Amplitude-assisted tagging of longitudinally polarised bosons using wide neural networks

Michele Grossi[1]*, Massimiliano Incudini[2]†, Mathieu Pellen[3]‡ Giovanni Pelliccioli[4]§

[1] *European Organisation for Nuclear Research (CERN)*
Geneva 1211, Switzerland

[2] *Department of Computer Science, University of Verona,*
Strada Le Grazie 15, 37129 Verona, Italy

[3] *Physikalisches Institut, Universität Freiburg,*
Hermann-Herder-Str. 3, 79104 Freiburg, Germany

[4] *Max-Planck-Institut für Physik,*
Föhringer Ring 6, 80805 München, Germany

Abstract

Extracting longitudinal modes of weak bosons in LHC processes is essential to understand the electroweak-symmetry-breaking mechanism. To that end, we propose a general method, based on wide neural networks, to properly model longitudinal-boson signals and hence enable the event-by-event tagging of longitudinal bosons. It combines experimentally accessible kinematic information and genuine theoretical inputs provided by amplitudes in perturbation theory. As an application we consider the production of a Z boson in association with a jet at the LHC, both at leading order and in the presence of parton-shower effects. The devised neural networks are able to extract reliably the longitudinal contribution to the unpolarised process. The proposed method is very general and can be systematically extended to other processes and problems.

---

*E-mail: michele.grossi@cern.ch
†E-mail: massimiliano.incudini@univr.it
‡E-mail: mathieu.pellen@physik.uni-freiburg.de
§E-mail: gpellicc@mpp.mpg.de (corresponding author)

arXiv:2306.07726v2 [hep-ph] 30 Aug 2023

Contents

1 Introduction

Accessing the polarisation of electroweak (EW) bosons at high-energy colliders is crucial to gain insights in the electroweak-symmetry-breaking mechanism (EWSB), whose nature is currently explained by the Higgs mechanism [1–3]. By means of the EWSB, the W and Z bosons are given a mass and a longitudinal-polarisation state. Therefore, any deviation in the production of longitudinal bosons in scattering processes would suggest the presence of new-physics effects, implying a different realisation of the EWSB compared to the Standard Model (SM) one.

The investigation of polarised-boson signals in LHC processes is becoming an important part of the analysis programme of the ATLAS and CMS collaborations with Run-2 data, as shown in recent measurements of di-boson production and vector-boson scattering (VBS) [4–7]. The increase in statistics of Run-3 and the High-Luminosity phase will drastically improve the precision of current analyses and give access to polarised signals in complex multi-boson processes [8–10].

The analysis paradigm for the measurement of polarisations of EW bosons with Run-1 data at 7/8 TeV was the evaluation of angular coefficients of the boson decay rate, which are related to the polarisation fractions. The extraction of angular coefficients in LHC processes was proposed in seminal phenomenological works [11–13] and applied in experimental analyses of W+j [14, 15] and Z+j [16, 17] events, as well as of top-quark decays [18–20]. Owing to its simplicity, this strategy has been further investigated and extended in more recent phenomenological studies [21–28]. However, its application is limited [11–13, 25, 26, 29] to inclusive decays (*i.e.* without selections on single decay products of bosons) and to two-body decays [*i.e.* without radiative corrections to the decay (*e.g.* EW corrections)].

More recently, a different approach was proposed [29] to interpret LHC Run-2 data in terms of polarisations of EW bosons, relying on the direct simulation of intermediate polarised bosons in Monte Carlo codes. This method has been automated and applied to several processes at leading order [29–32] and extended to higher-order EW and QCD corrections [33–39]. The existing results in the literature in this direction concern vector-boson scattering (VBS) [29–32], inclusive di-boson [33–36, 38–40], Higgs-boson decays [41–43] and W+j production [37]. The great advantage of this approach is that, upon subtraction of backgrounds, LHC signal events can be fitted with polarised templates in order to extract polarisation fractions in a differential way from the LHC data, accounting properly for interference effects and spin correlations. This has become the new analysis paradigm with Run-2 data, as demonstrated by the pioneering measurements performed by ATLAS and CMS in di-boson inclusive production [4, 6, 7] and VBS [5].

A number of recent studies have been carried out targeting the polarisation extraction in the presence of hadronic decays of the weak bosons [44–48, 40], both with the polarised-template method and with machine-learning techniques. The usage of machine learning was also proposed to extract polarisation fractions in VBS, starting from the kinematic structure of the events [49–53, 45].

Independently of the specific approach that may be used for the interpretation, the polarisation state of an unstable particle, like EW bosons, is not directly accessible in the detectors. Therefore, the information about it can only be reconstructed (in a probabilistic way) from the stable decay products. In other words, the polarisation of EW bosons is a pseudo-observable. On the other hand, from a theoretical perspective, the whole information regarding the fundamental quantum field theory and therefore the polarisation properties is encoded in the amplitude. Therefore, accessing the amplitude of scattering processes at experiments would give the maxi-

mal information possible, *i.e.* the maximal predictive power. However, in a realistic experimental environment only the momenta of the visible final states can be reconstructed. This means that the momenta of the initial states as well as their parton type, both needed for the exact evaluation of the amplitude, is a priori unknown. As explained in the rest of the article, machine learning (ML) can actually be used to approximate well the amplitudes based only on the partial information available experimentally.

It is worth pointing out that in this article we exclusively refer to *amplitudes* and not *matrix elements*, which are actually equivalent for our purposes. Our method, though close in spirit, should not to be confused with the *matrix-element method* [54, 55] and the *optimal-observable method* [56–58].

The article is organised as follows. In Sect. 2 we explain the difficulties in tagging longitudinal bosons and propose a solution based on wide neural networks and amplitudes. A concrete application of the proposed method to Z+j at the LHC is then detailed and discussed in Sect. 3. In Sect. 4 we draw the conclusions of our work.

2 Tagging longitudinal bosons

2.1 Definition of the problem

A generic (unpolarised) amplitude featuring a resonant gauge boson decaying into a lepton-neutrino pair can be written as follows (in the unitary gauge),

$$\mathcal{M} = \mathcal{M}_\mu^{\mathcal{P}} \frac{i}{k^2 - M_V^2 + i\Gamma_V M_V} \left(-g^{\mu\nu} + \frac{k^\mu k^\nu}{M_V^2} \right) \mathcal{M}_\nu^{\mathcal{D}}, \tag{1}$$

where $\mathcal{M}^{\mathcal{P}}$ and $\mathcal{M}^{\mathcal{D}}$ describe the production and decay part of the amplitude, respectively. The quantities M_V and Γ_V represent the gauge-boson mass and width, respectively. In particular, the tensor part of the propagator can be cast into the following form,

$$-g^{\mu\nu} + \frac{k^\mu k^\nu}{M^2} = \sum_{\lambda=1}^{4} \varepsilon_\lambda^\mu(k)\varepsilon_\lambda^{\nu*}(k) , \tag{2}$$

where the $\{\varepsilon_\lambda^\mu(k)\}$ represent polarisation vectors of the massive gauge boson. The sum runs over four polarisation states, namely the three physical states and a fourth one, whose structure depends on the EW-gauge choice, and is thus is unphysical[1]. Throughout the article, we use the labels L, +, and − for the longitudinal, right-handed, and left-handed states, respectively. Notice that the polarisation vectors are defined in such a way that they are transverse w.r.t. the boson four momentum, but do not transform as Lorentz covariants, therefore they must be defined in a specific Lorentz frame. Also, we would like to emphasise that the definition of the polarisation of the massive gauge bosons is only meaningful when gauge bosons are on the mass shell or when the resonant contributions are treated in the narrow-width [59, 60] or pole approximation [61–66]. The reason for this is to guarantee gauge invariance.

The amplitude in Eq. (1), including both production and decay parts, can therefore be written as,

$$\mathcal{M} = \sum_\lambda \mathcal{M}_\lambda , \qquad \lambda = \text{L}, +, - , \tag{3}$$

---

[1]It is important to notice that this unphysical contribution always cancels out against Goldstone-boson contributions at any order in perturbation theory.

where $\mathcal{M}_\lambda$ is the amplitude with a polarised intermediate gauge boson (with state λ),

$$\mathcal{M}_\lambda = \left[\mathcal{M}_\mu^{\mathcal{P}} \varepsilon_\lambda^\mu(k)\right] \frac{i}{k^2 - M_V^2 + i\Gamma_V M_V} \left[\varepsilon_\lambda^{\nu^*}(k)\mathcal{M}_\nu^{\mathcal{D}}\right]. \tag{4}$$

Hence, squaring the unpolarised amplitude leads to

$$|\mathcal{M}|^2 = \sum_\lambda |\mathcal{M}_\lambda|^2 + \sum_{\lambda \neq \lambda'} \mathcal{M}_\lambda^* \mathcal{M}_{\lambda'}, \qquad \lambda, \lambda' = \mathrm{L}, +, -, \tag{5}$$

where the first sum represents the incoherent sum over polarised squared amplitudes, while the second one includes all interference terms. For phenomenological purposes it is convenient [30] to define a transverse (T) contribution as the coherent sum of the left- and right-handed contributions to the squared amplitude,

$$|\mathcal{M}_\mathrm{T}|^2 = |\mathcal{M}_+|^2 + |\mathcal{M}_-|^2 + 2\,\mathrm{Re}\left(\mathcal{M}_+^* \mathcal{M}_-\right), \tag{6}$$

leading to a simpler structure of Eq. (5),

$$|\mathcal{M}|^2 = |\mathcal{M}_\mathrm{L}|^2 + |\mathcal{M}_\mathrm{T}|^2 + 2\,\mathrm{Re}\left(\mathcal{M}_\mathrm{L}^* \mathcal{M}_+\right) + 2\,\mathrm{Re}\left(\mathcal{M}_\mathrm{L}^* \mathcal{M}_-\right). \tag{7}$$

The term $|\mathcal{M}_\mathrm{L}|^2$ defines the longitudinally polarised squared amplitude which is the focus of the present work. Note that the interference terms of Eq. (5) [or of Eq. (7)] are in general non vanishing and can take either positive or negative values.

Hence, the fully differential unpolarised and longitudinal cross sections schematically read,

$$\mathrm{d}\sigma_\mathrm{unp} = \frac{1}{F}\,|\mathcal{M}|^2\,\mathrm{d}\Phi, \qquad \mathrm{d}\sigma_\mathrm{L} = \frac{1}{F}\,|\mathcal{M}_\mathrm{L}|^2\,\mathrm{d}\Phi, \tag{8}$$

where the flux factor and phase-space measure are denoted by F and $\mathrm{d}\Phi$, respectively. The differential longitudinal fraction in a generic observable $\mathcal{O}$, that is ought to be extracted experimentally, is therefore defined as,

$$f_\mathrm{L}(\mathcal{O}) = \frac{\mathrm{d}\sigma_\mathrm{L}}{\mathrm{d}\mathcal{O}} \bigg/ \frac{\mathrm{d}\sigma_\mathrm{unp}}{\mathrm{d}\mathcal{O}}. \tag{9}$$

Problem The main challenge is to extract the longitudinal fraction experimentally for arbitrary observables, *i.e.* in a fully differential way. In other words, we would like to answer the question: how can we infer on an event-by-event basis the probability for an LHC event to be longitudinally polarised? In what follows, we will address exactly this problem.

As briefly mentioned in Sect. 1, a number of methods have been proposed in past and recent years to address the issue of longitudinal-event tagging. In the rest of this section, we review some of them.

For the interpretation of LHC Run-1 data, the *angular-coefficient method* was typically applied with polarisation-extraction purposes. It relies on the functional structure of the tree-level decay rate of EW bosons, that can be written as follows [11, 12],

$$\begin{aligned}
\frac{\mathrm{d}^3\sigma}{\mathrm{d}\cos\theta^*\,\mathrm{d}\phi^*\,\mathrm{d}\mathcal{O}} =\ & \frac{\mathrm{d}\sigma}{\mathrm{d}\mathcal{O}}\frac{3}{16\pi}\Bigg[1 + \cos^2\theta^* + A_0\frac{1 - 3\cos^2\theta^*}{2} + A_4\cos\theta^* + 2A_1\cos\theta^*\sin\theta^*\cos\phi^* \\
& + 2A_6\cos\theta^*\sin\theta^*\sin\phi^* + A_3\sin\theta^*\cos\phi^* + A_5\sin\theta^*\sin\phi^* \\
& + \frac{1}{2}A_2\sin^2\theta^*\cos 2\phi^* + A_7\sin^2\theta^*\sin 2\phi^*\Bigg],
\end{aligned} \tag{10}$$

4

where θ^* and ϕ^* are the polar and azimuthal decay angle of a decay lepton in the decayed-boson rest frame, calculated w.r.t. the boson trajectory in a certain Lorentz frame (the one where polarisation states are defined). The coefficients $\{A_0, \ldots, A_7\}$, which are functions of the observable $\mathcal{O}$ (independent of decay angles), are related to polarisation fractions f_L and $f_\pm$ via linear combinations. Projecting Eq. (10) onto suitable spherical harmonics of rank 2, the polarisation fractions can be then easily extracted. This strategy is valid for a single boson, in the absence of radiative corrections to the decay, and in a fully inclusive decay-phase-space measure, *i.e.* without any cut on individual decay products. Applying the projections in the presence of transverse-momentum and/or rapidity cuts on decay products, as done in many experimental analyses, may give results that are far from describing the polarisation structure of a process [12, 13, 29, 22, 26]. Extensions of the method to account for multi-boson spin correlations [23, 24, 27, 67] are also limited due to similar reasons.

The extraction of angular coefficients from decay rates can also be applied differentially in any LHC observables, providing a way to reweight unpolarised LHC events according to polarisation fractions and therefore split the events into longitudinal and transverse samples. This approximate method has been applied for Run-1 $V + j$ events [14–17] but was proven to fail in certain kinematic regimes, especially due to the wrong assumption that polarisation fractions are the same in the presence and absence of decay-product selections [30].

The most prominent way to extract polarised signals out of Run-2 LHC data is the so-called *polarised-template method*. Building on a theoretically sound definition of polarised signals at amplitude level [29] that can be systematically extended to higher orders in perturbation theory [33–39], the method relies on separate templates for each physical polarisation state and for the interference terms. Upon a previous subtraction of reducible and irreducible backgrounds, the (unpolarised) signal events are simultaneously fitted with fully independent polarised templates. This can be applied differentially to any LHC observable. In practice, these fits are restricted to those observables that are thought to be the most sensitive to discriminate between longitudinal and transverse modes. This does not guarantee that the polarisation information is fully exploited from the available data. In addition, the fitting procedure requires quite intensive theoretical calculations that only recently achieved (N)NLO accuracy [33–39]. To be of any use, such calculations should be performed in a fiducial phase-space volume which is exactly the one used in the experimental analysis, a task that can turn out to be far from trivial.

The idea of using ML methods to facilitate the extraction of polarisation fractions has been already explored in the literature. It has been applied in the case of EW bosons produced in VBS and inclusive di-boson production, both with leptonic [49–51, 45, 52, 53] and hadronic decays [45, 46]. The proposed ML approaches typically rely on kinematic observables approximating decay angles in the case of leptonic decays of W bosons, and on jet-substructure observables to treat hadronic decays, with the aim of performing an event-by-event classification, possibly accounting for new-physics effects that may distort the underlying dynamics.

The method we propose to tag longitudinal bosons lies somehow at the intersection amongst the aforementioned methods, complementing accessible kinematic information of LHC events with a genuine theoretical input given by amplitudes describing the process dynamics.

As a last remark before detailing our strategy, we stress that the polarisation structure of a process is model dependent, owing to possibly different dynamics at production and decay level. The advantage of the solution we present in the following is that the model dependence is uniquely encoded in the amplitudes.

2.2 A machine-learning-based solution

Equation (9) implies that at the phase-space–point level, the polarisation fraction is equal to the ratio of the longitudinally polarised squared amplitude over the unpolarised one,

$$r_{\mathrm{L}} = \frac{|\mathcal{M}_{\mathrm{L}}|^2}{|\mathcal{M}|^2}. \tag{11}$$

This statement is exact at leading order (LO) in perturbation theory for each partonic channel occurring in the process. It means that computing r_{L} for a given process requires the knowledge of the full kinematics as well as of the flavours of the partonic-process external particles.

Considering an unpolarised event sample, this ratio can be computed for each event separately. One can therefore tag each event as longitudinally polarised or not by sampling on the value of r_{L}. From the unpolarised event sample, one can therefore obtain a longitudinally polarised sample. This procedure is completely equivalent to generating a longitudinal sample from scratch. It follows that the fully differential knowledge of r_{L} allows for the assessment of longitudinal polarisation on a even-by-event basis. We note that while Eq. (5) does not guarantee this ratio to be comprised between 0 and 1, in practice it is and allows a straightforward sampling without requiring to determine the minimum and maximum of r_{L} beforehand.

As mentioned above, this procedure is exact at LO accuracy. It can actually be extended to a LO sample with parton-shower (PS) corrections using a similar procedure. Considering an unpolarised sample at LO+PS accuracy, one can compute r_{L} with the original event (before showering) and tag the event after showering based on the value of r_{L}. Again, this procedure is equivalent to generate a longitudinal sample at LO+PS accuracy from scratch[2].

Key concept Given the possibility to evaluate r_{L}, one can tag events as longitudinally polarised. While doable theoretically, this is unfortunately not possible experimentally as the evaluation of r_{L} at the event level requires the knowledge of all momenta and flavours of the initial and final partons, an information which is not available experimentally. Instead, what is available experimentally is the knowledge of the final-state momenta which constitutes therefore only a partial information. The central idea is therefore to bypass this lack of information by using a neural network (NN) to obtain an approximate value of r_{L} called $\tilde{r}_{\mathrm{L}}$ which in turn depends only on the experimental information available. In other words, the NN is trained to mimic r_{L} with an incomplete information. Later, we show that this method is applicable in practice.

It implies, therefore, that one can use $\tilde{r}_{\mathrm{L}}$ to tag experimental events as longitudinal. The longitudinal fraction extracted in this way can then be compared against theoretical predictions. If r_{L} is computed within the SM, an agreement between the theoretical predictions and the extracted value of the fraction indicates that the data is compatible with SM expectations. On the other hand, a disagreement would be the sign of a failure of the SM to describe the physics at hand. The procedure can be applied not only to the SM but also to any *UV-complete model* as well as to rather model-independent frameworks like *simplified models* or *effective-field theories*.

The method we propose does not require any fitting procedure. It is by definition multidimensional and therefore ensures that all possible information available experimentally is used. It is also very flexible with respect to the phase-space requirements. In fact, if the training is

---

[2]Note that instead of Eq. (11), we have also considered a variation of it, namely: $r_{\mathrm{D,L}} = |\mathcal{M}_{\mathrm{D,L}}|^2 / |\mathcal{M}_{\mathrm{D}}|^2$, where the subscript D denotes the decay of the gauge bosons. It turned out that this variable is not sufficient to reproduce the full longitudinal fraction.

done in an inclusive phase space, the trained model can be used in any fiducial volume that is more restrictive.

In summary, the key idea of our approach is to relate the tagging of LHC events to a single theoretically clean quantity, using machine learning to cope with incomplete information.

3   Application: Z+jet

In order to illustrate the newly devised method, we apply it to the extraction of the longitudinal polarisation of a Z boson in Z+j production at the LHC. We would like to emphasis that, in spite of possibly different input features for the training of the NN for a different process, our method is fully general and can be applied to any process featuring one or several Z or W boson(s). The process we consider is,

$$pp \to j + Z(\to \mu^+ \mu^-) + X \,. \tag{12}$$

While providing a non-trivial test-bed, this reaction is particularly suited for polarisation studies as it has a very high cross section and allows for the full reconstruction of the final state.

Note that the production of a muon-antimuon pair is mediated by a photon and a Z boson. However, since we aim at extracting the polarisation of an intermediate Z boson, the photon contribution is regarded as an irreducible background to be subtracted before any polarisation analysis [36, 40]. In the presence of a cut on the lepton-pair invariant mass around the Z pole mass, the photon background (as well as the photon-Z interference) is typically small. In the setups considered here (see Sect. 3.1), this irreducible background is at the level of 1%, estimated from a comparison between the Z-mediated signal of Eq. (12) and the full off-shell calculation of $pp \to j + \mu^+ \mu^-$.

Since we are interested in polarised signals, we choose to define polarisation vectors in the Lorentz frame where the Z boson and the jet are back to back, which coincides (at LO) with the partonic centre-of-mass frame. This reference frame is the one where the $2 \to 2$ scattering happens and can be entirely reconstructed up to experimental uncertainties. Therefore, this choice is well motivated both from a theoretical and from an experimental viewpoint. We stress that any polarisation extraction from simulated events or experimental data is frame dependent. This means that, although the general strategy we propose can be applied to any polarisation-frame definition, the application considered here depends on the specific choice of the polarisation frame. In practice, the r_L quantity defined in Eq. (11) takes different values when computed for the same phase-space point but for different polarisation-frame choices, therefore the NN-training stage is tailored to the specific choice of polarisation frame.

3.1   Input parameters and event selections

In this section, we list the input SM parameters used for the numerical computations and the event selections considered for the phenomenological analysis.

The simulations are performed at a centre-of-mass energy of $\sqrt{s} = 13.6 \, \text{TeV}$ for proton-proton collisions at the LHC. The parton distribution function `NNPDF31_nlo_as_0118` [68] has been utilised thanks to LHAPDF [69]. The renormalisation and factorisation scales are fixed to

$$\mu_R = \mu_F = M_Z. \tag{13}$$

The EW coupling is fixed through the G_μ scheme [65, 70] is used for the electroweak coupling as

$$\alpha = \frac{\sqrt{2}}{\pi} G_\mu M_W^2 \left(1 - \frac{M_W^2}{M_Z^2} \right) \qquad \text{with} \qquad G_\mu = 1.16639 \times 10^{-5} \, \text{GeV}^{-2} \,. \tag{14}$$

7

The following masses and widths have been taken,

$$M_Z = 91.188\,\text{GeV}, \qquad \Gamma_Z = 2.49877\,\text{GeV},$$
$$M_W = 80.419\,\text{GeV}, \qquad \Gamma_W = 2.09291\,\text{GeV}. \tag{15}$$

The masses or widths of all other particles do not play a role in this process or have been set to zero. Note that these parameters are essentially the default ones in MG5_AMC@NLO [71].

A number of different event selections are used in this work. The first one, which we label *generation-level*, is characterised by a transverse-momentum and rapidity cut on the leading jet, as well as an invariant-mass cut on the charged-lepton pair,

$$p_{\text{T,j}} > 10\,\text{GeV}, \qquad |y_{\text{j}}| < 5, \qquad \text{and} \qquad 76\,\text{GeV} < M_{\mu^+\mu^-} < 106\,\text{GeV}. \tag{16}$$

With this setup we have generated the initial parton-level event samples for both unpolarised and longitudinally polarised Z bosons.

The second selection, labeled *inclusive*, used for some of the phenomenological results with and without PS effects, is characterised by slightly more restrictive cuts, in order to avoid biasing the PS application, namely,

$$p_{\text{T,j}} > 20\,\text{GeV}, \qquad |y_{\text{j}}| < 4, \qquad \text{and} \qquad 81\,\text{GeV} < M_{\mu^+\mu^-} < 101\,\text{GeV}. \tag{17}$$

Notice that both the *generation-level* and *inclusive* setups avoid any additional cut on the Z-boson decay products, making the selections not realistic in a collider environment. However, it enables to be as inclusive as possible for the training of NNs, ensuring that any realistic selection will be enclosed in the phase-space region.

Finally, the third selection, which we dub *fiducial*, is then used to mimic a realistic setup at the LHC. In addition to the cuts in Eq. (17), transverse-momentum and rapidity cuts are applied on the charged leptons,

$$p_{\text{T},\mu^\pm} > 20\,\text{GeV} \qquad \text{and} \qquad |y_{\mu^\pm}| < 2.7. \tag{18}$$

3.2 Tools

For the generation of longitudinal and unpolarised parton-level events, we have used version 2.7.3 of MG5_AMC@NLO [32], which enables to select intermediate-resonance helicity states in the narrow-width approximation [60]. As a validation of the longitudinal signal, we have compared the MG5_AMC@NLO results against those obtained with the private Monte Carlo framework MoCaNLO, that uses the pole-approximation approach detailed in Refs. [72, 33, 34, 36, 40]. Good agreement has been found.

In order to compute PS effects, we have used version 8.244 of the Pythia8 program [73] with standard settings. The space-like and time-like shower have been applied with both QCD and QED effects. For what concerns QED effects, we veto further photon splittings into fermion pairs in the shower. Note that we have not included multi-parton interactions and hadronisation effects. The reason is to keep this example, while non-trivial, as simple as possible. We argue that including these extra effects could simply be included upon performing a new training of the NN. The principle of the ML technique we propose would not be hampered by different scale choices, matching and PS settings that may be needed especially when including higher-order effects.
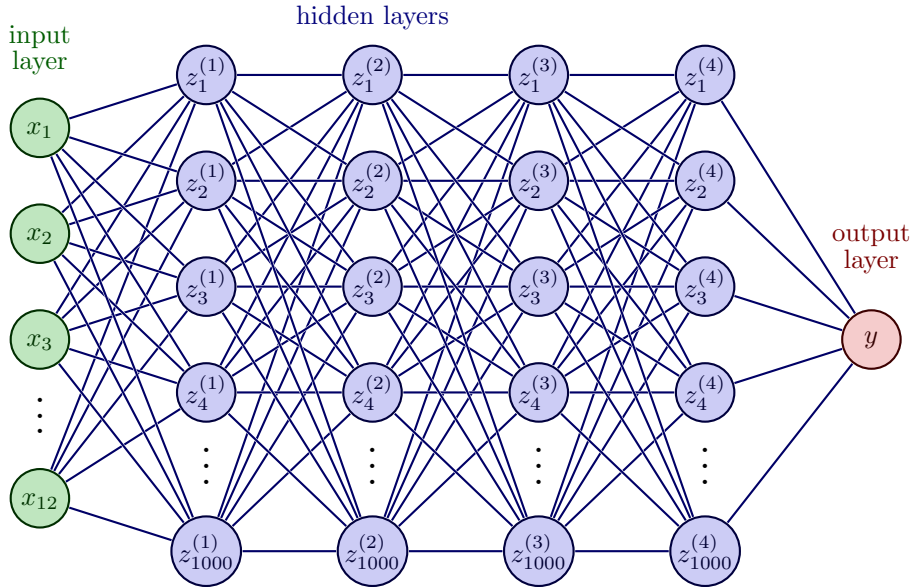
Figure 1: Structure of the neural network. Picture adapted from Ref. [76].

Finally, in order to compute r_L for the various partonic channels, we have used the matrix-element provider RECOLA 1 [74, 75] [3].

To obtain an approximate value $\tilde{r}_\mathrm{L}$, a machine learning approach was employed, utilising a feed-forward NN. The model is built from the dataset characterised by the following twelve input features, namely the four momenta of the (leading) jet, the antimuon, and the muon:

$$E_\mathrm{j}, \ p_{x,\mathrm{j}}, \ p_{y,\mathrm{j}}, \ p_{z,\mathrm{j}}, \ E_{\mu^+}, \ p_{x,\mu^+}, \ p_{y,\mu^+}, \ p_{z,\mu^+}, \ E_{\mu^-}, \ p_{x,\mu^-}, \ p_{y,\mu^-}, \ p_{z,\mu^-} \ , \tag{19}$$

along with the quantity r_L defined in Eq. (11), representing the continuous label.

The dataset used for training and testing containes 286 073 and 285 187 elements, respectively. As part of the data-preparation process, the dataset has been standardised according to a general procedure where each feature is normalised, with the subtraction of the average value of the feature and divided by its standard deviation. This approach is more suited than a min-max normalisation, due to its lower sensitivity to outliers. The architecture of the NN is wide, and this represents a crucial aspect of the proposed technique. Four hidden layers, consisting of 1000 nodes each, were employed. The mathematical representation of this NN involves a series of transformations. The input x has a dimensionality of $\mathbb{R}^{12\times1}$. The subsequent layers, indexed as i ranging from 0 to 4, were computed using the formula,

$$\begin{aligned}
z^{(0)}(x;W) &= x, \\
z^{(i+1)}(x;W) &= W_{i+1}\sigma(z^{(i)}(x;W)) + b_{i+1}, \\
y(x;W) &= z^{(5)}(x;W).
\end{aligned} \tag{20}$$

The weights are represented by $W = (W_1, W_2, W_3, W_4, W_5, b_1, b_2, b_3, b_4, b_5)$, where $W_1 \in \mathbb{R}^{1000\times12}$, while $W_2, W_3, W_4 \in \mathbb{R}^{1000\times1000}$, and $W_5 \in \mathbb{R}^{1\times1000}$. Additionally, $b_1, b_2, b_3, b_4 \in \mathbb{R}^{1000\times1}$, and

---

[3]We used the most recent RECOLA 1 release (version 1.4.3) which supports helicity selections for intermediate resonances at tree and one-loop level. Further documentation on the usage of polarisation-related subroutines can be found at https://recola.gitlab.io/recola2/api/polsel.html#polsel.

$b_5 \in \mathbb{R}^{1 \times 1000}$. The activation function used was the Rectified Linear Unit (ReLU) [77], defined as $\sigma(x) = \max\{0, x\}$. The structure of this NN is depicted in Figure 1.

The design of the machine learning model in this study follows the principles outlined in Ref. [78]. According to the theory presented, the effectiveness of a NN is influenced by the dynamics of its training process. It suggests that NNs with a "deep and narrow" architecture exhibit chaotic dynamics during training, while those with a "shallow and wide" architecture are easier to train. In the asymptotic case, infinitely wide NNs possess a convex loss landscape, enabling the optimal solution to be found through gradient descent. However, such models essentially become linear, losing the non-linear expressivity of the original network and potentially limiting its representational capacity. Therefore, a compromise must be made between the ease of training and the network's expressivity. In our experiments, we achieved satisfactory performance by employing a wide NN with a width-to-depth ratio of 200. Such a choice has been made by empirical experimentation and refinement.

The training of the NN model was performed using the RMSprop algorithm [79], an adaptive learning rate optimisation method specifically designed for mini-batch learning. The algorithm's parameters were set as follows: learning rate $\eta = 0.001$, smoothing constant $\alpha = 0.99$, weight decay of 0, and momentum of 0. The training process spanned 1000 epochs, with batches of size 500. The code implementing the model was developed in Python 3, making use of the PyTorch library [80].

3.3   Results

The key quantity in this application is the ratio r_{L} defined in Eq. (11). It encodes the whole information on the longitudinal-polarisation dynamics, relatively to the polarisation balance in the unpolarised process. As such, it determines the shape and normalisation of the kinematic distributions for the longitudinal signal. It is therefore a multi-dimensional function with as many dimensions as the number of random variables needed to generate the momenta of the full final state. As explained above, this ratio is actually different for each partonic channel. It means that in order to evaluate it on an event-by-event basis, one does not only need the full kinematic but also the knowledge of the partonic channel. In order to have a feeling about the structure of r_{L}, we show in Fig. 2 the differential distribution in r_{L} for unpolarised events, in the generation-level setup. We also show, for comparison purposes, the distribution in the corresponding r_{T} and r_{int} ratios, that are respectively defined as,

$$r_{\mathrm{T}} = \frac{|\mathcal{M}_{\mathrm{T}}|^2}{|\mathcal{M}|^2} \quad \text{and} \quad r_{\mathrm{int}} = \frac{2 \operatorname{Re}(\mathcal{M}_{\mathrm{L}}^* \mathcal{M}_{\mathrm{T}})}{|\mathcal{M}|^2} = 1 - r_{\mathrm{L}} - r_{\mathrm{T}}. \tag{21}$$

From Eq. (21) and Fig. 2, it is clear that, since all amplitudes are complex numbers, the interference term can take negative values, and both r_{L} and r_{T} can exceed the unit. Owing to a peak at 1, the r_{T} distribution in Fig. 2 suggests that in the considered process the transverse-polarisation component is way larger than the longitudinal one. This is a well-known result in the SM [11, 12]. It also shows that the proposed method is particularly efficient as it can make full use of this discriminating power.

r_{L}**-reweighting/tagging**   The first key observation that we have made above is that unpolarised event samples can be reweighted/tagged[4] using r_{L} to obtain longitudinally polarised

---

[4]In the following, we use indistinguishably *reweighting* and *tagging* unless otherwise stated. In the reweighting approach, the event weights r_{L} (or $\tilde{r}_{\mathrm{L}}$) are directly used to compute longitudinal distributions. In the tagging
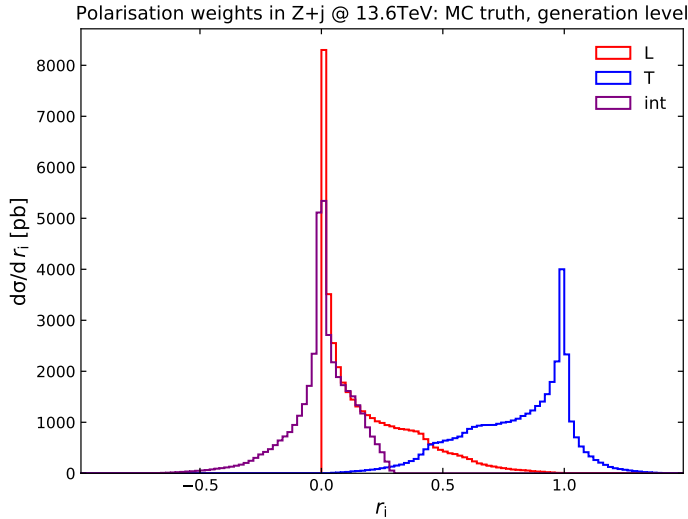
Figure 2: Distributions in the r_L, r_T and r_{int} quantities defined in Eqs. (11) and (21), all normalised to the unpolarised total cross section. The generation-level setup is understood.

accuracy	MC truth	r_L-reweighting
inclusive setup		
LO	0.1704(4)	0.1703(3)
LO + PS	0.1722(4)	0.1725(3)
fiducial setup		
LO	0.1879(6)	0.1883(6)
LO + PS	0.1889(6)	0.1894(6)

Table 1: Longitudinal-polarisation fraction determined from MC-truth longitudinal events and from r_L-reweighting of unpolarised events, in the inclusive and fiducial setups. Monte Carlo uncertainties on the fractions are shown in parentheses.

samples. This statement does not only hold at LO but also when including PS effects, thanks to the factorisation of the radiative corrections as implemented in a PS, *i.e.* adding multiple QCD and QED radiations in the collinear approximation at leading-logarithmic accuracy. This can be seen in Fig. 3, where two differential distributions are shown at LO and LO+PS accuracy. The distributions obtained with the r_L-reweighting reproduce very well those obtained with longitudinal events generated with the Monte Carlo (MC truth). This confirms that a longitudinally polarised sample can be obtained by simply reweighting an unpolarised one with r_L factors. In addition, one can observe that while the PS corrections are sizeable (boh in the overall normalisation and in the distribution shapes), the statement about the reweighting is equally true in the presence of PS corrections. From the results of Table 1, it can also be appreciated how the r_L-reweighting performs well both in inclusive setups and in the presence of more exclusive selection cuts. The differential results analogous to those of Fig. 3 but in the fiducial setup

---

approach, the longitudinal sample is extracted from the unpolarised one by means a one-dimensional sampling according to r_L (or $\tilde{r}_L$) weights. The selected events are then used to compute the longitudinal distributions. Notice that the two methods are equivalent within statistical uncertainties.
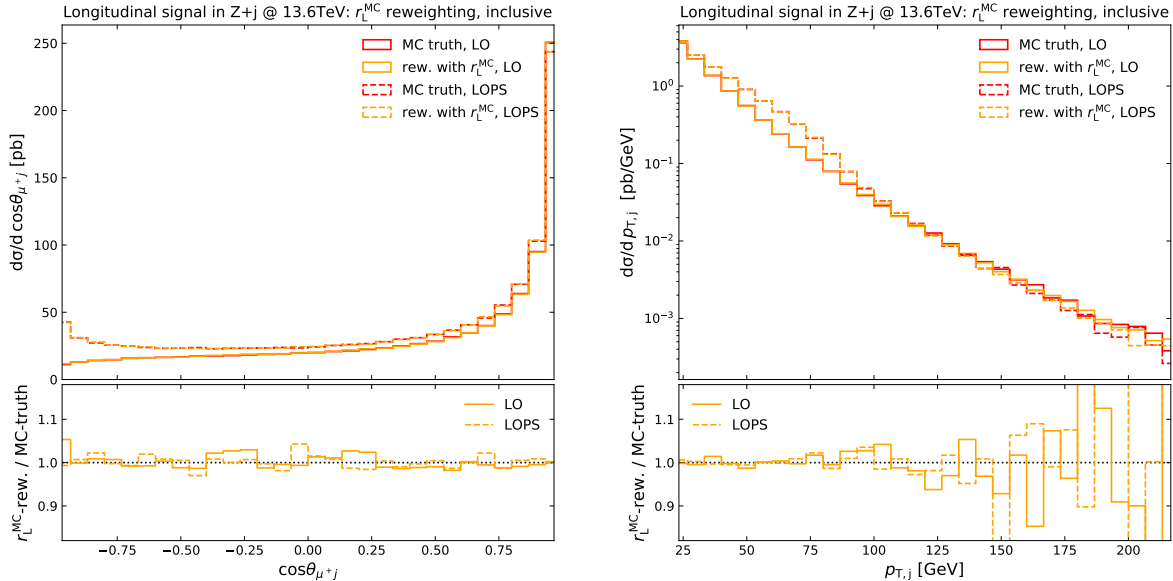
Figure 3: Longitudinal reweighting of unpolarised events with r_L (orange) compared to MC-truth longitudinal events (red) at LO (solid) and LO+PS (dashed). Absolute differential cross sections are shown in the top panel, ratios of reweighted results over MC-truth ones are shown in the bottom panel. The following observables are considered: cosine of the angular separation between the antimuon and the leading jet (left), leading-jet transverse momentum (right). The inclusive setup is understood here.

(not shown here) also highlight an almost perfect behaviour of the reweighting method as in the inclusive setup.

Leading order Turning the problem around, the results detailed in the previous paragraph imply that experimental data (here idealised by LO+PS unpolarised events) can be used to extract polarisation fraction provided that r_L is known and can be computed on an event-by-event basis. Actually, r_L cannot be computed from experimental data, which do not give access to the full kinematic dependence (including the initial state) and to the flavour of all external particles. To bypass this issue, one can use NNs to obtain an approximation $\tilde{r}_L$ of the true ratio, based on an incomplete information, namely the one available experimentally which consists in the visible final-state momenta. Along this line, the first step is therefore to check if one can obtain a good approximation of r_L at LO by training a NN in a supervised setting, as described above, with r_L as input label and the final-sate jet momentum and lepton momenta as incomplete information for the training features.

In Table 2 and Figs. 4–5, two different NNs, predicting $\tilde{r}_L$ factors as approximations of the r_L ones, are compared against the true Monte Carlo results, both at the level of polarisation fractions and at the level of differential cross sections. The first network (labeled NN_1, purple curves in plots) underwent training for 10^6 epochs, employing a batch size of 100 and a learning rate of 10^{-4}. The second network (labeled NN_2, olive curves in plots) was trained for 2×10^3 epochs, using a batch size of 500 and the same learning rate. The generation-level events were used for the training.

The results of Table 2 show a good performance of the NN_1 model in reproducing the polarisation fractions both at inclusive and fiducial level with sub-per-cent accuracy. The NN_2
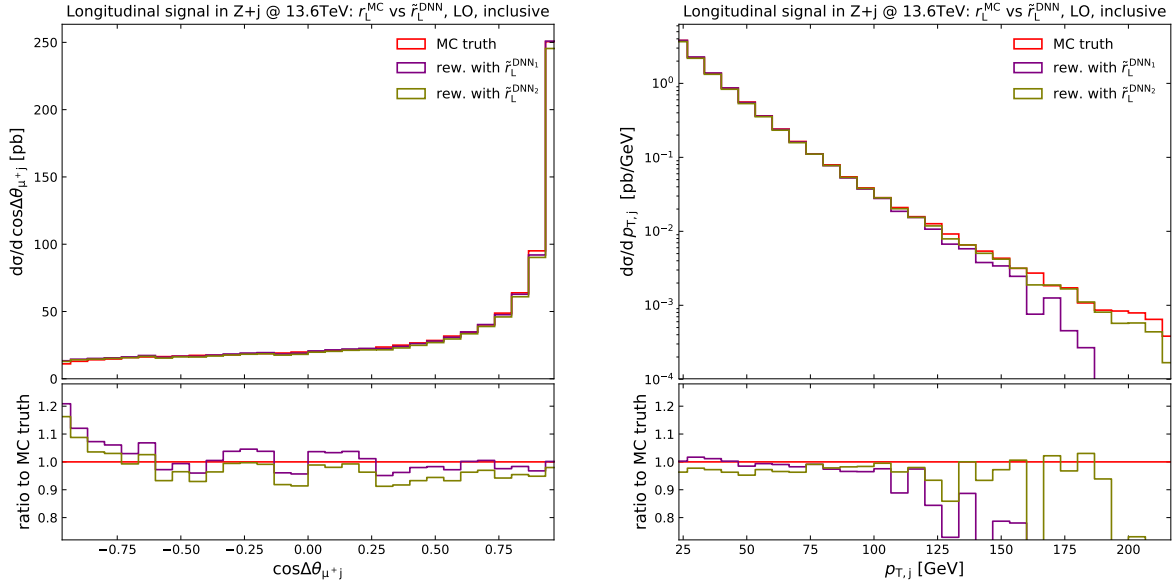
Figure 4: Longitudinal reweighting of unpolarised events with $\tilde{r}_L$ predicted by two different NN models (olive and purple curves) compared to MC-truth longitudinal events (red curve) at LO. Absolute differential cross sections are shown in the top panel, ratios of reweighted results over MC-truth ones are shown in the bottom panel. The following observables are considered: cosine of the angular separation between the antimuon and the leading jet (left), leading-jet transverse momentum (right). The inclusive setup is understood.
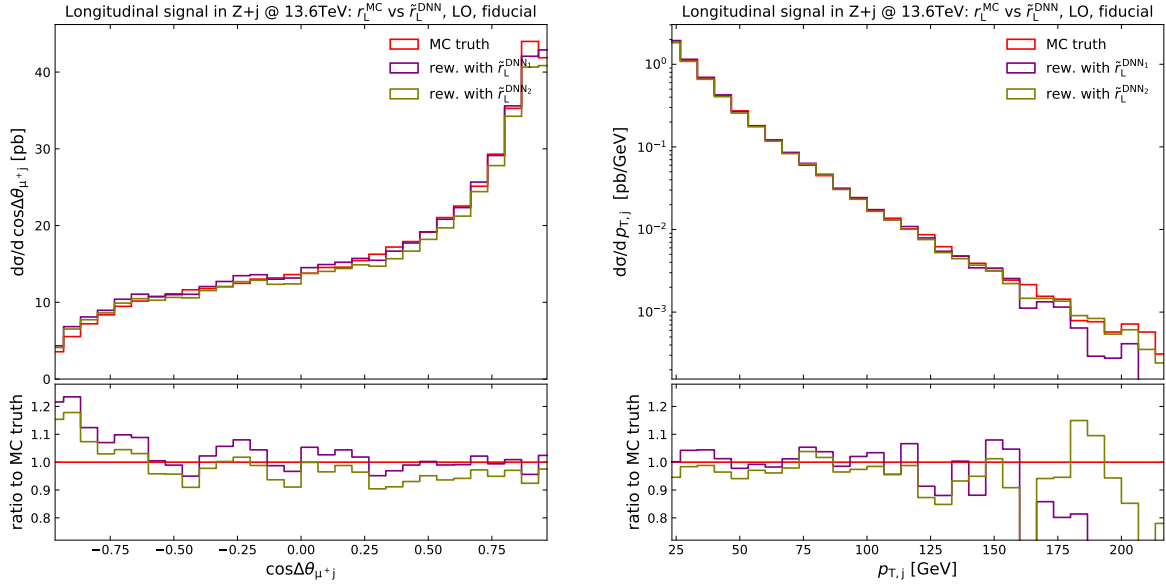
Figure 5: Same structure as Fig. 4. The fiducial setup is understood.

setup	MC	NN$_1$	NN$_1$/MC	NN$_2$	NN$_2$/MC
inclusive	0.1704(4)	0.1705(5)	100.1%	0.1649(5)	96.8%
fiducial	0.1879(6)	0.1904(8)	101.3%	0.1812(8)	96.4%

Table 2: Longitudinal-polarisation fractions at LO determined from Monte Carlo-truth longitudinal events (MC) and from reweighting of unpolarised events with NN-predicted $\tilde{r}_L$ (NN), in the inclusive and fiducial setups. Monte Carlo uncertainties on the fractions are shown in parentheses. The NN-predicted fractions are assigned Monte–Carlo-like uncertainties according to the number of events at testing level.

model underestimates the longitudinal fraction by 3-4%.

In Figs. 4 and 5, the cosine of the angular difference between the positive lepton and the jet as well as the transverse momentum of the jet are shown. The two figures differ in their phase-space regions: Fig. 4 is for the inclusive setup while Fig. 5 is for the fiducial one. One observes that the first NN is reproducing better the true result. In general, the agreement is at the per-cent level for the phenomenologically relevant part of the phase space and therefore good enough for our purpose. Also, it is worth pointing out that in suppressed regions of phase space where the statistics is low, the agreement degrades substantially. The limited statistics used for the training stage in this suppressed region does not constrain strongly enough the NN model, leading therefore to a systematic error in the NN-model prediction for $\tilde{r}_L$. For example, above 150 GeV for the transverse momentum of the jet in Fig. 5, the agreement is worth than 20%. This is nonetheless not an issue given that this region is suppressed by two orders of magnitude, meaning that it contributes to about 1% to the cross section and therefore introduces only a per-mille error or less in total.

Finally, we note that the results for the inclusive and fiducial setups are equally good. The only difference that one can notice is that the fiducial results suffer from larger fluctuations. This can be attributed to the lower statistics used in the fiducial case ($\approx$ 220k events), owing to more restrictive selections that cut away more than half of unpolarised events used in the inclusive setup ($\approx$ 480k events).

Parton-shower effects Overall, the above results prove that the method is reliable also in typical experimental regions. Nonetheless, this is a simplified version of the problem as this exercise was performed at LO meaning for events of identical multiplicity. A more realistic description of the data necessarily requires PS corrections. Indeed LHC events are typically affected by several effects such as multi-particle interactions, beam remnants, hadronisation, extra QCD and QED radiations etc. These phenomena are well described by multi-purpose PS programs like PYTHIA [73]. In our case, we have included QCD and QED radiations but other effects could equally be included.

In order to account for PS effects in our method, one can try to use the previous NN trained with LO events and apply it to events modeled with PS corrections. The results of this procedure are shown in Fig. 6 for the cosine of the angular separation between the antimuon and the leading jet, and for the leading-jet transverse momentum. Notice that in this case, we have only included effects from QCD PS, avoiding further photon radiations. From the plots, it is rather clear that this approach is failing. The reason for this is that the PS generates more QCD radiations leading to a sizeable distortion of the event kinematics. In fact, comparing Fig. 6 with the fixed-order results in Fig. 3, one observes that applying the NN trained with LO events to LO+PS events
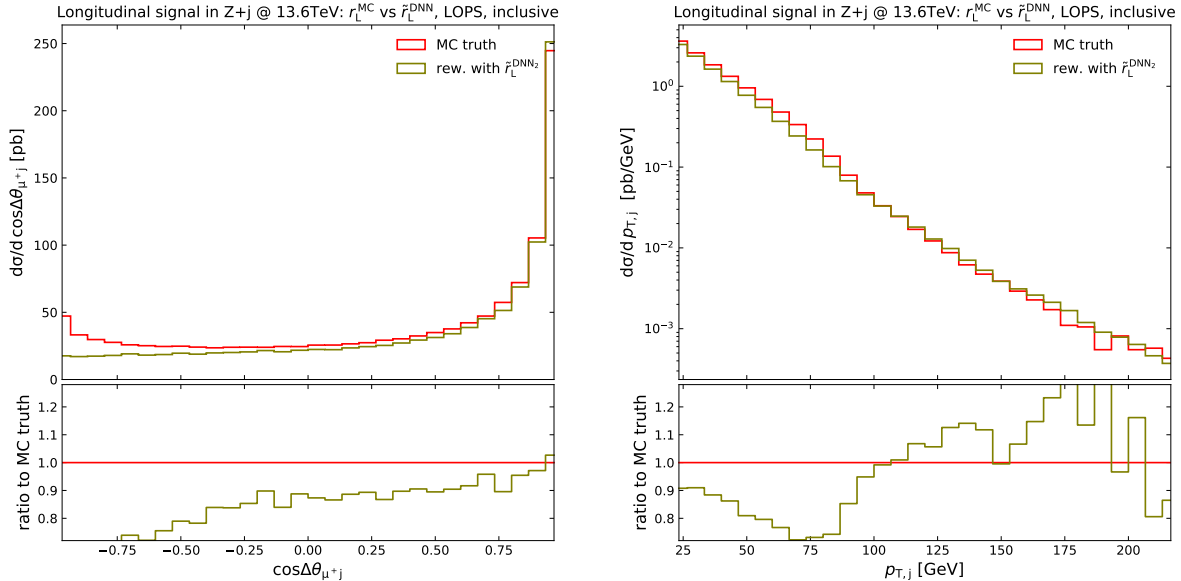
14

Figure 6: Longitudinal reweighting of unpolarised LO+PS events with $\tilde{r}_L$ from the NN model trained with LO events (olive curve), compared to MC-truth longitudinal events (red curve) at LO+PS (QCD shower only). Absolute differential cross sections are shown in the top panel, ratios of reweighted results over MC-truth ones are shown in the bottom panel. The following observables are considered: cosine of the angular separation between the antimuon and the leading jet (left), leading-jet transverse momentum (right). The inclusive setup is understood.

tends to reproduce the LO distribution shapes rather than the LO+PS ones.

As shown previously, one can apply a reweighting of the unpolarised sample and then apply the PS procedure or viceversa in order to obtain a longitudinally polarised sample with PS effects. This also means that for each showered event one can compute r_L with the original LO momenta before PS and therefore associate a meaningful r_L to each showered event. One can therefore train a new NN with the original r_L (computed before showering) along with the momenta after showering. Given that showered events possess more than one jet, only the four momentum of the jet with the largest transverse momentum is provided as a feature for the NN-model training.

In order to tackle this problem, we adopted the following training strategy. First, a wide neural network (labeled NN_{ws}) is trained using a *warm-start initialisation* [81]. Utilising the warm-start approach in training NNs involves initialising the model with weights from a previously trained model. This strategy potentially accelerates convergence, enhances performance, and reduces the need for extensive data, creating an efficient framework for model training. From a physics intuition, in the generation chain that starts with LO process and move to LO+PS, this step does not represents a completely new learning task rather it is a perturbation of the original process. It follows that the procedure involved utilising the configuration of the network previously trained on the LO events and starting the training with the LO+PS events from its optimal configuration in terms of architecture of the NN and its relative weights. This procedure, which is well-known in other machine-learning applications, has not been yet fully exploited in high-energy physics. Skipping a complete NN architecture optimisation procedure is advantageous because of the faster identification of the best model and of the computational-resource saving. For the sake of comparison, a second general NN (labeled NN_{nows}) is built from scratch, looking for the best depth-to-width ratio with a randomly chosen initial configuration.
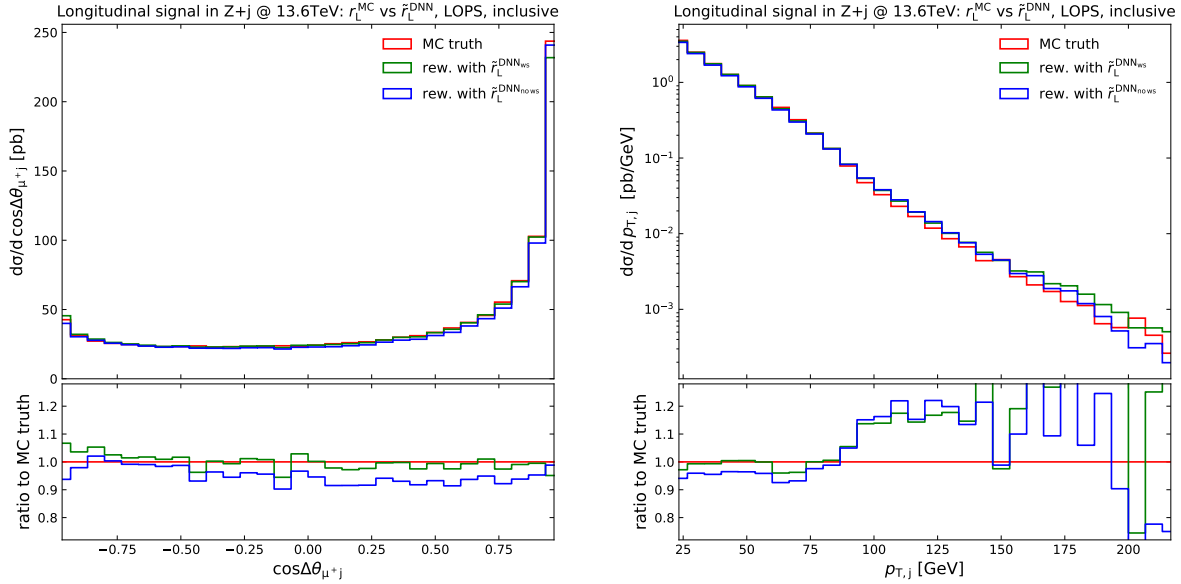
Figure 7: Longitudinal reweighting of unpolarised events with $\tilde{r}_L$ from two different NN models (blue and green curves) compared to MC-truth longitudinal events (red curve) at LO+PS (both QCD and QED showers included). Absolute differential cross sections are shown in the top panel, ratios of reweighted results over MC-truth ones are shown in the bottom panel. The following observables are considered: cosine of the angle between the antimuon and the leading jet (left), transverse momentum of the leading jet (right). The inclusive setup is understood.
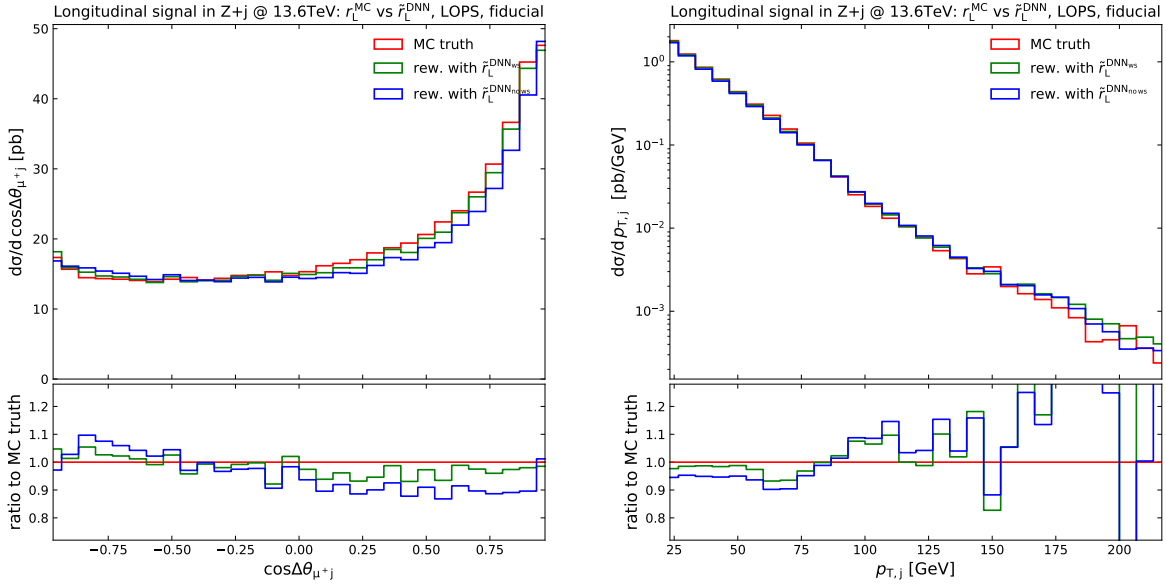
Figure 8: Same structure as Fig. 7. The fiducial setup is understood.

setup	MC	NN_{ws}	NN_{ws}/MC	$NN_{no\,ws}$	$NN_{no\,ws}$/MC
inclusive	0.1722(4)	0.1705(3)	99.0%	0.1646(3)	95.6%
fiducial	0.1889(6)	0.1853(5)	98.1%	0.1791(5)	94.8%

Table 3: Longitudinal-polarisation fractions at LO+PS determined from Monte Carlo-truth longitudinal events (MC) and from reweighting of unpolarised events with NN-predicted $\tilde{r}_L$ (NN), in the inclusive and fiducial setups. Monte Carlo uncertainties on the fractions are shown in parentheses. The NN-predicted fractions are assigned Monte–Carlo-like uncertainties according to the number of events at testing level.

The results provided by these two models are reported in Table 3 and Figs. 7–8 (green curves for NN_{ws}, blue curves for $NN_{no\,ws}$). As one can see from both integrated and differential results that the NN with warm start is outperforms the one built from scratch. From the results, it is clear that the warm start has beneficial effects on the NNs. Firstly, it biases the training towards solving a similar task, allowing the network to adjust its parameters to the new data, which limits the search space and leads to faster convergence. Additionally, as the problem becomes easier to solve, the quality of the solution improves. From a physics viewpoint, the good behaviour of the NN_{ws} implies that the LO step is actually of crucial importance to be able to use this method in an experimental analysis. In particular, the results are per-cent accurate at the level of polarisation fractions, which is good enough for the level of precision of this study. Considering differential observables, Fig. 7 refers to the inclusive case while Fig. 8 refers to the fiducial case. The same conclusions as at fixed order hold, namely that the limited statistics do play a role in the accuracy of the method, as can be observed in the far tails of the transverse-momentum distribution in Fig. 8 or in other phase-space regions which are the least populated ones. Nonetheless, at 100 GeV in the transverse-momentum distribution of the leading jet, a 10-20% mismodelling can be observed. These effects cannot be solely attributed to the statistics but should be considered as a systematic error of the NN. This mismodelling might originate from PS effect as shown in Fig. 3 where the region around 100 GeV marks a quantitative change in the PS corrections. This could also be interpreted as a limitation of the NN model to capture all features. Nonetheless, these 10-20% discrepancies appears in bins that are suppressed by almost two orders of magnitude and therefore they are not physically significant when integrating over the whole transverse-momentum spectrum. Hence, the method proposed here is still per-cent accurate.

Event tagging with $\tilde{r}_L$ As already discussed in Sect. 2.2, the probabilistic interpretation of r_L leads to the expectation that the $\tilde{r}_L$ predicted by the NN models is positive. However, the NN models have no physics insights about this constraint and $\tilde{r}_L$ is not always positive.

It turns out that at LO+PS level, the NNs are able to predict positive $\tilde{r}_L$ for more than 99% of the event, both in the inclusive and in the fiducial setup. Interestingly, at fixed order, the performances are worse, with positive longitudinal weights predicted for only roughly 95% of the events. The events for which the NN predicts negative weights give a harder p_{T,j_1} spectrum compared to the events with positive $\tilde{r}_L$, highlighting that in order to improve the accuracy of the NN also in boosted regimes, a dedicated training with boosted events is needed. We checked that discarding the events with negative $\tilde{r}_L$, in spite of a partial improvement in the reproduction of the transverse-momentum shapes, overestimates by several per-cent the overall longitudinal fraction.

setup	MC	NN$_{\text{ws}}$-sampl	NN$_{\text{ws}}$-sampl/MC
inclusive	0.1722(4)	0.1716(5)	99.7%
fiducial	0.1889(6)	0.1850(8)	97.9%

Table 4: Longitudinal-polarisation fractions at LO+PS determined from MC-truth longitudinal events (MC) and sampling (NN-sampl) of unpolarised events according to $\tilde{r}_{\text{L}}$ predicted with the NN$_{\text{ws}}$ model, in the inclusive and fiducial setups. Monte Carlo uncertainties on the fractions are shown in parentheses. The NN-predicted fractions are assigned Monte–Carlo-like uncertainties according to the number of events at testing level.

A viable strategy could be to include a suppression function for negative $\tilde{r}_{\text{L}}$ at the level of the last layer of the NN models, as a small step toward physics-informed approaches. These ones have already been applied for classification tasks in particle physics, where enforcing symmetries conservation for transformations under the Lorentz group, provides a much more physically interpretable model [82]. However, there is no guarantee of improved accuracy in the NN. In our specific case, enforcing the positivity of the label actually worsens the overall performance. This is due to the introduction of constraints complicating the landscape of the loss function, resulting in more challenging geometries with multiple local minima. As a result, training becomes less effective, leading to poorer predictions from the model. We have refrained from investigating this aspect further, as the LO+PS results are satisfactory for the present application.

So far we have indistinguishably used the expression *reweighting* and *tagging*. However, while the reweighting strategy can be applied also in the presence of negative weights, the event tagging is not well defined anymore in that case. In other terms, performing a longitudinal tagging according to the NN predictions is not possible for events with negative $\tilde{r}_{\text{L}}$. While at LO accuracy this means throwing away 5% of the events, at LO+PS accuracy, which is the most important case as it mimics the experimental environment, less than a per cent of the events have to be thrown away, which is good enough for our purposes.

To illustrate the applicability of the method, we show final results for the event tagging, which turns out to be equivalent to the reweighting ones, at LO+PS accuracy. In Table 4, the polarisation fractions obtained with tagging are compared to the MC-truth ones. As expected, these results are good and almost equivalent to the reweigthing results provided in Table 3, since only very few events have a negative $\tilde{r}_{\text{L}}$. Notice that for this comparison we have only considered the NN that employs the warm-start approach (NN$_{\text{ws}}$). The results are equally good at the differential level, as can be observed in Figs. 9 and 10 for the inclusive and fiducial setup, respectively. In particular, in these plots, the MC-truth longitudinal distributions are compared with those obtained reweighting and tagging according to $\tilde{r}_{\text{L}}$. Both are equivalent up to statistical fluctuations. This finally demonstrates that using $\tilde{r}_{\text{L}}$ with experimental inputs enables an actual longitudinal-polarisation tagging on an event-by-event basis.

3.4 Discussion

With this non-trivial LHC application detailed in Sect. 3, we have shown that one can assert the polarisation fraction on an event-by-event basis using amplitude information by reverting to machine learning. The method is per-cent accurate and particularly versatile. In this section we discuss limitations of the methods as well as possible extensions, generalisations, and further applications.
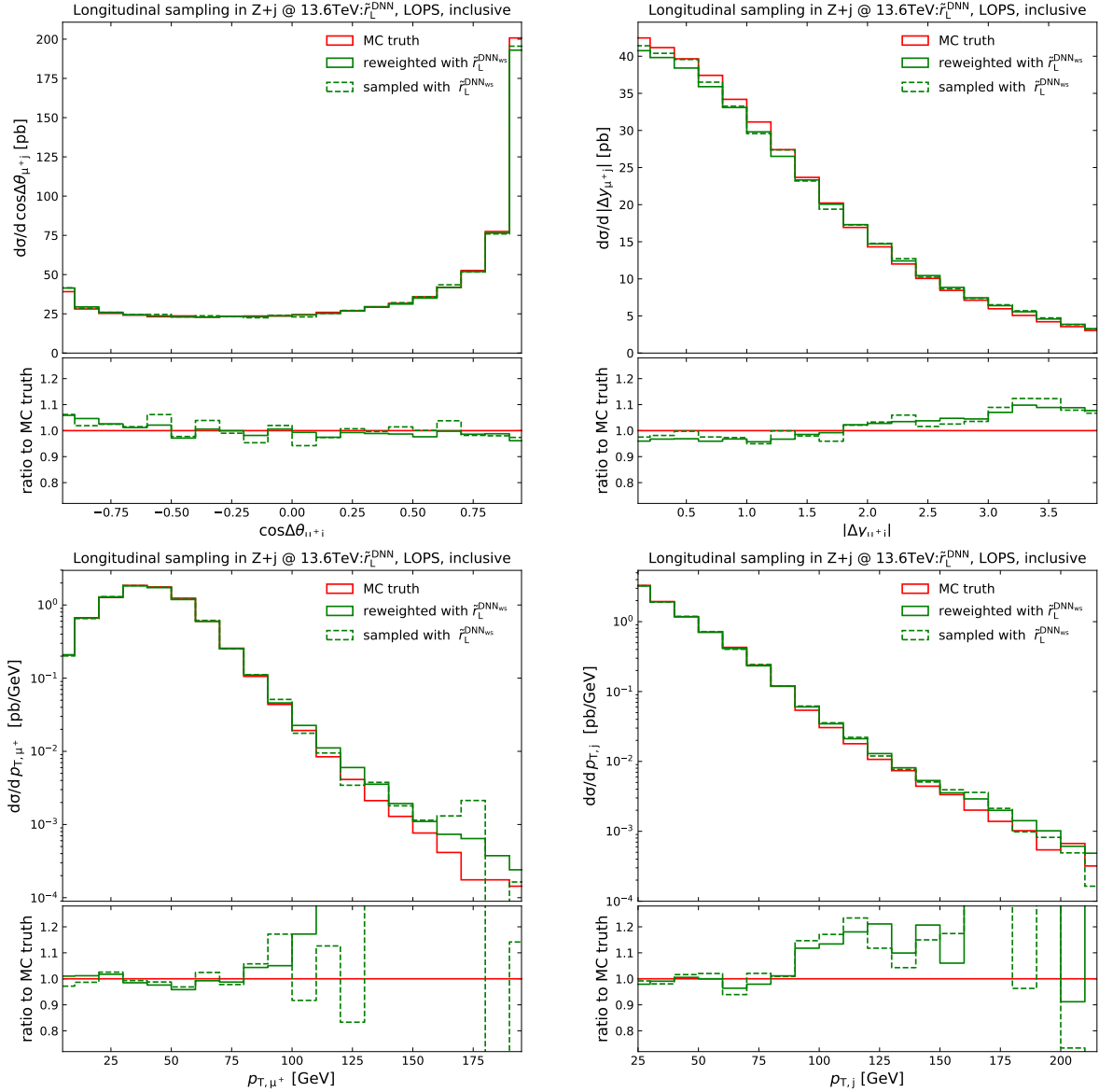
Figure 9: Longitudinal reweighting (solid green) and tagging (dashed green) of unpolarised events with $\tilde{r}_L$ predicted by the NN_{ws} model compared to MC-truth longitudinal events (solid red) at LO+PS (both QCD and QED showers included). Absolute differential cross sections are shown in top panels, ratios over MC-truth ones are shown in bottom panels. The following observables are considered: cosine of the angle between the antimuon and the leading jet (top left), rapidity separation between the antimuon and the leading jet (top right), tranverse momentum of the antimuon (bottom left), transverse momentum of the leading jet (bottom right). The inclusive setup is understood.
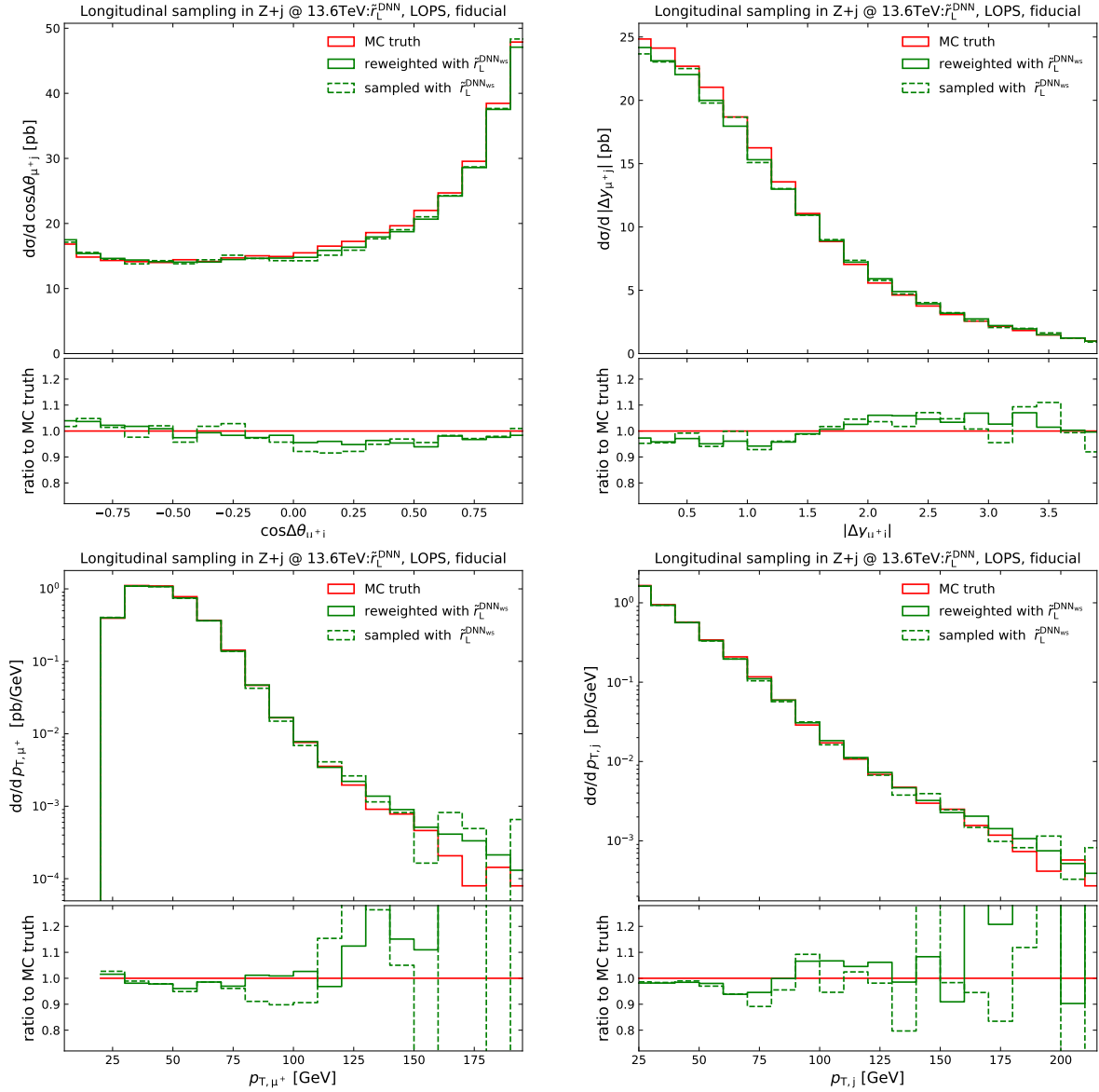
Figure 10: Same structure as Fig. 9. The fiducial setup is understood.

Validity of the method In the present example the training phase has been performed on events spanning a very inclusive sample. The trained models have then be used on a reduced phase-space as for typical experimental analyses. This ensures that the method is used in its region of validity. We therefore recommend to always perform the training on a more inclusive phase-space than the one actually used in the analysis. If this is not the case, it is not guaranteed that the method will still work as the network has not been trained (and thus validated) in the whole region considered at testing level. While it is not excluded that the NN can perform some extrapolation outside its training region, this has to be carefully verified. In particular, using the extrapolation power of the NN might require a different NN and a potentially a dedicated study on out-of-support extrapolation problem into a problem of within-support generalization.

While the specific application considered in this work concerns pp → Z + j at the LHC, the general idea of training a NN with experimentally accessible kinematic information and squared

polarised amplitudes can be applied to any single- or multi-boson process at colliders. We stress that in order to apply this strategy to another process, a new NN has to be constructed, relying on the corresponding input features that depend on the experimental signature. For example, for processes with final-state neutrinos, whose momenta cannot be fully reconstructed, the input feature would be the missing transverse momentum instead of the complete momenta of the neutrinos.

Error propagation As formulated here, the method would provide a numerical value corresponding to the experimentally-extracted fraction that can be compared against theoretical predictions. Nonetheless this extracted value has uncertainty of different sources: the accuracy of the theory prediction it relies on, the limited statistics of the training data set, and the experimental accuracy (both statistical and systematic) of the data.

Usually, the theoretical uncertainty on the prediction is assessed by means of scale variations of the factorisation and renormalisation scale. The envelop of the values of r_L extracted for different scale combinations would then provide the theory uncertainty associated to r_L. This quantity being a ratio of squared amplitudes, we expect the correlated scale uncertainties to be rather small, owing to cancellations between the numerator (longitudinal matrix element) and the denominator (unpolarised).

The uncertainty related to the finite size of the training sample can be inferred by performing the training with different sample sizes or by performing error propagation in the NN. The same applies to the experimental error associated to the reconstructed event kinematics, and it can be estimated by repeating the method using pseudo-data.

It is important to consider that NNs are complex models, and training them using stochastic gradient descent over a non-convex landscape does not guarantee optimal parameter quality upon convergence. In contrast, linear and kernel models can be trained more efficiently due to their convex and low-dimensional loss landscapes, albeit resulting in simpler predictors. The non-linearity of NNs allows them to learn new and more effective representations of the data, a process known as *feature learning* [78]. This feature learning effect makes NNs more powerful but also presents challenges in their training process. In our approach, we have opted for wide NNs that strike a balance between linear and nonlinear models. While infinite-width NNs are equivalent to linear models and enjoy convex optimization landscapes [83], wide networks with finite width exhibit a slightly more challenging training landscape. Nevertheless, training wide NNs remains effective, with the difficulty of the landscape increasing as the depth-to-width ratio grows. Considering all these aspects related to the complexity behavior of ML models, we refrain from assigning any intrinsic uncertainty to the predicted output.

Model independence As already mentioned, the polarisation of weak bosons is a pseudo-observable and its extraction necessarily understands some degree of model dependence. In the method we propose, the model dependence is encoded into the r_L function. In the present work, the SM is considered: it means that events are tagged according to SM expectations and the longitudinal fractions extracted should be compared against the one of theoretical predictions within the SM. Such a model dependence is impossible to avoid. In fact, even a simplified version of r_L relying only on the boson-decay matrix elements still depends on the polarisation fractions determined by the model-specific production mechanism. However, since the method proposed in this work is model agnostic, the same procedure can be performed with more general models, *i.e.* simplified models or effective field theories, or even with UV-finite theories.

Extension to higher orders   In the present work, we have restricted our analysis to LO+PS accuracy. Nonetheless, it is in principle possible to extend this to higher orders in perturbation theory, at least for QCD corrections to processes with leptonically decaying bosons. If one can produce a sample of unweighted events at a given order in perturbation theory, the presented method can be extended.

Having unweighted events at fixed order implies having events with different jet multiplicity (depending on the order considered). It means that for each multiplicity $i = 0, 1, ...$, the exact r_L^i can be computed using loop and/or tree amplitudes depending on the accuracy of the sample. As ratios, they should actually be free of infrared singularities if QCD dependencies factorise from the polarisation effects. As in the presented application, the NN can learn a single approximate $\tilde{r}_L$ based on the experimental input available and in particular by feeding only the leading jet(s) in the transverse momentum. Adding PS or further corrections can then be achieved as shown in the previous sections.

We stress that we have not explicitly tested this method with higher orders and therefore that if ones wants to use this proposal, it should be carefully checked first. In particular, the main assumption here is that QCD corrections and polarisation effects factorise to a large extend (as PS and polarisation effects). This implies that the inclusion of EW corrections would probably requires a more refined analysis given that they are known not to factorise.

Generalisation to other problems   As highlighted several times, the key aspect of the method is to encode the whole physics problem in one single ratio (in the present application r_L) which can be approximately reconstructed using incomplete information thanks to NN methods. It therefore implies that the method can be applied to any physics problem that can be cast in this form. The only requirement being that the key quantity is bounded as it is the case for ratios of amplitudes. It also means that appropriate problems for this method are the extraction of a signal over a background which is very common in experimental particle physics.

4   Conclusions

The polarisation of heavy gauge bosons encodes the intricate structure of the electroweak sector of the Standard Model. The theoretical study and the experimental extraction of such pseudo-observables is thus of prime importance for the present and upcoming physics programme of the LHC. It is therefore key to combine our theoretical understanding to make use of all the information available in experimental data in order to probe the structure of the Standard Model at the deepest.

In this work, we have designed an original method to extract polarisation fractions using the maximal information encoded in the amplitude thanks to the versatility of neural networks. The key feature is that all information is encapsulated in a single number which can be computed on a event-by-event basis. In particular, the neural network is able to construct a particularly good approximation of this quantity which can then be evaluated with incomplete information, namely the one available in experiments. This number allows to assert whether an event is most likely longitudinally polarised or not. In this way, all information *i.e.* the fully differential information is exploited and not only the information contained in one or several observables as it is the case for other methods. It also means that no fitting procedure is required. Another advantage is that the theory dependence is clearly identified as it is only encoded in the amplitude. Finally, the amplitude considered can be the one of arbitrary-general or -specific models of quantum field theory.

To illustrate the method, we have applied it to the extraction of the longitudinal polarisation of a Z boson in the hadronic process pp → Z + j, in the leptonic decay channel at the LHC. We have demonstrated that the idea is working with a per-cent accuracy by reverting to the sequential training of a neural network. In particular, when being used in actual experimental analyses, the closure tests that we have presented here should be carried out to ensure the correctness of the results.

Finally, we point out that the method we have developed is very general. It can therefore be applied to other problems and/or generalised. In particular, the method seems to be particularly appropriate for the extraction of signals over irreducible or even reducible background.

Acknowledgements

References

[1] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons.* Phys. Rev. Lett. **13** (1964) 321–323.

[2] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons.* Phys. Rev. Lett. **13** (1964) 508–509.

[3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles.* Phys. Rev. Lett. **13** (1964) 585–587.

[4] **ATLAS** Collaboration, M. Aaboud *et al.*, *Measurement of $W^{\pm}Z$ production cross sections and gauge boson polarisation in pp collisions at $\sqrt{s} = 13\,TeV$ with the ATLAS detector.* Eur. Phys. J. C **79** (2019) no. 6, 535, arXiv:1902.05759 [hep-ex].

[5] **CMS** Collaboration, A. M. Sirunyan *et al.*, *Measurements of production cross sections of polarized same-sign W boson pairs in association with two jets in proton-proton collisions at $\sqrt{s} = 13\,TeV$.* Phys. Lett. B **812** (2021) 136018, arXiv:2009.09429 [hep-ex].

[6] **CMS** Collaboration, A. Tumasyan *et al.*, *Measurement of the inclusive and differential WZ production cross sections, polarization angles, and triple gauge couplings in pp collisions at $\sqrt{s} = 13\,TeV$.* JHEP **07** (2022) 032, arXiv:2110.11231 [hep-ex].

[7] **ATLAS** Collaboration, *Observation of gauge boson joint-polarisation states in $W^{\pm}Z$ production from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.* arXiv:2211.09435 [hep-ex].

[8] **CMS** Collaboration, *Vector Boson Scattering prospective studies in the ZZ fully leptonic decay channel for the High-Luminosity and High-Energy LHC upgrades* Tech. Rep. CMS-PAS-FTR-18-014, CERN, Geneva, 12, 2018.

[9] P. Azzi *et al.*, "Report from Working Group 1: Standard Model Physics at the HL-LHC and HE-LHC," in *Report on the Physics at the HL-LHC,and Perspectives for the HE-LHC*, A. Dainese, M. Mangano, A. B. Meyer, A. Nisati, G. Salam, and M. A. Vesterinen, eds., vol. 7, pp. 1–220. CERN, 12, 2019. `arXiv:1902.04070 [hep-ph]`.

[10] J. Roloff, V. Cavaliere, M.-A. Pleier, and L. Xu, *Sensitivity to longitudinal vector boson scattering in semileptonic final states at the HL-LHC.* Phys. Rev. D **104** (2021) no. 9, 093002, `arXiv:2108.00324 [hep-ex]`.

[11] Z. Bern *et al.*, *Left-handed W bosons at the LHC.* Phys. Rev. **D84** (2011) 034008, `arXiv:1103.5445 [hep-ph]`.

[12] W. J. Stirling and E. Vryonidou, *Electroweak gauge boson polarisation at the LHC.* JHEP **07** (2012) 124, `arXiv:1204.6427 [hep-ph]`.

[13] A. Belyaev and D. Ross, *What Does the CMS Measurement of W-polarization Tell Us about the Underlying Theory of the Coupling of W-Bosons to Matter?* JHEP **08** (2013) 120, `arXiv:1303.3297 [hep-ph]`.

[14] **CMS** Collaboration, S. Chatrchyan *et al.*, *Measurement of the Polarization of W Bosons with Large Transverse Momenta in W+Jets Events at the LHC.* Phys. Rev. Lett. **107** (2011) 021802, `arXiv:1104.3829 [hep-ex]`.

[15] **ATLAS** Collaboration, G. Aad *et al.*, *Measurement of the polarisation of W bosons produced with large transverse momentum in pp collisions at $\sqrt{s} = 7\,TeV$ with the ATLAS experiment.* Eur. Phys. J. **C72** (2012) 2001, `arXiv:1203.2165 [hep-ex]`.

[16] **CMS** Collaboration, V. Khachatryan *et al.*, *Angular coefficients of Z bosons produced in pp collisions at $\sqrt{s} = 8\,TeV$ and decaying to $\mu^+\mu^-$ as a function of transverse momentum and rapidity.* Phys. Lett. **B750** (2015) 154–175, `arXiv:1504.03512 [hep-ex]`.

[17] **ATLAS** Collaboration, G. Aad *et al.*, *Measurement of the angular coefficients in Z-boson events using electron and muon pairs from data taken at $\sqrt{s} = 8\,TeV$ with the ATLAS detector.* JHEP **08** (2016) 159, `arXiv:1606.00689 [hep-ex]`.

[18] **ATLAS** Collaboration, M. Aaboud *et al.*, *Measurement of the W boson polarisation in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 8\,TeV$ in the lepton + jets channel with ATLAS.* Eur. Phys. J. C **77** (2017) no. 4, 264, `arXiv:1612.02577 [hep-ex]`. [Erratum: Eur. Phys. J. C 79 (2019) 19].

[19] **CMS** Collaboration, V. Khachatryan *et al.*, *Measurement of the W boson helicity fractions in the decays of top quark pairs to lepton + jets final states produced in pp collisions at $\sqrt{s} = 8\,TeV$.* Phys. Lett. **B762** (2016) 512–534, `arXiv:1605.09047 [hep-ex]`.

[20] **CMS, ATLAS** Collaboration, G. Aad *et al.*, *Combination of the W boson polarization measurements in top quark decays using ATLAS and CMS data at $\sqrt{s} = 8\,TeV$.* JHEP **08** (2020) no. 08, 051, `arXiv:2005.03799 [hep-ex]`.

[21] R. Gauld, A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, and A. Huss, *Precise predictions for the angular coefficients in Z-boson production at the LHC.* JHEP **11** (2017) 003, `arXiv:1708.00008 [hep-ph]`.

[22] J. Baglio and N. Le Duc, *Fiducial polarization observables in hadronic WZ production: A next-to-leading order QCD+EW study*. JHEP **04** (2019) 065, `arXiv:1810.11034 [hep-ph]`.

[23] R. Rahaman and R. K. Singh, *Anomalous triple gauge boson couplings in ZZ production at the LHC and the role of Z boson polarizations*. Nucl. Phys. B **948** (2019) 114754, `arXiv:1810.11657 [hep-ph]`.

[24] R. Rahaman and R. K. Singh, *Unravelling the anomalous gauge boson couplings in $ZW^{\pm}$ production at the LHC and the role of spin-1 polarizations*. JHEP **04** (2020) 075, `arXiv:1911.03111 [hep-ph]`.

[25] J. Baglio and N. Le Duc, *Polarization observables in WZ production at the 13 TeV LHC: Inclusive case*. Commun. Phys. **30** (2020) no. 1, 35–47, `arXiv:1910.13746 [hep-ph]`.

[26] R. Frederix and T. Vitos, *Electroweak corrections to the angular coefficients in finite-p_T Z-boson production and dilepton decay*. Eur. Phys. J. C **80** (2020) no. 10, 939, `arXiv:2007.08867 [hep-ph]`.

[27] R. Rahaman and R. K. Singh, *Breaking down the entire spectrum of spin correlations of a pair of particles involving fermions and gauge bosons*. Nucl. Phys. B **984** (2022) 115984, `arXiv:2109.09345 [hep-ph]`.

[28] M. Pellen, R. Poncelet, A. Popescu, and T. Vitos, *Angular coefficients in $W + j$ production at the LHC with high precision*. Eur. Phys. J. C **82** (2022) no. 8, 693, `arXiv:2204.12394 [hep-ph]`.

[29] A. Ballestrero, E. Maina, and G. Pelliccioli, *W boson polarization in vector boson scattering at the LHC*. JHEP **03** (2018) 170, `arXiv:1710.09339 [hep-ph]`.

[30] A. Ballestrero, E. Maina, and G. Pelliccioli, *Polarized vector boson scattering in the fully leptonic WZ and ZZ channels at the LHC*. JHEP **09** (2019) 087, `arXiv:1907.04722 [hep-ph]`.

[31] A. Ballestrero, E. Maina, and G. Pelliccioli, *Different polarization definitions in same-sign WW scattering at the LHC*. Phys. Lett. B **811** (2020) 135856, `arXiv:2007.07133 [hep-ph]`.

[32] D. Buarque Franzosi, O. Mattelaer, R. Ruiz, and S. Shil, *Automated predictions from polarized matrix elements*. JHEP **04** (2020) 082, `arXiv:1912.01725 [hep-ph]`.

[33] A. Denner and G. Pelliccioli, *Polarized electroweak bosons in W^+W^- production at the LHC including NLO QCD effects*. JHEP **09** (2020) 164, `arXiv:2006.14867 [hep-ph]`.

[34] A. Denner and G. Pelliccioli, *NLO QCD predictions for doubly-polarized WZ production at the LHC*. Phys. Lett. B **814** (2021) 136107, `arXiv:2010.07149 [hep-ph]`.

[35] R. Poncelet and A. Popescu, *NNLO QCD study of polarised W^+W^- production at the LHC*. JHEP **07** (2021) 023, `arXiv:2102.13583 [hep-ph]`.

[36] A. Denner and G. Pelliccioli, *NLO EW and QCD corrections to polarized ZZ production in the four-charged-lepton channel at the LHC*. JHEP **10** (2021) 097, `arXiv:2107.06579 [hep-ph]`.

[37] M. Pellen, R. Poncelet, and A. Popescu, *Polarised W+j production at the LHC: a study at NNLO QCD accuracy.* JHEP **02** (2022) 160, `arXiv:2109.14336 [hep-ph]`.

[38] D. N. Le and J. Baglio, *Doubly-polarized WZ hadronic cross sections at NLO QCD + EW accuracy.* Eur. Phys. J. C **82** (2022) no. 10, 917, `arXiv:2203.01470 [hep-ph]`.

[39] D. N. Le, J. Baglio, and T. N. Dao, *Doubly-polarized WZ hadronic production at NLO QCD+EW: calculation method and further results.* Eur. Phys. J. C **82** (2022) no. 12, 1103, `arXiv:2208.09232 [hep-ph]`.

[40] A. Denner, C. Haitz, and G. Pelliccioli, *NLO QCD corrections to polarized diboson production in semileptonic final states.* Phys. Rev. D **107** (2023) no. 5, 053004, `arXiv:2211.09040 [hep-ph]`.

[41] J. Brehmer, J. Jaeckel, and T. Plehn, *Polarized WW Scattering on the Higgs Pole.* Phys. Rev. D **90** (2014) no. 5, 054023, `arXiv:1404.5951 [hep-ph]`.

[42] E. Maina, *Vector boson polarizations in the decay of the Standard Model Higgs.* Phys. Lett. B **818** (2021) 136360, `arXiv:2007.12080 [hep-ph]`.

[43] E. Maina and G. Pelliccioli, *Polarized Z bosons from the decay of a Higgs boson produced in association with two jets at the LHC.* Eur. Phys. J. C **81** (2021) no. 11, 989, `arXiv:2105.07972 [hep-ph]`.

[44] S. De, V. Rentala, and W. Shepherd, *Measuring the polarization of boosted, hadronic W bosons with jet substructure observables.* `arXiv:2008.04318 [hep-ph]`.

[45] M. Grossi, J. Novak, B. Kersevan, and D. Rebuzzi, *Comparing traditional and deep-learning techniques of kinematic reconstruction for polarization discrimination in vector boson scattering.* Eur. Phys. J. C **80** (2020) no. 12, 1144, `arXiv:2008.05316 [hep-ph]`.

[46] T. Kim and A. Martin, *A $W^{\pm}$ polarization analyzer from Deep Neural Networks.* `arXiv:2102.05124 [hep-ph]`.

[47] A. Dey and T. Samui, *Jet Substructure and Multivariate Analysis Aid in Polarization Study of Boosted, Hadronic W Fatjet at the LHC.* `arXiv:2110.02773 [hep-ph]`.

[48] L. Ricci and M. Riembau, *Energy correlators of hadronically decaying electroweak bosons.* Phys. Rev. D **106** (2022) no. 11, 114010, `arXiv:2207.03511 [hep-ph]`.

[49] J. Searcy, L. Huang, M.-A. Pleier, and J. Zhu, *Determination of the WW polarization fractions in $pp \rightarrow W^{\pm}W^{\pm}jj$ using a deep machine learning technique.* Phys. Rev. D **93** (2016) no. 9, 094033, `arXiv:1510.01691 [hep-ph]`.

[50] J. Lee, N. Chanon, A. Levin, J. Li, M. Lu, Q. Li, and Y. Mao, *Polarization fraction measurement in same-sign WW scattering using deep learning.* Phys. Rev. D **99** (2019) no. 3, 033004, `arXiv:1812.07591 [hep-ph]`.

[51] J. Lee, N. Chanon, A. Levin, J. Li, M. Lu, Q. Li, and Y. Mao, *Polarization fraction measurement in ZZ scattering using deep learning.* Phys. Rev. D **100** (2019) no. 11, 116010, `arXiv:1908.05196 [hep-ph]`.

[52] J. Li, C. Zhang, and R. Zhang, *Polarization measurement for the dileptonic channel of W+W- scattering using generative adversarial network.* Phys. Rev. D **105** (2022) no. 1, 016005, `arXiv:2109.09924 [hep-ph]`.

[53] K. Ozdemir, *Study of Kinematics of the Jets in VBS Process at the Large Hadron Collider.* `arXiv:2305.16552 [hep-ex]`.

[54] K. Kondo, *Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models.* J. Phys. Soc. Jap. **57** (1988) 4126–4140.

[55] K. Kondo, *Dynamical likelihood method for reconstruction of events with missing momentum. 2: Mass spectra for $2 \to 2$ processes.* J. Phys. Soc. Jap. **60** (1991) 836–844.

[56] M. Diehl and O. Nachtmann, *Optimal observables for the measurement of three gauge boson couplings in $e^+e^- \to W^+W^-$.* Z. Phys. C **62** (1994) 397–412.

[57] M. Diehl and O. Nachtmann, "Optimal observables for measuring three gauge boson couplings in $e^+e^- \to W^+W^-$," in *Physics with e^+e^- Linear Colliders (The European Working Groups 4 Feb - 1 Sep 1995: Session 2)*, pp. 301–304. 3, 1996. `arXiv:hep-ph/9603207`.

[58] P. Janot, *Top-quark electroweak couplings at the FCC-ee.* JHEP **04** (2015) 182, `arXiv:1503.01325 [hep-ph]`.

[59] C. Uhlemann and N. Kauer, *Narrow-width approximation accuracy.* Nucl. Phys. B **814** (2009) 195–211, `arXiv:0807.4112 [hep-ph]`.

[60] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations.* JHEP **03** (2013) 015, `arXiv:1212.3460 [hep-ph]`.

[61] R. G. Stuart, *General renormalization of the gauge invariant perturbation expansion near the Z^0 resonance.* Phys. Lett. B **272** (1991) 353–358.

[62] R. G. Stuart, *Gauge invariance, analyticity and physical observables at the Z^0 resonance.* Phys. Lett. B **262** (1991) 113–119.

[63] A. Aeppli, G. J. van Oldenborgh, and D. Wyler, *Unstable particles in one loop calculations.* Nucl. Phys. **B428** (1994) 126–146, `arXiv:hep-ph/9312212 [hep-ph]`.

[64] A. Aeppli, F. Cuypers, and G. J. van Oldenborgh, *$\mathcal{O}(\Gamma)$ corrections to W pair production in e^+e^- and $\gamma\gamma$ collisions.* Phys. Lett. **B314** (1993) 413–420, `arXiv:hep-ph/9303236 [hep-ph]`.

[65] A. Denner, S. Dittmaier, M. Roth, and D. Wackeroth, *Electroweak radiative corrections to $e^+e^- \to WW \to 4$ fermions in double pole approximation: The RACOONWW approach.* Nucl. Phys. **B587** (2000) 67–117, `arXiv:hep-ph/0006307 [hep-ph]`.

[66] A. Denner, S. Dittmaier, M. Roth, and L. H. Wieders, *Electroweak corrections to charged-current $e^+e^- \to 4$ fermion processes: Technical details and further results.* Nucl. Phys. **B724** (2005) 247–294, `arXiv:hep-ph/0505042 [hep-ph]`. [Erratum: Nucl. Phys. **B854** (2012) 504].

[67] F. Boudjema and R. K. Singh, *A Model independent spin analysis of fundamental particles using azimuthal asymmetries.* JHEP **07** (2009) 028, `arXiv:0903.4705 [hep-ph]`.

[68] **NNPDF** Collaboration, R. D. Ball *et al.*, *Parton distributions from high-precision collider data.* Eur. Phys. J. C **77** (2017) no. 10, 663, `arXiv:1706.00428 [hep-ph]`.

[69] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *LHAPDF6: parton density access in the LHC precision era.* Eur. Phys. J. **C75** (2015) 132, `arXiv:1412.7420 [hep-ph]`.

[70] S. Dittmaier and M. Krämer, *Electroweak radiative corrections to W-boson production at hadron colliders.* Phys. Rev. **D65** (2002) 073007, `arXiv:hep-ph/0109062 [hep-ph]`.

[71] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations.* JHEP **07** (2014) 079, `arXiv:1405.0301 [hep-ph]`.

[72] A. Denner, J.-N. Lang, M. Pellen, and S. Uccirati, *Higgs production in association with off-shell top-antitop pairs at NLO EW and QCD at the LHC.* JHEP **02** (2017) 053, `arXiv:1612.07138 [hep-ph]`.

[73] 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 PYTHIA 8.2.* Comput. Phys. Commun. **191** (2015) 159–177, `arXiv:1410.3012 [hep-ph]`.

[74] S. Actis, A. Denner, L. Hofer, A. Scharf, and S. Uccirati, *Recursive generation of one-loop amplitudes in the Standard Model.* JHEP **04** (2013) 037, `arXiv:1211.6316 [hep-ph]`.

[75] S. Actis, A. Denner, L. Hofer, J.-N. Lang, A. Scharf, and S. Uccirati, *RECOLA: REcursive Computation of One-Loop Amplitudes.* Comput. Phys. Commun. **214** (2017) 140–173, `arXiv:1605.01090 [hep-ph]`.

[76] I. Neutelings, *Tikz.org: Neural networks.* Available at `https://tikz.net/neural_networks/` under Creative Commons 2.5 Licence.

[77] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, vol. 15 of *Proceedings of Machine Learning Research*, pp. 315–323. PMLR, Apr, 2011.

[78] D. A. Roberts, S. Yaida, and B. Hanin, *The principles of deep learning theory.* Cambridge University Press, 2022. `arXiv:2106.10165 [cs.LG]`.

[79] S. Ruder, *An overview of gradient descent optimization algorithms.* `arXiv:1609.04747 [cs.LG]`.

[80] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`. Accessed: 7 Apr 2023.

[81] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, *How transferable are features in deep neural networks?* Advances in neural information processing systems **27** (2014) 3320–3328, `arXiv:1411.1792 [cs.LG]`.

[82] A. Bogatskiy, B. Anderson, J. T. Offermann, M. Roussi, D. W. Miller, and R. Kondor, *Lorentz Group Equivariant Neural Network for Particle Physics*, 2020. `arXiv:2006.04780 [hep-ph]`.

[83] A. Jacot, F. Gabriel, and C. Hongler, *Neural tangent kernel: Convergence and generalization in neural networks.* Advances in neural information processing systems **31** (2018) .