla.ac.uk,
christoph.englert@glasgow.ac.uk, david.j.miller@glasgow.ac.uk,
p.stylianou.1@research.gla.ac.uk

ABSTRACT: We explore the potential of Graph Neural Networks (GNNs) to improve the performance of high-dimensional effective field theory parameter fits to collider data beyond traditional rectangular cut-based differential distribution analyses. In this study, we focus on a SMEFT analysis of $pp \to t\bar{t}$ production, including top decays, where the linear effective field deformation is parametrised by thirteen independent Wilson coefficients. The application of GNNs allows us to condense the multidimensional phase space information available for the discrimination of BSM effects from the SM expectation by considering all available final state correlations directly. The number of contributing new physics couplings very quickly leads to statistical limitations when the GNN output is directly employed as an EFT discrimination tool. However, a selection based on minimising the SM contribution enhances the fit's sensitivity when reflected as a (non-rectangular) selection on the inclusive data samples that are typically employed when looking for non-resonant deviations from the SM by means of differential distributions.

Contents

1 Introduction

The search for new physics, albeit so far unsuccessful at the Large Hadron Collider (LHC) remains a cornerstone of the collider phenomenology programme. The lack of direct evidence for new states beyond the Standard Model (BSM) can be interpreted as an indication of a separation between the electroweak scale and the scale of new physics Λ. Integrating out BSM states can then be cast into a consistent and systematic extension of the SM by higher dimensional operators $\mathcal{O}$ with dimensions $[\mathcal{O}] > 4$ [1]. Turning the argument around, constraints on any extension of the SM can be obtained by agnostically reflecting all a priori allowed operator deformations in the interpretation of LHC results up to a considered order in the Λ^{-1} expansion. At operator level up to dimension six in Standard Model Effective Theory (SMEFT) [2–7]

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{C_i}{\Lambda^2} \mathcal{O}_i \,, \tag{1.1}$$

this programme has been comprehensively addressed from a theoretical perspective (see e.g. [8–14] and [15] for a recent review) and provides a theoretically consistent interpretation at this order in the Λ^{-1} expansion. Related phenomenological (proof-of-principle) analyses have focussed on (combinations of) Higgs physics [16–23], electroweak precision observables [24–26], and the top sector [27–34], owing to good statistical and systematic control at past and present collider experiments, and their expected roles in BSM physics in general.

The experimental implementation of such a strategy is far from trivial. The number of involved and independent effective interactions can be large, thus potentially limiting the

sensitivity of a single, specific analysis. In parallel, systematic uncertainties can lead to weak constraints, giving only loose and perhaps non-perturbative limits when understood as UV constraints in concrete matching calculations.

There are two avenues to improve phenomenological sensitivity. Firstly, decreasing the theoretical and experimental uncertainties alongside systematics of the EFT and SM hypotheses, possibly via data-driven approaches, will lead to improved limits when more data become available (assuming agreement with the SM prevails). Lower limits on the direct evidence of new states, e.g., via s-channel production are predominantly driven by the available LHC centre-of-mass energy. Therefore, the lower limit on Λ in eq. (1.1), which is driven by the LHC's energy coverage, will not change dramatically in the future. Thus, any modelling improvement at scales $|Q^2| \ll \Lambda^2$ where the EFT expansion can be considered reliable will be reflected in improved constraints on the Wilson coefficients (WCs) C_i (modulo remaining blind directions).

Secondly, we can resolve to a more comprehensive extraction of information from experimental data. Such strategies are highlighted in the recent resurgence of machine learning (ML) applications to particle physics [35–45] (in particular focusing also on experimental improvements [46, 47]). 'Traditional' collider observables such as transverse momenta, angles and (pseudo)rapidities, alongside rectangular cuts on these, might not fully capture the exclusion potential when all ad hoc modifications of correlations are considered, which is the key motivation of the EFT approach (in particular this extends to the inclusion of systematic uncertainties [48]).

The latter of these two avenues is the focus of this work. We focus on EFT parameter constraints for the top sector [27–34]. In particular, we focus on $pp \to t\bar{t}$ production with semi-leptonic top decays as this provides a relatively clean channel with good statistical control to discuss ML-improved EFT strategies (see also ref. [39]). To reflect expected correlations between the final state (i.e. fully showered and hadronised) objects, we employ Graph Neural Networks (GNNs) with Edge Convolution [49–52]. This setup exploits the structure of data as well as the correlations ('edges') of different intermediate and final state particles ('nodes') and is therefore well motivated for particle physics applications [53–64].

This paper is organised as follows. In section 2, we review the EFT operators relevant for this case study. We also detail our simulation, analysis and fit setup of $t\bar{t}$ production. section 3 is devoted to the machine learning aspects of this study: we briefly outline our baseline cuts (taking the experimental analysis of [65] as guidance), review our ML setup, and discuss input parameters, training and classification. We highlight the performance improvements of a ML-informed top sector fit in section 4 and conclude in section 5.

2 Effective interactions for top pair production with leptonic decays

Any differential cross section that follows from eq. (1.1) can be written as

$$\mathrm{d}\sigma = \mathrm{d}\sigma_{\mathrm{SM}} + \frac{C_i}{\Lambda^2}\mathrm{d}\sigma_i^{(1)} + \frac{C_i C_j}{\Lambda^4}\mathrm{d}\sigma_{ij}^{(2)}, \tag{2.1}$$

where the C_i are the Wilson Coefficients (WCs) and Λ is the generic scale of new physics (NP). The first term shows the contribution from the Standard Model (SM) only, while

Distribution	Observable	Binning
$\frac{1}{\sigma}\frac{d\sigma}{d|y_t^h|}$	$|y_t^h|$	$[0.0, 0.2, 0.4, 0.7, 1.0, 1.3, 1.6, 2.5]$
$\frac{1}{\sigma}\frac{d\sigma}{d|y_t^l|}$	$|y_t^l|$	$[0.0, 0.2, 0.4, 0.7, 1.0, 1.3, 1.6, 2.5]$
$\frac{1}{\sigma}\frac{d\sigma}{d|y_{t\bar{t}}|}$	$|y_{t\bar{t}}|$	$[0.0, 0.2, 0.4, 0.6, 0.9, 1.3, 2.3]$
$\frac{1}{\sigma}\frac{d\sigma}{dp_\perp^{t,h}}$	$p_\perp^{t,h}$	$[0, 45, 90, 135, 180, 225, 270, 315, 400, 800]$ GeV
$\frac{1}{\sigma}\frac{d\sigma}{dp_\perp^{t,l}}$	$p_\perp^{t,l}$	$[0, 45, 90, 135, 180, 225, 270, 315, 400, 800]$ GeV
$\frac{1}{\sigma}\frac{d\sigma}{dm_{t\bar{t}}}$	$m_{t\bar{t}}$	$[300, 375, 450, 530, 625, 740, 850, 1100, 2000]$ GeV
$\frac{1}{\sigma}\frac{d\sigma}{d|y_{t\bar{t}}|d|m_{t\bar{t}}|}$	$|y_{t\bar{t}}|$	$[0.0, 0.2, 0.4, 0.6, 0.9, 1.3, 2.3]$
	$m_{t\bar{t}}$	$[300, 375, 450, 625, 850, 2000]$ GeV
$\frac{1}{\sigma}\frac{d\sigma}{dp_\perp^{t,h}d|y_t^h|}$	$p_\perp^{t,h}$	$[0, 45, 90, 135, 180, 225, 270, 315, 400, 800]$ GeV
	$|y_t^h|$	$[0.0, 0.5, 1.0, 1.5, 2.5]$

Table 1. Distributions provided in ref. [65] and included in the fit in this work.

the second term is the contributions from the interference of the EFT and the SM terms. The third term represents the contribution from the EFT squared or cross-terms which are Λ^4 suppressed. In the following, we will limit ourselves to dimension 6 (differential) cross sections $\sim \Lambda^{-2}$ that result from interference of the EFT and SM amplitudes. While this is a theoretically consistent approach, it also constitutes a conservative case for EFT limit setting: contributions $\sim \Lambda^{-4}$ typically show a dramatic momentum-transfer enhanced behaviour and are therefore relatively easy to constrain, even using standard approaches. Put differently, any sensitivity improvement that we can identify for the linearised approach will generalise to the inclusion of the $\sim \Lambda^{-4}$ terms in eq. (2.1).

2.1 Analysis setup and fit methodology

We use the SMEFTSim [66, 67] implementation to include the effective operators, which is then interfaced with MadGraph5 [68] via FeynRules [69] and UFO [70] to generate the event samples at leading order (LO)[1] for

$$pp \to t\bar{t} \to \ell b\bar{b}j + \not{E}_T \,. \tag{2.2}$$

We use a $\sqrt{s} = 13$ TeV analysis by the CMS collaboration [65] as inspiration to investigate (correlated) differential measurement results and representative data binning as given in table 1. SM predictions are injected as mock reference data for the luminosity $\mathcal{L}_{\text{ref}} = 2.3$ fb^{-1} of ref. [65] and we scale statistical uncertainties relative to this luminosity, using

---

[1]In this work, we focus on GNN performance of EFT parameter fits and limit ourselves to a leading order analysis. We note that including higher order contributions for the SM hypothesis is crucial to obtain consistency with the measured data, but will not impact the qualitative results of this work. We have checked that the results of table 2 are qualitatively reproduced by a full NLO fit using a forthcoming version of TOPFITTER [71].
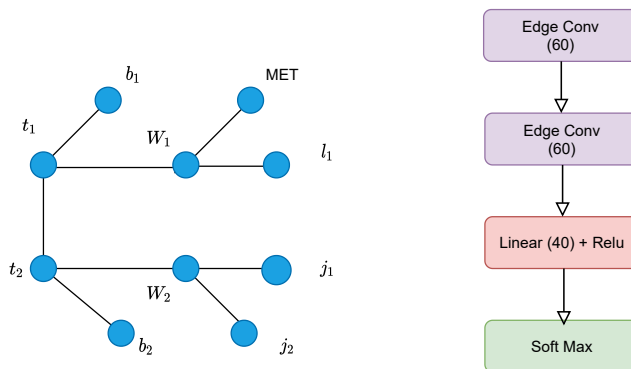
Figure 1. Representative diagram for the input graph and the network architecture used in this paper.

$\sqrt{\mathcal{L}_{\text{ref}}/\mathcal{L}}$ for extrapolations. Our implementation relies on Rivet [72, 73], which processes events after showering with Pythia8 [74] before feeding them into the fit.

To avoid imposing any assumptions as to correlations — and remove the chance that double-counting of events would artificially inflate sensitivity to EFT contributions — a single distribution is used where bin-to-bin correlations are included, and a single bin is used where they are not. In the absence of a full reference correlation/covariance matrix the selection of the bin/distribution is made on a coefficient-by-coefficient basis, with the input with maximum deviation from a fixed point on that axis being selected. This maximum sensitivity, minimum correlation assumption input subset is then used to determine individual and profiled bounds for the coefficient being studied. Where a normalised distribution is used we must drop a bin, as otherwise the covariance matrix will be singular. The dropped bin is chosen such that we obtain the most stable covariance matrix, with the bin with the largest uncertainty being dropped if there are multiple bins leading to an equivalently well-conditioned covariance matrix.[2]

In the following we will consider bounds for all relevant operators using the dimensionless 'bar' notation

$$\bar{C}_i = C_i \frac{v^2}{\Lambda^2}, \tag{2.3}$$

with the electroweak expectation value $v \simeq 246$ GeV.

In many standard analyses, cut-and-count techniques are often used to restrict the phase space region in such a way that the SM contamination is minimised, and as a result this yields an increased new-physics sensitivity. However rectangular cuts often yield inferior sensitivity compared to the methodically selected regions by means of machine learning classifiers. In our scenario an efficient event-by-event classification using GNNs, separating the generated events into either pure SM or the SMEFT operators that sourced them, could lead to improvements on the bounds of WCs after imposing cuts on the output score of the network.

---

[2]For details on statistical inference we refer interested readers to refs. [27, 28].

3 Graph representation of events

In order to use a GNN as a classifier, the events need to be embedded in a graph structure with nodes, edges and features associated to observables of final states or reconstructed objects. While various different approaches are possible to construct a graph from the IR-safe, calibratable and detectable final states, we employ a physics-motivated strategy, creating graphs similar to the tree of the chain of eq. (2.2).[3] Concretely, we pre-process the data samples and require at least two jets of transverse momentum $p_T(j) > 20$ GeV and pseudorapidity $|\eta| < 5$ that are not b-tagged. The event is vetoed if there are not at least two b-jets and one lepton ℓ in the central part of the detector ($|\eta(\ell)| < 2.5$), where the b-jets must also satisfy $p_T(b) > 20$ GeV. Subsequently, we embed the passed events into graphs using the following steps (see also figure 1):

(i) Nodes. Firstly, the missing transverse momentum (MTM) is identified by balancing the net visible momenta, $-p(\text{visible})$, neglecting the longitudinal components. A node is added corresponding to MTM. Then, for each lepton, we attempt to reconstruct the W four-momentum as a sum of the lepton's four-momentum and the MTM. The invariant mass of the W candidate is calculated and if it falls within $[65, 95]$ GeV a node is added, labelled W_1, as well as one for the b-jet b_1 that has the smallest separation $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ from W_1. In the case where there are more lepton-MTM combinations with compatible invariant mass, the one closest to the W boson mass is selected. The top from the leptonic decay chain t_1 is finally reconstructed from the four-momenta of ℓ, b_1 and MTM and obtains its respective node. Following a similar procedure, we consider combinations of jets to find a pair with dijet invariant mass 70 GeV $\leq m(jj) \leq$ 90 GeV. If a pair is found we add nodes for the two jets j_1, j_2 and for the second boson W_2, otherwise we only add nodes for the two leading jets. From the remaining b-jets a node is added for the leading one, b_2, as well as for the second top t_2 whose four-momentum is reconstructed using b_2, j_1 and j_2. We scan over the remaining particles and if any are within $\Delta R < 0.8$ of any of the identified or reconstructed objects we add a node that will be connected only to the nearby object.

(ii) Edges. The connections between the nodes create the adjacency matrix of the graph and the nodes of the final states are connected to the ones of the reconstructed objects from which they are derived. We first connect the MTM and lepton to W_1 and subsequently, W_1 and b_1 are connected to the first top quark node. If a W_1 was not created then the aforementioned final states connect directly to t_1.[4] Similarly, for the other leg of the decay chain, if W_2 was successfully reconstructed, we join its node with the two jets used to reconstruct it, and then W_2 and b_2 are connected to the top node. The jets are directly connected to the top if there is no node for W_2. Any node originating from the remaining final states is connected to the node of the object that satisfied $\Delta R < 0.8$.

---

[3]We also checked the fully connected graph and found low performance for the given network as the number of edges increases, which carry much less physics information. Hence the decay chain-like structure in the graph gives good performance.

[4]We expect that this will lead to a further enhancement of sensitivity when the Λ^{-4} non-resonant contributions are considered.

(iii) Node features. After constructing the node and edges, we associate each node with a feature vector $[p_T, \eta, \phi, E, m, \text{PID}]$, which represent transverse momentum, pseudorapidity, azimuthal angle, energy, mass and particle identification number respectively.

3.1 Graph neural network with edge convolution

Convolution networks have seen a range of developments in the past few years. These have created the capability to employ multi-scale localised spatial features. However, Convolutional Neural Networks (CNNs) are limited to work on regular Euclidean-data like images. Recent GNN developments have overcome this limitation through generalising CNNs to operate on graph structured data, facilitating the exploration of non-Euclidean domains of the data [75]. This was formalised as Message Passing Neural Networks (MPNNs) in ref. [51] for supervised learning applications. We briefly describe the general paradigm of the MPNN, which we will generalise later for the edge convolution (EdgeConv) network used in this paper. MPNNs have two main components: a message-passing phase and a graph readout layer. The message passing is defined as a mathematical operation between two nodes i and j. We define $x_i^{(l)}$ as the ith node's features and $\vec{e}_{ij}^{(l)}$ as the edge connecting the nodes i and j at the lth time-step, where the vector sign represents the directed graph. A graph can be undirected or directed; we have used bi-directed graphs for this study. During the message-passing phase a message $\vec{m}_{ij}^{(l)}$ is calculated between the two nodes by the following operation,

$$\vec{m}_{ij}^{(l)} = \vec{M}^{(l)}(\vec{x}_i^{(l)}, \vec{x}_j^{(l)}, \vec{e}_{ij}^{(l)}). \tag{3.1}$$

The message function can be a linear activation function or a multilayer-perceptron (MLP), which is shared between the edges and is analogous to convolution operation (here we use a linear activation function for the message function). Once the messages between all connected nodes have been calculated in a layer, each node feature is updated using an aggregation function

$$\vec{x}_i^{(l+1)} = \vec{A}(\vec{x}_i^{(l)}, \{\vec{m}_{ij}^{(l)} | j \in \mathcal{N}(i)\}), \tag{3.2}$$

where $\mathcal{N}(i)$ are the nodes which are connected to ith node and $\vec{A}$ is the permutation invariant function (for instance 'max', 'sum', or 'mean'). The vector $\vec{x}^{(l+1)}$ is the input to the next message passing layer. For graph classification, after some message passing operation L we perform a permutation invariant graph readout operation $\square$ on the final node features $x_i^{(L)}$,

$$\vec{X} = \square(\vec{x}_i^{(L)} | i \in G), \tag{3.3}$$

where G denotes the input graph. This gives us fixed length representation of (possibly) variable length graphs, and feeds into a downstream neural network.

We use an EdgeConv network in this study, which is an ideally suited network for exploiting the edge features from given node features. The edge convolution operation is defined with the following message-passing function

$$\vec{x}_i^{(l+1)} = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \text{RELU}\left(\Theta(\vec{x}_j^{(l)} - \vec{x}_i^{(l)}) + \Phi(\vec{x}_i^{(l)})\right), \tag{3.4}$$

where aggregation for each node is done using 'mean', after which the features of each node are updated. The linear layers Θ and Φ take the inputs and map them to identical dimensional spaces. We use $L = 2$ and mean graph-readout.

3.2 Network architecture and training

We use the `Deep Graph Library` [76] and `PyTorch` [77] to construct the graphs and the networks that classify the different EFT signal contributions and the SM 'background'. Models with different architectures are trained on data samples that consist of 70000 events for each class, with 80%, 10% and 10% used for training, validation and testing respectively. The network models considered, incorporate EdgeConv layers followed by hidden linear layers and ReLU is used as the activation function for each layer. Probabilities for each class can be obtained from the output layer by applying the softmax function. We choose the categorical cross-entropy loss function for the multi-class classification problem and use the Adam optimiser with a learning rate of 0.001 to minimise the loss function. The learning rate decays with a factor of 0.1 if the loss function has not decreased for three consecutive epochs. We train the models for 100 epochs in mini-batches of 100 graphs and an early stopping condition when no loss decrease has occurred for ten epochs.

By varying the amount of layers and nodes, and training the different models on the data, we find that the configuration of two EdgeConv layers of 60 nodes and one hidden linear layer of 40 nodes performs particularly well for our scenario. Any event used during training or validation is not considered further in any other part of this work. The loss and accuracy curves for the classifying events have been checked to avoid overtraining. It is worth highlighting that we observe signs of overtraining when we consider deeper networks. The good performance of a relatively shallow network signifies that non-resonant physics is characterisable by relatively few phenomenological properties, which is consistent with the findings of traditional differential EFT fits (see in particular ref. [19]). This observation will form the baseline of the qualitative discussion of a two-operator example in the next section.

4 GNN-improved Wilson coefficient constraints

4.1 A minimal example

For illustration purposes, we first limit our study to a three-class classification problem. The network output in this example returns the probability of an event belonging to each of the three classes. An event is then assigned to the EFT/SM class with the greatest corresponding probability. Generalising this to a higher and critical number of WCs will be the focus of section 4.2.

The restriction employed in this section is motivated from the generic modifications that can be expected from EFT interactions. Momentum-dependent interactions will typically enhance the tails of momentum-dependent distributions compared to the SM, while interactions that modify SM couplings (feeding into, e.g., a modified top quark width) will predominantly lead to a modified inclusive rate with momentum-related distributions
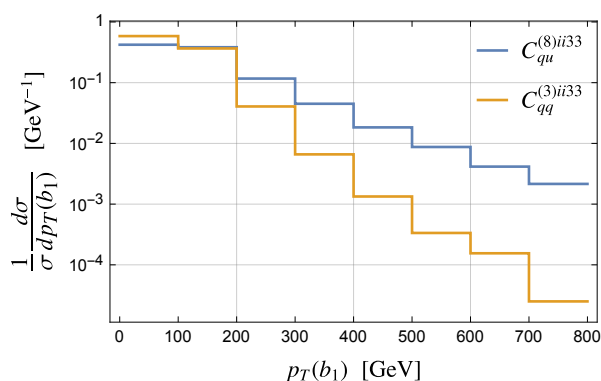
Figure 2. The normalised $p_T(b_1)$ distributions at the 13 TeV LHC for the two operators of the three-class example, eq. (4.1).
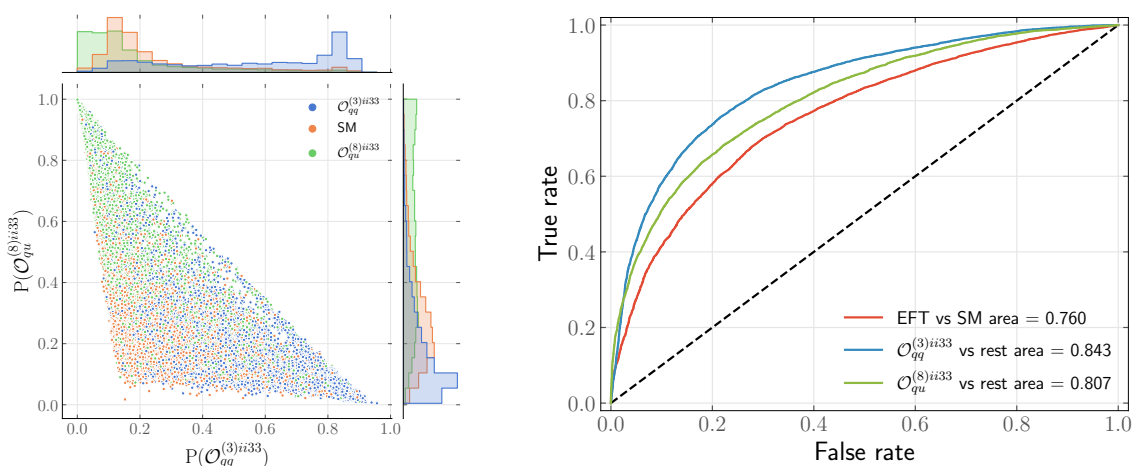

Figure 3. The probabilities calculated for each event to be a result of each SMEFT insertion is shown. On the right the Receiver Operator Characteristic (ROC) curves are shown. We calculate these in a one-vs-rest scheme for each operator.

similar to the SM. We reflect this in our choice of operators for this section:

$$\begin{aligned}
\mathcal{O}_{qu}^{(8)ii33} &= (\bar{q}_i\gamma_\mu T^A q_i)(\bar{u}_3\gamma^\mu T^A u_3)\,, \\
\mathcal{O}_{qq}^{(3)ii33} &= (\bar{q}_i\gamma_\mu \tau^I q_i)(\bar{q}_3\gamma^\mu \tau^I q_3)\,.
\end{aligned} \tag{4.1}$$

The distributions of the hardest b jet for these operators are given in figure 2. Correlated with the events hardness are more central final states and characteristically modified angular and rapidity separations. Identifying the most appropriate superposition of physical observables is therefore critical for a particularly sensitive EFT analysis. We consider the two operators of eq. (4.1) as they exhibit a particularly distinguishable phenomenology, but they will also allow us to discuss the limitations of using different approaches to a ML-informed limit setting.

In figure 3 (left), the probabilities calculated for each event to be a result of each SMEFT insertion are shown. As can be expected, events arising from $\mathcal{O}_{qu}^{8(ii33)}$ are more commonly located in the upper left region (large $\mathcal{O}_{qu}^{8(ii33)}$ probability, small $\mathcal{O}_{qq}^{3(ii33)}$ probability) while events from $\mathcal{O}_{qq}^{3(ii33)}$ are in the bottom right. In contrast, the SM events usually
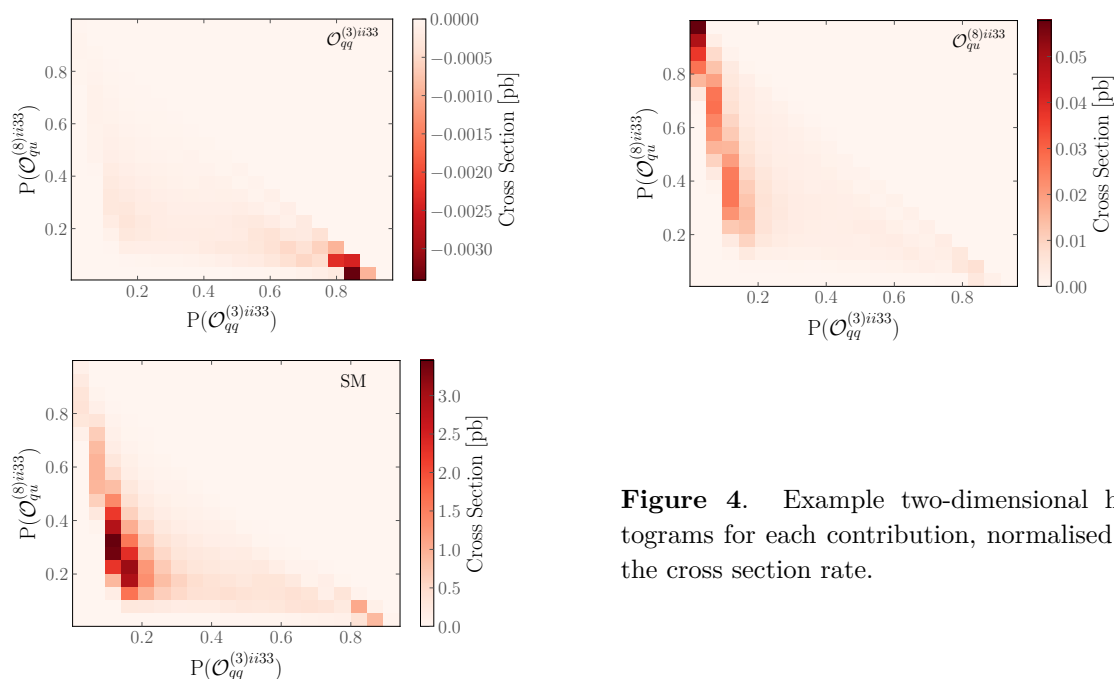
Figure 4. Example two-dimensional histograms for each contribution, normalised to the cross section rate.

end up in a region where the probabilities for $\mathcal{O}_{qu}^{8(ii33)}$ and $\mathcal{O}_{qq}^{3(ii33)}$ are both low (and the probability of belonging to the SM is high due to the normalisation of probabilities). The network is able to discriminate efficiently among the three classes and different regions can be efficiently removed by cuts on the two output probabilities. On the right in figure 3 the Receiver Operator Characteristic (ROC) curves are shown. We calculate these in a one-vs-rest scheme by first binarising the labels and using the network score output for each WC. We also show an EFT vs SM ROC curve where all EFT labels are marked as signal and the SM as background. We construct the ROC curve using the summed scores for each new physics WC, which we later generalise when more than two contributions are on.

To examine the improvement of the network performance for this simplified test case of two WCs modifying SM production, we performed a χ^2 fit for each operator to yield bounds on the WCs. To construct the χ^2 (for details see ref. [28]), we use the distribution $p_T(b_1)$, the transverse momentum of the leading b-jet. To gain as much statistical control as possible, we also extrapolate the results to an integrated luminosity of 3 ab^{-1}, in line with the expected performance of the High-Luminosity (HL) LHC. The qualitative pattern of results, however, is independent of the luminosity chosen. Performing this analysis on the full datasets gives the contours shown in black in figure 5, establishing a baseline against which we can evaluate the improvement in the constraints from applying the GNN results.

To demonstrate the power of the GNN approach, we cut on the datasets, based on the probability assigned by the network of belonging to a given class; only events with a probability greater than an optimised value of belonging to one operator class are used in the χ^2 fit. The correlation of figure 3 (left) allows us to select a threshold probability to cut on, which has the effect of substantially reducing the SM background and the contamination from the other operator, resulting in a relatively stronger signal effect and thus a tighter
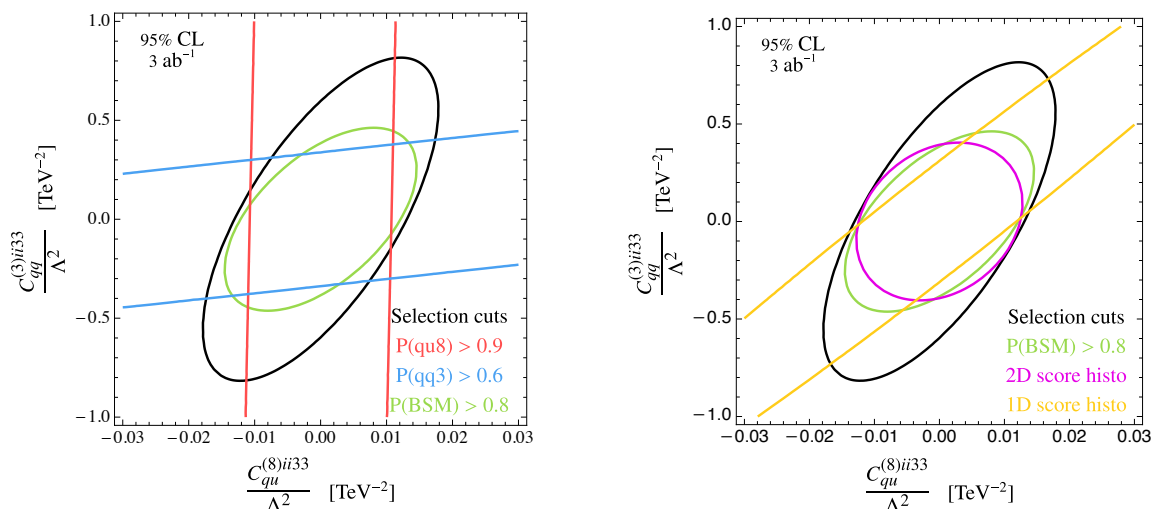
Figure 5. WC constraint contours at the 95% C.L. from χ^2 fitting; in black from the data of the baseline selection of section 2 which also passes the network requirements. The left plot shows the contours from cuts on the NN scores at the optimal value of these score cuts, with the analysis performed using $p_T(b_1)$ distributions. The right plot shows the BSM score cut as in the left plot, along with the contour from the 2D score histogram of figure 4 (with no score cuts) analysis, as well as an analysis using the 1D BSM score histogram. For details see text.

constraint on the WC for the operator for which the cut is performed. This is shown in the blue and red contours in figure 5, where the values of the cuts have been tuned to give maximal performance for each operator respectively, whilst avoiding completely depleting bins in the SM $p_T(b_1)$ distribution, as to do so would lead to unrealistic bounds on the WCs as statistical control is lost.

Due to this optimisation the bounds on individual coefficients improve, yet the other coefficient is essentially free, with expectedly far worse performance than in the original case with the full dataset. To resolve this and improve the combined bounds, we consider the probability P(BSM), which is simply the sum of the network assigned probabilities of each operator, i.e.

$$\mathrm{P(BSM)} = \mathrm{P}(\mathcal{O}_{qu}^{(8)ii33}) + \mathrm{P}(\mathcal{O}_{qq}^{(3)ii33}) \tag{4.2}$$

for the two operator classification considered here. This does indeed result in a combined bound that is superior to the original analysis.

An alternative approach to formulating constraints is to directly employ the output of the GNN, i.e. using 2D histograms of the probabilities from the network (see, for example, the individual histograms from each contribution in figure 4), in place of the $p_T(b_1)$ distributions of figure 2. A d-dimensional classification can be converted into a $d-1$ dimensional probability histogram. This can act as a template for limit setting using the information that has condensed down the phenomenologically available information into the operator classification. Considering again $\mathcal{O}_{qu}^{(8)ii33}$ and $\mathcal{O}_{qq}^{(3)ii33}$, this is demonstrated in figure 4, where the three histograms can be used to construct a χ^2, in the same way as the 1D $p_T(b_1)$ distributions, allowing the information of all three histograms to contribute. The resulting
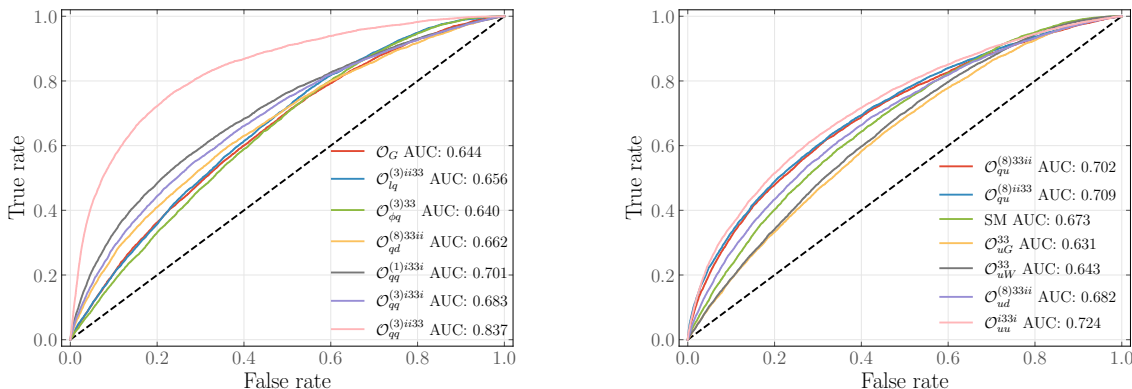
Figure 6. ROC curves for the scenario where multi-class classification is performed on thirteen SMEFT operators and the SM.

contour from this method is shown on the right plot of figure 5. This method also improves the bounds on the WCs compared to the original $p_T(b_1)$ distribution analysis with no cuts on the probabilities required. This approach is feasible when we consider only a small (sub)set of the relevant interactions. Turning to the full $d - 1$ dimensional histogram very quickly increases the statistical uncertainty. As can be seen from the qualitative similarity of the two approaches, a minimisation of

$$P(SM) = 1 - P(BSM) \tag{4.3}$$

appears to be adequate for multi-dimensional EFT analyses, particularly at luminosities below 3 ab^{-1}.

It should be noted that the one-dimensional P(BSM) histogram could be used to construct a χ^2 as well, in order to obtain the contours on the $C_{qq}^{(3)ii33} - C_{qu}^{(8)ii33}$ plane. However, the sensitivity is limited compared to the other approaches along certain directions, as shown in figure 5, due to the loss of information in the projection of the two-dimensional output to a one-dimensional score. We therefore have not explored this approach further.

4.2 Fit constraints with GNN selections

Extending the qualitative discussion of the previous section to the thirteen dimensional SMEFT parameter space, we show the Receiver Operator Characteristic (ROC) curves of the full classification in figure 6. The ROC curves are calculated with the generalised procedure discussed above. Again we see that the network[5] is capable of distinguishing operators adequately.

Starting from the baseline sensitivity as quoted in table 2 (see also section 2), we first show how contributing operators are impacted by imposing ML score cuts in figure 6. Sizeable improvements can be obtained when the momentum enhancement is present (e.g.

---

[5]By optimizing the hyperparameters for this scenario we conclude that the architecture used for the two operators case continues to perform particularly well. Deeper networks do not significantly improve the performance and often suffer from longer training times and overtraining.
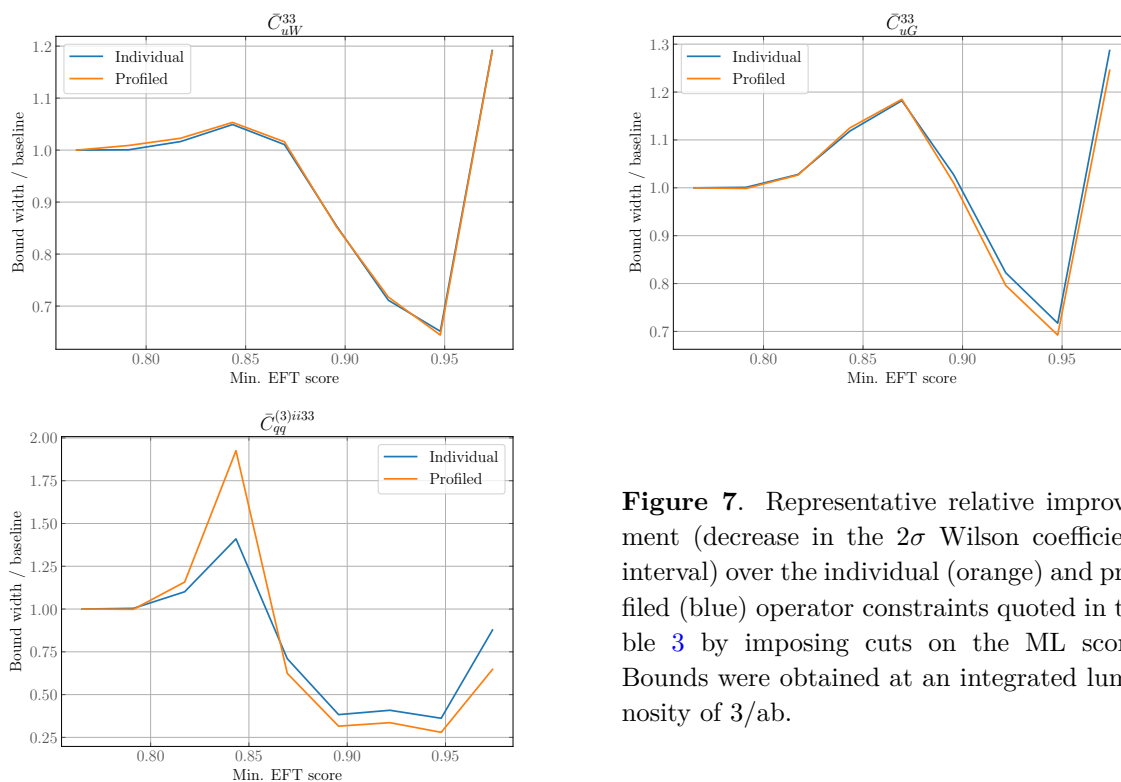
Figure 7. Representative relative improvement (decrease in the 2σ Wilson coefficient interval) over the individual (orange) and profiled (blue) operator constraints quoted in table 3 by imposing cuts on the ML score. Bounds were obtained at an integrated luminosity of 3/ab.

	2.3 fb^{-1}		3 ab^{-1}	
	Individual	Profiled	Individual	Profiled
$\bar{C}_G$	$(-0.0543, 0.0535)$	$(-0.1785, 0.1776)$	$(-0.0015, 0.0015)$	$(-0.0047, 0.0047)$
$\bar{C}_{\varphi q}^{(3)33}$	$(-0.0317, 0.0326)$	$(-0.0806, 0.0758)$	$(-0.0009, 0.0009)$	$(-0.0022, 0.0022)$
$\bar{C}_{uG}^{33}$	$(-0.0253, 0.0247)$	$(-0.0622, 0.0655)$	$(-0.0007, 0.0007)$	$(-0.0017, 0.0017)$
$\bar{C}_{uW}^{33}$	$(-0.0234, 0.0228)$	$(-0.0544, 0.0580)$	$(-0.0006, 0.0006)$	$(-0.0015, 0.0016)$
$\bar{C}_{qd}^{(8)33ii}$	$(-0.1543, 0.1558)$	$(-0.3789, 0.3698)$	$(-0.0043, 0.0043)$	$(-0.0104, 0.0104)$
$\bar{C}_{qq}^{(1)i33i}$	$(-0.0202, 0.0204)$	$(-0.0495, 0.0484)$	$(-0.0006, 0.0006)$	$(-0.0014, 0.0014)$
$\bar{C}_{qq}^{(3)i33i}$	$(-0.0101, 0.0102)$	$(-0.0247, 0.0241)$	$(-0.0003, 0.0003)$	$(-0.0007, 0.0007)$
$\bar{C}_{qq}^{(3)ii33}$	$(-3.2964, 3.3259)$	—	$(-0.0917, 0.0917)$	$(-0.3045, 0.3046)$
$\bar{C}_{qu}^{(8)33ii}$	$(-0.0867, 0.0875)$	$(-0.2127, 0.2079)$	$(-0.0024, 0.0024)$	$(-0.0058, 0.0058)$
$\bar{C}_{qu}^{(8)ii33}$	$(-0.0577, 0.0583)$	$(-0.1416, 0.1383)$	$(-0.0016, 0.0016)$	$(-0.0039, 0.0039)$
$\bar{C}_{ud}^{(8)33ii}$	$(-0.1598, 0.1613)$	$(-0.3923, 0.3824)$	$(-0.0044, 0.0044)$	$(-0.0107, 0.0107)$
$\bar{C}_{uu}^{i33i}$	$(-0.0225, 0.0228)$	$(-0.0553, 0.0540)$	$(-0.0006, 0.0006)$	$(-0.0015, 0.0015)$
$\bar{C}_{lq}^{(3)ii33}$	—	—	$(-0.3289, 0.3288)$	$(-1.8493, 1.8930)$

Table 2. Baseline 2σ bounds for different luminosities.

	2.3 fb^{-1}		3 ab^{-1}	
	Individual	Profiled	Individual	Profiled
$\bar{C}_G$	0.07%	14.12%	0.07%	11.09%
$\bar{C}_{\varphi q}^{(3)33}$	33.74%	34.19%	33.73%	33.48%
$\bar{C}_{uG}^{33}$	28.29%	32.18%	28.28%	30.74%
$\bar{C}_{uW}^{33}$	34.86%	35.35%	34.85%	35.53%
$\bar{C}_{qd}^{(8)33ii}$	4.71%	4.68%	4.71%	4.76%
$\bar{C}_{qq}^{(1)i33i}$	3.50%	3.45%	3.50%	4.73%
$\bar{C}_{qq}^{(3)i33i}$	4.35%	4.28%	4.35%	5.00%
$\bar{C}_{qq}^{(3)ii33}$	63.83%	—	63.83%	71.91%
$\bar{C}_{qu}^{(8)33ii}$	3.45%	3.51%	3.45%	3.48%
$\bar{C}_{qu}^{(8)ii33}$	3.74%	3.72%	3.74%	3.77%
$\bar{C}_{ud}^{(8)33ii}$	4.62%	4.46%	4.62%	4.79%
$\bar{C}_{uu}^{i33i}$	3.38%	3.35%	3.38%	1.95%
$\bar{C}_{lq}^{(3)ii33}$	—	—	10.57%	35.52%

Table 3. Maximum improvements in 2σ bounds via a cut on the ML score.

in case of $\bar{C}_{uG}^{33}$). Similarly, the graph network performs well in discriminating the non-resonant top decay contributions, e.g. in case $\bar{C}_{uW}^{33}$. Improvements ranging between 5% and 60% are achievable in such instances (see table 3), depending on the operators under consideration, however, always at stringent cuts on the ML score to achieve a generic BSM-sensitive selection (before losing statistical control for score cuts approaching unity). Representative operator improvements as a function of the ML score are given in figure 7. Operators showing a relatively small improvement are already under relatively good control via the inclusive rate and the baseline selection, which establishes good sensitivity to such non-SM interactions. In particular this holds for the $\bar{C}_G$ direction (which can be constrained in more adapted ways by exploiting multi-jet production [78, 79]).

Since individual constraints focus on one operator fixing the rest of the WCs to zero, it is common practice to profile over the rest of the WCs by determining their value such that the χ^2 function is minimised. In the scenario where the analysis is particularly sensitive to the presence of any additional operator, a significant decrease in sensitivity will arise. We calculate the improvement in the case of profiled WCs which, as shown in figure 7, remains similar to the individual WCs case. This is expected as the network selection removes background contributions but keeps new-physics effects. However, we note that the improvement on profiled bounds can be greater than on individual ones as in figure 7. This occurs when the cut on the EFT score selects a region where the impact on the bounds of a particular operator by the presence of additional ones is reduced, even though the robustness of one class against variations of others is not taken into account in our work.

5   Summary and outlook

The absence of direct evidence for new physics beyond the Standard Model at the LHC is as surprising as it is challenging for particle physics. Turning to effective field theory methods with the aim of fingerprinting new physics through the observation of modifications of expected SM correlations in the plethora of LHC data is a well-motivated approach to experimentally challenge, and perhaps, overcome the current status quo. The multitude of ad hoc new physics interactions in the SMEFT approach demands tailored approaches to achieve the most sensitive limit setting. In this sense, limiting analyses to a handful of, albeit motivated, differential distributions is not beneficial for enhancing the sensitivity. Conversely, employing machine learning techniques that fingerprint and exploit correlations in data provides a highly adaptive avenue to enhance the overall sensitivity that can be achieved at the LHC but also other (future) collider experiments.

In this work, we have focused employing on GNNs for EFT limit setting. GNNs are particularly motivated approaches for this purpose as they allow us to directly reflect the graph structure which is imposed by EFT interactions in the classification and eventual limit setting. We base our analysis on the semileptonic $t\bar{t}$ final states, as this is a motivated phenomenological arena for the presence of new interactions, but also because we face a critically large Wilson coefficient parameter space for multi-label classification. We find that large improvements of the sensitivity become achievable when correlations are not yet fully exploited in the inclusive base selection. This demonstrates that machine learning of multi-labelled collider data provides an excellent avenue towards improving the sensitivity of EFT-related measurements at colliders. We find that this improvement translates from individual to profiled bounds; our results also indicate a strategic approach to improve profiled constraints by tensioning operators against each other, which is not directly accessible by minimising the SM probability, but highlights the relative operator probabilities as another avenue for future investigations. Along these lines, we also note that optimisations of the ML score can be achieved via different weightings of the individual class probabilities. This way more model-specific (i.e. matched) interpretations of EFT constraints can be included to the machine learning stage, which should lead, in principle, to further sensitivity enhancements.

We note that the results of our exploratory study presented here are based on a Monte Carlo analysis; the comparison of actual data with Monte Carlo predictions is affected by a range of theoretical and experimental uncertainties. While our results do not include such uncertainties, in principle it is possible to treat them via Generative Adversarial Neural Networks, e.g. [80, 81]. Such an approach would discriminate between the different (labelled) hypotheses when the data is well-described by individual classes or superpositions of classes, effectively removing modelled uncertainty parameters from the classifier score. In general, this will lead to a decreased sensitivity compared to the idealised situation of the proof-of-principle analysis presented in this work. There are examples of such approaches to treat theoretical [48] and experimental [82] uncertainties. We leave modifications of the architecture presented in this paper along these lines for future work.

Acknowledgments

References

[1] S. Weinberg, *Phenomenological Lagrangians*, *Physica A* **96** (1979) 327 [INSPIRE].

[2] W. Buchmüller and D. Wyler, *Effective Lagrangian analysis of new interactions and flavor conservation*, *Nucl. Phys. B* **268** (1986) 621 [INSPIRE].

[3] C.J.C. Burges and H.J. Schnitzer, *Virtual effects of excited quarks as probes of a possible new hadronic mass scale*, *Nucl. Phys. B* **228** (1983) 464 [INSPIRE].

[4] C.N. Leung, S.T. Love and S. Rao, *Low-energy manifestations of a new interaction scale: operator analysis*, *Z. Phys. C* **31** (1986) 433 [INSPIRE].

[5] K. Hagiwara, R.D. Peccei, D. Zeppenfeld and K. Hikasa, *Probing the weak boson sector in* $e^+e^- \to W^+W^-$, *Nucl. Phys. B* **282** (1987) 253 [INSPIRE].

[6] B. Grzadkowski, M. Iskrzynski, M. Misiak and J. Rosiek, *Dimension-six terms in the Standard Model Lagrangian*, *JHEP* **10** (2010) 085 [arXiv:1008.4884] [INSPIRE].

[7] A. Dedes, W. Materkowska, M. Paraskevas, J. Rosiek and K. Suxho, *Feynman rules for the Standard Model effective field theory in* R_ξ-*gauges*, *JHEP* **06** (2017) 143 [arXiv:1704.03888] [INSPIRE].

[8] R. Alonso, E.E. Jenkins, A.V. Manohar and M. Trott, *Renormalization group evolution of the Standard Model dimension six operators. Part III. Gauge coupling dependence and phenomenology*, *JHEP* **04** (2014) 159 [arXiv:1312.2014] [INSPIRE].

[9] E.E. Jenkins, A.V. Manohar and M. Trott, *Renormalization group evolution of the Standard Model dimension six operators. Part II. Yukawa dependence*, *JHEP* **01** (2014) 035 [arXiv:1310.4838] [INSPIRE].

[10] E.E. Jenkins, A.V. Manohar and M. Trott, *Renormalization group evolution of the Standard Model dimension six operators. Part I. Formalism and* λ *dependence*, *JHEP* **10** (2013) 087 [arXiv:1308.2627] [INSPIRE].

[11] J. Baglio, S. Dawson and S. Homiller, *QCD corrections in Standard Model EFT fits to WZ and WW production*, *Phys. Rev. D* **100** (2019) 113010 [arXiv:1909.11576] [INSPIRE].

[12] T. Corbett, A. Helset, A. Martin and M. Trott, *EWPD in the SMEFT to dimension eight*, *JHEP* **06** (2021) 076 [arXiv:2102.02819] [INSPIRE].

[13] S. Dawson and P.P. Giardino, *New physics through Drell-Yan standard model EFT measurements at NLO*, *Phys. Rev. D* **104** (2021) 073004 [`arXiv:2105.05852`] [ɪɴSPIRE].

[14] S. Dawson and P.P. Giardino, *Flavorful electroweak precision observables in the Standard Model effective field theory*, `arXiv:2201.09887` [ɪɴSPIRE].

[15] I. Brivio and M. Trott, *The Standard Model as an effective field theory*, *Phys. Rept.* **793** (2019) 1 [`arXiv:1706.08945`] [ɪɴSPIRE].

[16] T. Corbett, O.J.P. Eboli, D. Goncalves, J. Gonzalez-Fraile, T. Plehn and M. Rauch, *The Higgs legacy of the LHC run I*, *JHEP* **08** (2015) 156 [`arXiv:1505.05516`] [ɪɴSPIRE].

[17] T. Corbett, O.J.P. Eboli, D. Goncalves, J. Gonzalez-Fraile, T. Plehn and M. Rauch, *The non-linear Higgs legacy of the LHC run I*, `arXiv:1511.08188` [ɪɴSPIRE].

[18] C. Englert, R. Kogler, H. Schulz and M. Spannowsky, *Higgs coupling measurements at the LHC*, *Eur. Phys. J. C* **76** (2016) 393 [`arXiv:1511.05170`] [ɪɴSPIRE].

[19] C. Englert, R. Kogler, H. Schulz and M. Spannowsky, *Higgs characterisation in the presence of theoretical uncertainties and invisible decays*, *Eur. Phys. J. C* **77** (2017) 789 [`arXiv:1708.06355`] [ɪɴSPIRE].

[20] J. Ellis, V. Sanz and T. You, *The effective Standard Model after LHC run I*, *JHEP* **03** (2015) 157 [`arXiv:1410.7703`] [ɪɴSPIRE].

[21] J. de Blas et al., *Higgs boson studies at future particle colliders*, *JHEP* **01** (2020) 139 [`arXiv:1905.03764`] [ɪɴSPIRE].

[22] S. Dawson, C. Englert and T. Plehn, *Higgs physics: it ain't over till it's over*, *Phys. Rept.* **816** (2019) 1 [`arXiv:1808.01324`] [ɪɴSPIRE].

[23] I. Brivio, T. Corbett and M. Trott, *The Higgs width in the SMEFT*, *JHEP* **10** (2019) 056 [`arXiv:1906.06949`] [ɪɴSPIRE].

[24] J. Ellis, C.W. Murphy, V. Sanz and T. You, *Updated global SMEFT fit to Higgs, diboson and electroweak data*, *JHEP* **06** (2018) 146 [`arXiv:1803.03252`] [ɪɴSPIRE].

[25] J. De Blas, G. Durieux, C. Grojean, J. Gu and A. Paul, *On the future of Higgs, electroweak and diboson measurements at lepton colliders*, *JHEP* **12** (2019) 117 [`arXiv:1907.04311`] [ɪɴSPIRE].

[26] J. Ellis, M. Madigan, K. Mimasu, V. Sanz and T. You, *Top, Higgs, diboson and electroweak fit to the Standard Model effective field theory*, *JHEP* **04** (2021) 279 [`arXiv:2012.02779`] [ɪɴSPIRE].

[27] A. Buckley et al., *Global fit of top quark effective theory to data*, *Phys. Rev. D* **92** (2015) 091501 [`arXiv:1506.08845`] [ɪɴSPIRE].

[28] A. Buckley et al., *Constraining top quark effective theory in the LHC run II era*, *JHEP* **04** (2016) 015 [`arXiv:1512.03360`] [ɪɴSPIRE].

[29] M. Vos et al., *Top physics at high-energy lepton colliders*, `arXiv:1604.08122` [ɪɴSPIRE].

[30] N. Castro, J. Erdmann, C. Grunwald, K. Kröninger and N.-A. Rosien, *EFTfitter — a tool for interpreting measurements in the context of effective field theories*, *Eur. Phys. J. C* **76** (2016) 432 [`arXiv:1605.05585`] [ɪɴSPIRE].

[31] D. Barducci et al., *Interpreting top-quark LHC measurements in the standard-model effective field theory*, `arXiv:1802.07237` [ɪɴSPIRE].

[32] N.P. Hartland et al., *A Monte Carlo global analysis of the Standard Model effective field theory: the top quark sector*, *JHEP* **04** (2019) 100 [arXiv:1901.05965] [INSPIRE].

[33] I. Brivio et al., *O new physics, where art thou? A global search in the top sector*, *JHEP* **02** (2020) 131 [arXiv:1910.03606] [INSPIRE].

[34] SMEFIT collaboration, *Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC*, *JHEP* **11** (2021) 089 [arXiv:2105.00006] [INSPIRE].

[35] J. Brehmer, F. Kling, T. Plehn and T.M.P. Tait, *Better Higgs-CP tests through information geometry*, *Phys. Rev. D* **97** (2018) 095017 [arXiv:1712.02350] [INSPIRE].

[36] J. Brehmer, K. Cranmer, F. Kling and T. Plehn, *Better Higgs boson measurements through information geometry*, *Phys. Rev. D* **95** (2017) 073002 [arXiv:1612.05261] [INSPIRE].

[37] J. Brehmer, K. Cranmer, G. Louppe and J. Pavez, *Constraining effective field theories with machine learning*, *Phys. Rev. Lett.* **121** (2018) 111801 [arXiv:1805.00013] [INSPIRE].

[38] J. Brehmer, K. Cranmer, G. Louppe and J. Pavez, *A guide to constraining effective field theories with machine learning*, *Phys. Rev. D* **98** (2018) 052004 [arXiv:1805.00020] [INSPIRE].

[39] J. D'Hondt, A. Mariotti, K. Mimasu, S. Moortgat and C. Zhang, *Learning to pinpoint effective operators at the LHC: a study of the $t\bar{t}b\bar{b}$ signature*, *JHEP* **11** (2018) 131 [arXiv:1807.02130] [INSPIRE].

[40] S. Chatterjee, N. Frohner, L. Lechner, R. Schöfbeck and D. Schwarz, *Tree boosting for learning EFT parameters*, arXiv:2107.10859 [INSPIRE].

[41] J.Y. Araz and M. Spannowsky, *Combine and conquer: event reconstruction with Bayesian ensemble neural networks*, *JHEP* **04** (2021) 296 [arXiv:2102.01078] [INSPIRE].

[42] P. Konar, V.S. Ngairangbam and M. Spannowsky, *Energy-weighted message passing: an infra-red and collinear safe graph neural network algorithm*, *JHEP* **02** (2022) 060 [arXiv:2109.14636] [INSPIRE].

[43] G. Karagiorgi, G. Kasieczka, S. Kravitz, B. Nachman and D. Shih, *Machine learning in the search for new fundamental physics*, arXiv:2112.03769 [INSPIRE].

[44] C.W. Bauer, M. Freytsis and B. Nachman, *Simulating collider physics on quantum computers using effective field theories*, *Phys. Rev. Lett.* **127** (2021) 212001 [arXiv:2102.05044] [INSPIRE].

[45] M. Feickert and B. Nachman, *A living review of machine learning for particle physics*, arXiv:2102.02770 [INSPIRE].

[46] E.A. Moreno et al., *JEDI-net: a jet identification algorithm based on interaction networks*, *Eur. Phys. J. C* **80** (2020) 58 [arXiv:1908.05318] [INSPIRE].

[47] J. Pata, J. Duarte, J.-R. Vlimant, M. Pierini and M. Spiropulu, *MLPF: efficient machine-learned particle-flow reconstruction using graph neural networks*, *Eur. Phys. J. C* **81** (2021) 381 [arXiv:2101.08578] [INSPIRE].

[48] C. Englert, P. Galler, P. Harris and M. Spannowsky, *Machine learning uncertainties with adversarial neural networks*, *Eur. Phys. J. C* **79** (2019) 4 [arXiv:1807.08763] [INSPIRE].

[49] J. Zhou et al., *Graph neural networks: a review of methods and applications*, arXiv:1812.08434.

[50] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang and P.S. Yu, *A comprehensive survey on graph neural networks*, *IEEE Trans. Neural Networks Learn. Syst.* **32** (2021) 4 [arXiv:1901.00596].

[51] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals and G.E. Dahl, *Neural message passing for quantum chemistry*, in *International conference on machine learning*, PMLR, (2017), p. 1263 [arXiv:1704.01212].

[52] Y. Wang, Y. Sun, Z. Liu, S.E. Sarma, M.M. Bronstein and J.M. Solomon, *Dynamic graph CNN for learning on point clouds*, *ACM Trans. Graph.* **38** (2019) 1 [arXiv:1801.07829] [INSPIRE].

[53] F.A. Dreyer and H. Qu, *Jet tagging in the Lund plane with graph networks*, *JHEP* **03** (2021) 052 [arXiv:2012.08526] [INSPIRE].

[54] A. Blance and M. Spannowsky, *Unsupervised event classification with graphs on classical and photonic quantum computers*, *JHEP* **08** (2021) 170 [arXiv:2103.03897] [INSPIRE].

[55] O. Atkinson, A. Bhardwaj, C. Englert, V.S. Ngairangbam and M. Spannowsky, *Anomaly detection with convolutional graph neural networks*, *JHEP* **08** (2021) 080 [arXiv:2105.07988] [INSPIRE].

[56] H. Qu and L. Gouskos, *ParticleNet: jet tagging via particle clouds*, *Phys. Rev. D* **101** (2020) 056019 [arXiv:1902.08570] [INSPIRE].

[57] T. Dorigo, M. Fumanelli, C. Maccani, M. Mojsovska, G.C. Strong and B. Scarpa, *RanBox: anomaly detection in the Copula space*, arXiv:2106.05747 [INSPIRE].

[58] V. Mikuni and F. Canelli, *ABCNet: an attention-based method for particle tagging*, *Eur. Phys. J. Plus* **135** (2020) 463 [arXiv:2001.05311] [INSPIRE].

[59] O. Knapp, O. Cerri, G. Dissertori, T.Q. Nguyen, M. Pierini and J.-R. Vlimant, *Adversarially learned anomaly detection on CMS open data: re-discovering the top quark*, *Eur. Phys. J. Plus* **136** (2021) 236 [arXiv:2005.01598] [INSPIRE].

[60] V. Mikuni and F. Canelli, *Unsupervised clustering for collider physics*, *Phys. Rev. D* **103** (2021) 092007 [arXiv:2010.07106] [INSPIRE].

[61] G. DeZoort et al., *Charged particle tracking via edge-classifying interaction networks*, *Comput. Softw. Big Sci.* **5** (2021) 26 [arXiv:2103.16701] [INSPIRE].

[62] M. Abdughani, J. Ren, L. Wu and J.M. Yang, *Probing stop pair production at the LHC with graph neural networks*, *JHEP* **08** (2019) 055 [arXiv:1807.09088] [INSPIRE].

[63] J. Duarte and J.-R. Vlimant, *Graph neural networks for particle tracking and reconstruction*, arXiv:2012.01249 [INSPIRE].

[64] X. Ju and B. Nachman, *Supervised jet clustering with graph neural networks for Lorentz boosted bosons*, *Phys. Rev. D* **102** (2020) 075014 [arXiv:2008.06064] [INSPIRE].

[65] CMS collaboration, *Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at* 13 *TeV*, *Phys. Rev. D* **95** (2017) 092001 [arXiv:1610.04191] [INSPIRE].

[66] I. Brivio, Y. Jiang and M. Trott, *The SMEFTsim package, theory and tools*, *JHEP* **12** (2017) 070 [arXiv:1709.06492] [INSPIRE].

[67] I. Brivio, *SMEFTsim* 3.0 — *a practical guide*, *JHEP* **04** (2021) 073 [arXiv:2012.11343] [INSPIRE].

[68] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[69] A. Alloul, N.D. Christensen, C. Degrande, C. Duhr and B. Fuks, *FeynRules 2.0 — a complete toolbox for tree-level phenomenology*, *Comput. Phys. Commun.* **185** (2014) 2250 [arXiv:1310.1921] [INSPIRE].

[70] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer and T. Reiter, *UFO — the Universal FeynRules Output*, *Comput. Phys. Commun.* **183** (2012) 1201 [arXiv:1108.2040] [INSPIRE].

[71] S. Brown et al., *Updated results from TopFitter*, to appear.

[72] A. Buckley et al., *Rivet user manual*, *Comput. Phys. Commun.* **184** (2013) 2803 [arXiv:1003.0694] [INSPIRE].

[73] C. Bierlich et al., *Robust independent validation of experiment and theory: Rivet version 3*, *SciPost Phys.* **8** (2020) 026 [arXiv:1912.05451] [INSPIRE].

[74] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[75] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, *Gradient-based learning applied to document recognition*, *Proc. IEEE* **86** (1998) 2278.

[76] M. Wang et al., *Deep graph library: a graph-centric, highly-performant package for graph neural networks*, arXiv:1909.01315.

[77] A. Paszke et al., *Pytorch: an imperative style, high-performance deep learning library*, arXiv:1912.01703.

[78] H.K. Dreiner, A. Duff and D. Zeppenfeld, *How well do we know the three gluon vertex?*, *Phys. Lett. B* **282** (1992) 441 [INSPIRE].

[79] L.J. Dixon and Y. Shadmi, *Testing gluon selfinteractions in three jet events at hadron colliders*, *Nucl. Phys. B* **423** (1994) 3 [*Erratum ibid.* **452** (1995) 724] [hep-ph/9312363] [INSPIRE].

[80] G. Louppe, M. Kagan and K. Cranmer, *Learning to pivot with adversarial networks*, arXiv:1611.01046 [INSPIRE].

[81] R. Kansal et al., *Particle cloud generation with message passing generative adversarial networks*, arXiv:2106.11535 [INSPIRE].

[82] M. Bellagente, A. Butter, G. Kasieczka, T. Plehn and R. Winterhalder, *How to GAN away detector effects*, *SciPost Phys.* **8** (2020) 070 [arXiv:1912.00477] [INSPIRE].