Determination of the WW polarization fractions in $pp \to W^\pm W^\pm jj$ using a deep machine learning technique

Jacob Searcy,[1] Lillian Huang,[1] Marc-André Pleier,[2] and Junjie Zhu[1]

[1]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*

[2]*Brookhaven National Laboratory, Upton, New York 11973-5000, USA*

The unitarization of the longitudinal vector boson scattering (VBS) cross section by the Higgs boson is a fundamental prediction of the Standard Model which has not been experimentally verified. One of the most promising ways to measure VBS uses events containing two leptonically decaying same-electric-charge W bosons produced in association with two jets. However, the angular distributions of the leptons in the W boson rest frame, which are commonly used to fit polarization fractions, are not readily available in this process due to the presence of two neutrinos in the final state. In this paper we present a method to alleviate this problem by using a deep machine learning technique to recover these angular distributions from measurable event kinematics and demonstrate how the longitudinal-longitudinal scattering fraction could be studied. We show that this method doubles the expected sensitivity when compared to previous proposals.

PACS numbers: 13.88.+e,14.70.Fm

Studying longitudinal vector boson scattering (VBS) processes has long been a central goal of high energy colliders [1]. Without a Higgs boson, the scattering amplitude of longitudinal vector bosons ($V_L V_L \to V_L V_L$) increases with center-of-mass energy and eventually violates unitarity [2–4]. The discovery of a Higgs-like boson at the LHC [5, 6] was the first step towards understanding these interactions. However, if this Higgs boson's couplings to vector bosons deviate from the Standard Model (SM) expectation that the scattering amplitude of VBS processes can still increase with center-of-mass energy, which makes VBS a sensitive probe of anomalous Higgs couplings [7]. In addition, many new physics scenarios predict increases in VBS cross sections, through extended Higgs sectors or other new resonances [8–11]. VBS measurements hence are both a window to new physics and a constraint on fundamental properties of the Higgs boson. Measuring VBS processes at a hadron collider is experimentally challenging due to small cross sections and the difficulty of separating longitudinal states from transverse ones.

The ATLAS and CMS collaborations recently provided the first evidence for and study of a VBS process using events with two leptonically decaying same-electric-charge W bosons in association with two forward jets ($pp \to$

$pp \to W^\pm W^\pm jj$) [12, 13]. This final state has the advantage of relatively small SM background contributions compared to other VBS processes, paired with a production rate large enough to measure in early LHC data sets. While an ideal candidate for first observation of the VBS process, measuring the longitudinal fraction in these events is not straightforward since the presence of two neutrinos in the final state prevents full kinematic reconstruction of the events.

Recent studies have shown that advances in machine learning can improve the prospects for measurements at the Large Hadron Collider [14, 15]. In this paper we explore a machine learning technique that has not previously been used in the experimental high energy physics community: regression with deep neural networks. We apply this method to the difficult problem of measuring longitudinal VBS in $W^\pm W^\pm jj$.

In general, the polarization of a gauge boson can be determined from the angular distribution of its decay products. The differential cross section of a ~~leptonically-decaying~~ leptonically decaying W boson is related to the polarization fractions as [16] ~~:~~

$$\frac{d\sigma}{d\cos\theta^*} \propto \frac{3}{8}f_-(1\mp\cos\theta^*)^2 + \frac{3}{8}f_+(1\pm\cos\theta^*)^2$$
$$+\frac{3}{4}f_L(1-\cos^2\theta^*), \text{for } W^\pm, \tag{1}$$

where θ^* is the angle between the charged lepton in the boson rest frame and the W boson direction of motion. Fraction parameters f_-, f_+, and f_L denote the fractions of events with the three possible polarization states of the W boson, -1, $+1$, and 0, respectively. They are constrained via $f_- + f_+ + f_L = 1$. In order to measure θ^*, we need to fully reconstruct the direction of motion of the W boson.

Requiring both W bosons to decay leptonically in $pp \to W^\pm W^\pm jj$ events enables the determination of the electric charge of each W boson via the charged leptons. However, since the corresponding two neutrinos in the final state are not detected, the W boson rest frames cannot be directly measured. It is thus difficult to determine polarization fractions of each boson and the fraction of longitudinal scattering events in the $W^\pm W^\pm jj$ process.

Many proposals have been made to determine the longitudinal fractions in other VBS final states, such as ~~semi-leptonic~~ semileptonic W^+W^- [17] and $W^\pm Z$ [18] or ~~fully-leptonic~~ fully leptonic decay modes of $W^\pm Z$ and ZZ, where the full event kinematics can be reconstructed or estimated using the W boson mass constraint. However, these channels suffer from large SM backgrounds that are not present in the $W^\pm W^\pm jj$ channel. Attempts have been made to gain sensitivity through other variables [1, 18–21] than θ^* in the $W^\pm W^\pm jj$ channel. One example is the variable $R_{pT} = (p_T^{\ell 1} \times p_T^{\ell 2})/(p_T^{j1} \times p_T^{j2})$ [20], where ℓ_1 and ℓ_2 denote the two leptons in no particular order and j_1 and j_2 denote the two most energetic jets in the event. This single variable does not encompass all of the sensitivity to

longitudinal scattering, and better discrimination can be achieved by combining the available event information with a machine learning technique. Therefore, we develop a method to use a neural network (NN) to map measurable quantities to the true $\cos\theta^*$ values that contain the event's polarization information. This approximation mitigates the limitation of missing kinematic information in this final state, and makes $W^{\pm}W^{\pm}jj$ a promising channel for observing the behavior of longitudinal W boson scattering.

While it has become a common practice in high energy physics to use multivariate techniques to separate signal from background, multivariate regression is not commonly used to measure underlying physics quantities of interest. Unlike classification, where the goal of the neural network is to produce discrete assignments, for example signal and background, NN regression relies on the fact that a neural network is a universal approximator [22] to instead approximate an unknown continuous function. The goal of our NN is to find the best approximation of the two truth values of $\cos\theta^*$ (one for each W boson) present in each event, using measurable quantities. Similar techniques are currently in use to address the problem of estimating parton distribution functions [23].

It has also become common practice to train these multivariate methods using variables built from basic measurable event quantities to be sensitive to a given physics process [24]. These high-level variables add some understanding of the underlying physics to the training variables and can produce better results. However, it has also recently been observed that extending a single layer neural network to a "deep" network containing many layers can regain most of the sensitivity produced by these high-level variables [14, 15]. Since only a few high-level variables have been proposed for this process we choose to use deep neural networks for maximum sensitivity.

$W^{\pm}W^{\pm}jj$ events have the signature of two quarks, two same sign leptons, and two undetected neutrinos. We use all basic measurable object kinematics as input variables to the neural network: the transverse momentum (p_T), pseudorapidity (η), and azimuthal angle (ϕ) of the two leptons and two jets, and x and y components of the missing transverse energy ($\not{E}_T^x$ and $\not{E}_T^y$). The overall number of measurable quantities used hence is 14. The goal of the multivariate technique is to find the best mapping from these measurable quantities to the two truth values of $\cos\theta^*$ (one for each W boson) present in each event.

Training deep NNs has been the subject of intensive study and a good review of some of this work is presented in [25]. For completeness we list some of the properties of the NN that we utilize. We choose a multilayer neural network with a two node final output layer with linear activation to approximate the true $\cos\theta^*$ distribution of each W boson. The NN is implemented with the THEANO software packages [26, 27]. The cost function is

FIG. 1: A comparison of the truth level $\cos\theta^*$ and the NN output $\cos\theta^{NN}$ distributions for $--$, $++$, and LL events (a), R_{pT} templates for the corresponding polarization states with log scaling (b), and a comparison of the $\cos\theta^{NN}$ for the signal and dominant $WZjj$ background (c) are shown with arbitrary units (arb. units). (a), (c) Projections of the 2D distributions onto one of the two leptons, which are identical to the projections onto the other lepton due to arbitrary sorting.

defined as

$$
\mathcal{C} = \frac{1}{N} \sum_{i=1}^{N} [(\cos\theta_{1,i}^* - \cos\theta_{1,i}^{NN})^2 + (\cos\theta_{2,i}^* - \cos\theta_{2,i}^{NN})^2],
\tag{2}
$$

where N is the number of events per minibatch, $\cos\theta_{1/2,i}^*$ is the truth value of $\cos\theta^*$ for each W boson with random ordering for the ith event, and $\cos\theta_{1/2,i}^{NN}$ is the value of the two neural network outputs. A stochastic gradient descent algorithm is used to train the neural network by minimizing the cost function. Hyperparameters are tuned by hand and confirmed by a local grid search, with the best performance given by a deep network with 20 layers each with 200 hidden nodes. This network yields a 20% better cost value in a validation data set than the best single layer NN tested.

Signal $W^\pm W^\pm jj$ events are generated using the MADGRAPH event generator [28] at a proton-proton center-of-mass energy of 13 TeV. The invariant mass of the two outgoing partons was required to be greater than 150 GeV. The W bosons are decayed, assuming they are on shell and have no spin correlations, with the DECAY routine provided with MADGRAPH. We generated 500,000 events for training and another million for testing and validation.

Polarization fractions can then be obtained experimentally by fitting the two-dimensional distribution of the NN output $\cos\theta^{NN}$. In order to fit for these polarization fractions templates must be built for "pure" polarization states. These templates are created using generator level helicity information. In addition, a method was tested that used truth level reweighting and included all spin correlations and off-shell effects, and the results were found to be comparable.

Since there are two W bosons in each event, six distinct polarization states are possible. Events where both W bosons have a polarization of 0 are referred to as LL, of 1 as $++$, and of -1 as $--$. Events with differing polarizations of $(-1,1)$ or $(1,-1)$ are referred to as $+-$, $(-1,0)$ or $(0,-1)$ as $L-$, and likewise $(1,0)$ or $(0,1)$ as $L+$.

Figure 1(a) shows the comparison between the truth $\cos\theta^*$ and the NN output $\cos\theta^{NN}$ for $--$, $++$, and LL events ($+-/L+/L-$ are omitted for clarity, but closely resemble combinations of the templates shown). As expected,

FIG. 2: The two-dimensional distribution of $\cos\theta_1^{NN}$ versus $\cos\theta_2^{NN}$ for the two leptons in each event is shown for each of the three polarization states: TT, TL, and LL, respectively.

$\cos\theta^{NN}$ has less separation power for the different polarization states compared to $\cos\theta^*$ due to missing information for the two final state neutrinos. However, reasonable discrimination between each polarization state can clearly be seen from these distributions. Figure 1(b) shows the R_{pT} distribution for $--$, $++$, and LL events. The discrimination power is seen only for large values of R_{pT}, and is only apparent with a logarithmic scale. Figure 1(c) shows the $\cos\theta^{NN}$ distribution for the signal $W^\pm W^\pm jj$ process and an important background process: $WZjj$ production. Reasonable separation power is observed, which could be utilized in a combined fit. In an actual data analysis, the $WZjj$ component would be subtracted as background from the observed data before fitting the polarization fractions.

Having established templates for each polarization state and a distribution that is sensitive to different polarization states, we fit the two-dimensional $\cos\theta_1^{NN}$ versus $\cos\theta_2^{NN}$ distribution in pseudo data to derive each polarization fraction. Five equal-size bins are used for each $\cos\theta^{NN}$ variable ranging from -1 to 1. A maximum likelihood fit is performed within the RooFit framework [29]. We combine events with both W bosons transversely polarized as "TT" (the sum of the $--$, $+-$, and $++$ combinations), events with one W boson transversely polarized and one W boson longitudinally polarized as "TL" (the sum of the $L-$ and $L+$ combinations), and events with both W bosons longitudinally polarized as "LL." This reduces the free fitting parameters from five to two and allows for a better constraint on the LL scattering fraction of interest, under the assumption that the relative admixture of contributions within TT and TL does not change. The two-dimensional distributions for these templates are shown in Fig. 2. Statistical fit uncertainties are determined by randomly fluctuating data expectations within their Poisson uncertainties and repeating the fit, and confidence intervals are derived from these toy experiments.

Fits are performed in a range of integrated luminosities from 0.01–3 ab^{-1}. An example fit for 1 ab^{-1} is shown in Fig. 3 where the pseudo data are compared to the sum of the contributions from the TT, TL, and LL components. It is found that the LL fraction can be measured with a 68% confidence limit of $(6.7\pm1.4)\%$ with an ultimate luminosity of 3 ab^{-1}. When a similar fit is applied to the R_{pT} variable the precision of $(6.7^{+7.1}_{-6.7})\%$ is found to be consistent with 0. The regression technique hence greatly enhances the sensitivity to this process.

The above measurements are performed with parton level predictions. While they show encouraging results it is important to also check to see if this procedure will stand up to the experimental reality of finite detector resolution,

FIG. 3: One example fit where the pseudo data are compared to the sum of contributions from the TT, TL, and LL components. There are five groups with five bins inside each group. These five groups represent $\cos\theta_2^{NN}$ from -1 to 1 with a step size of 0.4, while five bins inside each group represent $\cos\theta_1^{NN}$ from -1 to 1 with a step size of 0.4.

and the event level selection that will be required to remove backgrounds from this analysis. To study the effects of event level cuts, we apply additional selection criteria as used by the ATLAS Collaboration [12] to obtain a tighter fiducial region which is dominated by the contribution from electroweak production of $W^{\pm}W^{\pm}jj$ events: jet $p_T > 30$ GeV, lepton $p_T > 25$ GeV, missing transverse energy $\not{E}_T > 40$ GeV, dijet mass $M_{jj} > 500$ GeV, and dijet pseudorapidity difference $|\Delta\eta_{jj}| > 2.4$.

To emulate the response of a typical general-purpose LHC detector, these events are passed through the application of parton showering in PYTHIA [30] and then through the response simulation of the CMS detector implemented in DELPHES [31]. This detector smearing adds some degree of realism, but could neglect various effects due to the large number of overlapping interactions during the high luminosity LHC runs. Since these effects are often mitigated with specific reconstruction techniques and need detailed detector modeling, we leave studies of this nature to dedicated efforts by the experiments. After detector simulation, parton level quantities are then approximated by taking the leading two jets clustered with an anti-k_t algorithm [32] with jet size parameter $R = 0.5$.

Backgrounds to the $W^{\pm}W^{\pm}jj$ process depend largely on experimental choice, and require detailed simulation. It can be seen in Fig. 1 that background from the $WZjj$ process (where one of the leptons from the Z boson decay is not detected) has a different $\cos\theta^{NN}$ shape than the $W^{\pm}W^{\pm}jj$ events, but it is likely event level cuts will still need to be applied to reduce this component. Determining the systematics on the background shapes will require significant work from the experiments, and is not treated here. However, as already mentioned the backgrounds in the $W^{\pm}W^{\pm}jj$ channel are relatively small compared to other channels.

We determine the precision that can be achieved for fractions of TT, TL, and LL components using three different scenarios increasing in realism: (a) using all generated events at the parton level; (b) using events passing the additional selection criteria used by ATLAS at the parton level; (c) using events processed with the DELPHES detector simulation and with reconstructed objects passing additional selection criteria used by ATLAS.

The precision for the three polarization fractions as a function of the integrated luminosity is presented in Fig. 4. TT components can be measured with great precision, whereas separating pure LL scattering from LT scattering is challenging. In the most difficult and realistic scenario, (c), the cuts and object efficiencies slightly increase the mean LL fraction to 7.0% and 68% confidence limits are found $7^{+19}_{-7}\%$ ($7^{+5}_{-6}\%$) for

an integrated luminosity of 0.1 (3) ab^{-1}. Equivalent limits from fits to the R_{pT} variable are found to be $7^{+29}_{-7}\%$ and $7^{+9}_{-7}\%$. The fit to the neural network almost doubles the ultimate precision to anomalous LL fractions in this scenario. Reaching the same statistical sensitivity using the R_{pT} variable would require approximately 10 ab^{-1}, more than three times the total expected luminosity of the LHC program. We have shown that large sensitivity gains can be made with NN regression. In scenario (c) our simple estimate falls short of the 5σ criteria for observation of longitudinal VBS, however, new physics beyond the SM could greatly enhance this fraction [8], making stringent limits or a future observation of this fraction very important. A comparison of scenarios (a) and (b) to (c) illustrates potential gains in sensitivity through cut optimization and improved detector performance. In addition, the authors hope that experiments can improve on this methodology by training on fully simulated events or by improving detector performance (e.g. through upgrades) to enhance sensitivity.

In conclusion, we present a method to determine the WW polarization fractions in $W^{\pm}W^{\pm}jj$ events by using a deep machine learning technique. This method allows us to recover the charged lepton angular distributions in the W boson rest frame from measurable event kinematics. The results obtained from this method show greatly enhanced sensitivity over the example R_{pT} variable. Cuts to reject backgrounds as well as finite detector resolutions reduce the sensitivity as expected, but the method remains a powerful tool for the study of polarization fractions in VBS events, almost doubling the ultimate precision when compared to R_{pT}.

FIG. 4: 68% (yellow) and 95% (green) expected confidence intervals for the measured LL fraction f_{LL} (top panels), the TL fraction f_{TL} (middle panels), and the TT fraction f_{TT} (bottom panels) as a function of the integrated luminosity for, from left to right, scenarios (a), (b), and (c) discussed in the text.

[1] D. A. Dicus, J. F. Gunion, and R. Vega, Phys. Lett. B **258**, 475 (1991).

[2] J. G. Veltman, Acta Phys. Pol. B **8**, 475 (1977).

[3] B. W. Lee, C. Quigg, and H. B. Thacker, Phys. Rev. Lett. **38**, 883 (1977).

[4] B. W. Lee, C. Quigg, and H. B. Thacker, Phys. Rev. D **16**, 1519 (1977).

[5] G. Aad et al. (ATLAS Collaboration), Phys. Lett. B **716**, 1 (2012).

[6] S. Chatrchyan et al. (CMS Collaboration), Phys. Lett. B **716**, 30 (2012).

[7] J. M. Campbell and R. K. Ellis, J. High Energy Phys. 04 (2015) 030.

[8] A. Alboteanu, W. Kilian, and J. Reuter, J. High Energy Phys. 11 (2008) 010.

[9] W. Kilian, T. Ohl, J. Reuter, and M. Sekulla, Phys. Rev. D **91**, 096007 (2015).

[10] S. Godfrey and K. Moats, Phys. Rev. D **81**, 075026 (2010).

[11] C.-W. Chiang, S. Kanemura, and K. Yagyu, Phys. Rev. D **90**, 115025 (2014).

[12] G. Aad et al. (ATLAS Collaboration), Phys. Rev. Lett. **113**, 141803 (2014).

[13] V. Khachatryan et al. (CMS Collaboration), Phys. Rev. Lett. **114**, 051801 (2015).

[14] P. Baldi, P. Sadowski, and D. Whiteson, Nat. Commun. **5**, 4308 (2014).

[15] P. Baldi, P. Sadowski, and D. Whiteson, Phys. Rev. Lett. **114**, 111801 (2015).

[16] R. K. Ellis, W.-J. Stirling, and B.-R. Webber, Cambridge Monogr. Part. Phys., Nucl. Phys. Cosmol. **8**, 1 (1996).

[17] T. Han, D. Krohn, L.-T. Wang, and W. Zhu, J. High Energy Phys. 03 (2010) 082.

[18] A. Ballestrero, D. B. Franzosi, and E. Maina, J. High Energy Phys. 06 (2011) 013.

[19] A. Freitas and J. S. Gainer, Phys. Rev. D **88**, 017302 (2013).

[20] K. Doroba, J. Kalinowski, J. Kuczmarski, S. Pokorski, J. Rosiek et al., Phys. Rev. D **86**, 036011 (2012).

[21] M. Fabbrichesi, M. Pinamonti, A. Tonero, and A. Urbano, Vector boson scattering at the LHC: A study of the $WW \to WW$ channels with the Warsaw cut, Phys. Rev. D **93**, 015004 (2016).

[22] G. Cybenko, Math Control Signals Syst. **2**, 303 (1989).

[23] S. Forte, L. Garrido, J.-I. Latorre, and A. Piccione, J. High Energy Phys. 05 (2002) 062.

[24] P.-C. Bhat, Annu. Rev. Nucl. Part. Sci. **61**, 281 (2011).

[25] J. Schmidhuber, Neural Netw. **61**, 85 (2015).

[26] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio, in *Proceedings of the 9th Python in Science Conference, Austin, TX, 2010*, edited by S. van der Walt and J. Millman (2010), p. 3.

[27] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I.-J. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley, and Y. Bengio, arXiv:1211.5590.

[28] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni *et al.*, J. High Energy Phys. 07 (2014) 079.

[29] W. Verkerke and D.-P. Kirkby, *Proceedings of the 2003 Computing in High Energy and Nuclear Physics (CHEP03), La Jolla, CA, 2003,* eConf C0303241, MOLT007 (2003).

[30] T. Sjöstrand, S. Mrenna, and P.-Z. Skands, Comput. Phys. Commun. **178**, 852 (2008).

[31] J. de Favereau *et al.* (DELPHES3 Collaboration), J. High Energy Phys. 02 (2014) 057.

[32] M. Cacciari, G.-P. Salam, and G. Soyez, J. High Energy Phys. 04 (2008) 063.