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-Andre 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*

(Dated: July 13, 2015)

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. In the first LHC run, both ATLAS and CMS collaborations presented first studies of VBS processes in events with two leptonically-decaying same-electric-charge W bosons produced in association with two jets. This channel has the advantage of smaller backgrounds compared to other VBS channels while still exhibiting a detectable production rate. 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 circumvent this problem by using a deep machine learning technique to recover these angular distributions from measurable event kinematics to determine the longitudinal fraction. We also compare the results obtained from this method and from other traditional methods. The method presented here could be used as well in other VBS channels with neutinos(s) in the final state.

PACS numbers: 14.70m

The discovery of a Higgs-like boson at the LHC [1, 2] was the first step toward a better understanding of the electroweak symmetry breaking (EWSB) mechanism. One important, and unverified, prediction of the Standard Model (SM) is that the scattering amplitude of longitudinal vector bosons ($V_L V_L \to V_L V_L$) is unitarized by the Higgs boson. Measuring VBS processes at a hadron collider, however, is experimentally challenging. The ATLAS and CMS collaborations recently provided the first evidence for and study of a VBS process using events with two same-electric-charge W bosons in association with two forward jets ($pp \to W^\pm W^\pm jj$) [3, 4]. 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 datasets. While an ideal candidate for first observation of the VBS process, measuring the longitudinal fraction of these events is not straight forward.

In general the polarization of a gauge boson can be determined from the angular distribution of its decay products. Since a W boson only couples to left-handed particles and right-handed anti-particles, the decayed charged lepton is expected to preferentially pointing along the boson direction of motion for a left-handed W^+ and along the opposite boson direction of motion for a left-handed W^-. The normalized differential cross section of a leptonically-decaying W boson can be written in terms of polarization fractions as [**?**]:

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

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 W events with three possible polarization states -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 the $pp \to W^\pm W^\pm jj$ events enables the determination of the electric charge of each W boson via the charged leptons. However, 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 W^+W^- [5], $W^\pm Z$ [**?**] and ZZ [**?**] or 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 either suffer from large SM backgrounds not present in the $W^\pm W^\pm jj$ channel or have relatively low production cross sections. Attempts have been made to gain sensitivity through other variables than θ^* in the $W^\pm W^\pm jj$ channel [**?**]. One example is the variable $R_{p_T} = (p_T^{\ell 1} \times p_T^{\ell 2})/(p_T^{j_1} \times p_T^{j_2})$ [6], 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. It is natural to assume that not all of the sensitivity to longitudinal scattering is encompassed in a single variable, and that better discrimination could be obtained by combining the available event information with a machine learning technique. In this paper we develop a method to use a neural network to map measurable quantities to the true $\cos\theta^*$ values that contain the events polarization information.

While it has become common practice in high energy physics to use multi-variate techniques to separate signal from background, to the author's knowledge multi-variate regression has not been used to directly predict underlying quantities **also not in H properties?**. Recent success with deep learning in other areas of HEP are presented in [7, 8].

For the $W^{\pm}W^{\pm}jj$ events, we use representative measurable quantities such as the transverse momentum (p_T) [**?**], pseudorapidity (η) and azimuthal angle (ϕ) of the two leptons and two jets, and x- and y-components of missing transverse energy ($\not{E}_T^x$ and $\not{E}_T^y$). The overall number of measurable quantities used is 14. The goal of the multi-variate 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. We choose a multi-layer neural network with a final output layer with linear activation. The neural network is implemented with the Theano software packages [**?**]. Hyper-parameters are tuned by hand, but undoubtedly could be improved. The cost function is defined as $\mathcal{C} = \sum_{i=1}^{N}[(\cos\theta^*_{1,i} - \cos\theta^{NN}_{1,i})^2 + (\cos\theta^*_{2,i} - \cos\theta^{NN}_{2,i})^2]/N$, where N is the number of events per mini-batch, $\cos\theta^*_{1/2,i}$ is the truth value of $\cos\theta^*$ for each W boson with random ordering for the i-th event, and $\cos\theta^{NN}_{1/2,i}$ is the value of the two neural network outputs. Stochastic gradient descent algorithm [**?**] is used to train the neural network.

Signal $W^{\pm}W^{\pm}jj$ events are generated using the MADGRAPH event generator [**?**] at a proton-proton center-of-mass energy of 13 TeV. The NN23l01 **find more correct name for this PDF set** parton distribution function [**?**] is used as the default. The following selection criteria are applied during the event generation: outgoing parton $p_T > 20$ GeV and $|\eta| < 5$, lepton $p_T > 10$ GeV and $|\eta| < 2.5$, and the invariant mass of the two outgoing partons must be greater than 150 GeV. The resulting cross section at 13 TeV is 8.4 fb, which is used to normalize the expected number of signal events**why did we remove this?**. To emulate the response of a typical general-purpose LHC detector, these events are parton-showered using PYTHIA[**?**] and passed through the response simulation of the CMS **it really is the CMS detector not ATLAS** detector implemented in DELPHES [**?**]. **mention jet alg/size etc. here** Events are split into three categories: 1/4 are used in a training sample, 1/4 are used for a validation test against over-training, and the remaining 1/2 are used to build templates and perform sensitivity studies. A ** layer neural network with ** hidden neurons and a learning rate of ** is used **I can move thie earlier, but I also didn't use a fixed learning rate** .

The dominant background **only true for ATLAS** to the $W^{\pm}W^{\pm}jj$ process comes from the $WZjj$ process where one of the leptons from the Z boson decay is not detected or reconstructed. We also use the MADGRAPH event generator to produce $WZjj$ events.

Polarization fractions can then be obtained by experimentally 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 by reweighting each event based on the truth $\cos\theta^*$ distribution. The event weight W_i for polarization state i is given by $W_i = F_i/n$, where n is used for the normalization and is defined as

$$n = [\frac{3}{8}f_-(1\mp\cos\theta_1^*)^2 + \frac{3}{8}f_+(1\pm\cos\theta_1^*)^2 + \frac{3}{4}f_L(1-\cos^2\theta_1^*)] \times [\frac{3}{8}f_-(1\mp\cos\theta_2^*)^2 + \frac{3}{8}f_+(1\pm\cos\theta_2^*)^2 + \frac{3}{4}f_L(1-\cos^2\theta_2^*)]. \quad (2)$$

The index i represents one of six possible polarization states for the two W bosons: Left-Left ($--$), Left-Right ($-+$), Right-Right($++$), Left-Longitudinal ($-L$), Right-Longitudinal ($+L$), or Longitudinal-Longitudinal (LL) **no reason to name them Left/Right if we are using +/-**. F_i is defined as

$$F_i \in \begin{pmatrix} --= f_-^2(1\mp\cos\theta_1^*)^2(1\mp\cos\theta_2^*)^2, \\ -+= f_-f_+[(1\mp\cos\theta_1^*)^2(1\pm\cos\theta_2^*)^2 \\ +(1\pm\cos\theta_1^*)^2(1\mp\cos\theta_2^*)^2], \\ ++= f_+^2(1\pm\cos\theta_1^*)^2(1\pm\cos\theta_2^*)^2, \\ -L = f_-f_L[(1\mp\cos\theta_1^*)^2(1-\cos^2\theta_2^*) \\ +(1-\cos^2\theta_1^*))(1\mp\cos\theta_2^*)^2], \\ +L = f_+f_L[(1\pm\cos\theta_1^*)^2(1-\cos^2\theta_2^*)) \\ +(1-\cos^2\theta_1^*)(1\pm\cos\theta_2^*)^2], \\ LL = f_L^2(1-\cos^2\theta_1^*))(1-\cos^2\theta_2^*) \end{pmatrix}. \quad (3)$$

Since no ordering is applied to the two bosons we require that the individual polarization fractions f_-,f_+,f_L are the same for both W bosons. For reweighting the original sample f_-, f_+, f_L are take as a function of the invariant mass of the diboson system (M_{WW}). Weights are calculated before any additional event level cuts are made, and the resulting templates are remade for each set of cuts explored. **Does this mean you got the MG5 scripts to work and add the helicity to the lhe? A. No this was never done removed this.**

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 omiited for clearity but closely resemble combinations of the templates shown). As expected, $\cos\theta^{NN}$ has less seperation power to 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_{p_T} distribution for $--$, $++$, and LL events. The discrimination power is seen only for large values of R_{p_T}, and only apparent with a logorithm scale.

Figure 1(c) shows the $\cos\theta^{NN}$ distribution between the signal $W^{\pm}W^{\pm}jj$ process and the background $WZjj$ process. Reasonable separation power can be found. In actual data analysis, component would be subtracted as background from observed data before fitting for the polarization fractions.

Having established templates for each polarization state and a distribution that is sensitive to different polarization states, all we have to do is fit the two-dimensional $\cos\theta_1^{NN}$ vs $\cos\theta_2^{NN}$ distribution in data to

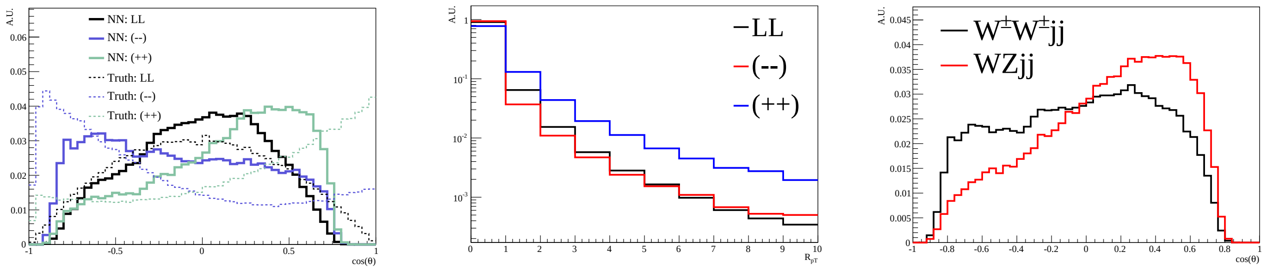

FIG. 1. Comparison of the truth level $\cos\theta^*$ and the NN output $\cos\theta^{NN}$ distributions for $--$, $++$ and LL events (Left), R_{p_T} templates for the corresponding polarization states (Middle), and comparison of the $\cos\theta^{NN}$ for the signal and dominant $WZjj$ background (Right).

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 [**?**]. Fit uncertainties are determined by randomly fluctuating data expectations within their Poisson errors and repeating the fit, and confidence intervals are derived from these toy experiments. **(need to put more description here about 68% and 95% level limits)**

We combine events with both W bosons transversely-polarized as "TT" (the sum of $--$, $-+$ and $++$ combinations), events with one W boson transversely-polarized and one W boson longitudinally-polarized as "TL" (the sum of $-L$ and $+L$ combinations), and events with both W bosons longitudinally-polarized as "LL". This reduces the free fitting parameters from 5 to 2 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.

Figure 2 shows one example fit where the pseudo data are compared to the sum of contributions from TT, TL and LL components. A total integrated luminosity of 1 ab^{-1} is assumed.

Studies shown above are performed using all events at the parton level and show the encouraging results to use neutral network to recover the lepton angular information and improve the separation between different polarization states. It is important to also check if this procedure will stand up to experimental realities of finite detector resolution. To further reduce SM backgrounds, we apply additional selection criteria as used by the ATLAS collaboration [3] 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 $|\Delta\eta_{jj}| > 2.4$. If these criterias are applied at the generator level, jet p_T, M_{jj} and $\Delta\eta_{jj}$ are calculated using the parton-level jets and $\not{E}_T$ are calculated using two generator-level neutrinos.

We determine the precision that can be achieved for fractions of TT, TL, and LL components using four different scenarios: (a) using all generated events at the parton level; (b) using events passed these additional selection criteria used by ATLAS at the parton level; (c) using all generated events passed through the DELPHES simulation; and (d) using events passed through the DELPHES simulation and also passed these additional selection criteria used by ATLAS. The precision for the three polarization fractions as a function of the integrated luminosity are presented in Fig. 3. Transverse components can be measured with great precision, whereas separating pure longitudinal-longitudinal scattering from longitudinal-transverse scattering is challenging. The precision for the LL fraction is **% (**%) for an integrated luminosity of 100 (3000) fb^{-1}. **from this plot it seems that we can not determine the LL fraction even using 3 ab-1, is that true?**

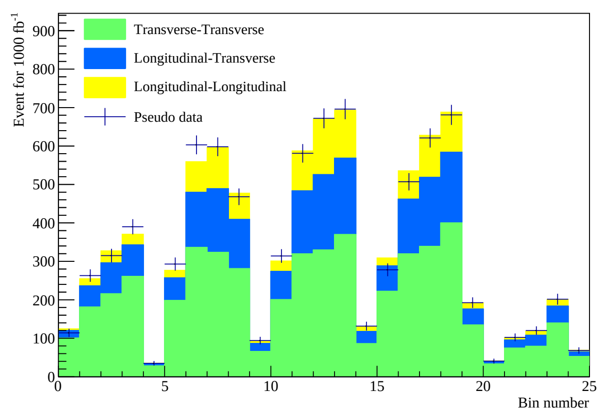

FIG. 2. One example fit where the pseudo data are compared to the sum of contributions from TT, TL and LL components. There are five groups with five bins inside each group. These five groups represent $\cos\theta_2^{NN}$ from -1 to 1 with a step size of 0.4, while five bins inside each group represent $\cos\theta_1^{NN}$ from -1 to 1 with a step size of 0.4.

We also determine the precision for TT, TL, and LL fractions by fitting the R_{p_T} distribution. The precision for the LL fraction is found to be **% (**%) for an integrated luminosity of 100 (3000) fb^{-1}. Better precision is obtained for $\cos\theta^{NN}$, which indicates that it is a more sensitive variable to different polarization states than R_{p_T}.

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 to recover the charged lepton angular distributions in the W boson rest frame from measurable event kinematics. We compare the results obtained from this method and from other traditional methods to illustrate the gain in sensitivity with our method. Cuts to reject backgrounds as well as finite detector resolutions reduces the sensitivity as expected, but the method remains a useful tool for the study of polarization fractions in VBS events.

[1] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **716**, 1 (2012) [arXiv:1207.7214 [hep-ex]].

[2] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **716**, 30 (2012) [arXiv:1207.7235 [hep-ex]].

[3] G. Aad *et al.* [ATLAS Collaboration], Phys. Rev. Lett. **113**, no. 14, 141803 (2014) [arXiv:1405.6241 [hep-ex]].

[4] V. Khachatryan *et al.* [CMS Collaboration], Phys. Rev. Lett. **114**, no. 5, 051801 (2015) [arXiv:1410.6315 [hep-ex]].

[5] T. Han, D. Krohn, L. T. Wang and W. Zhu, JHEP **1003**, 082 (2010) [arXiv:0911.3656 [hep-ph]].

[6] K. Doroba, J. Kalinowski, J. Kuczmarski, S. Pokorski, J. Rosiek, M. Szleper and S. Tkaczyk, Phys. Rev. D **86**, 036011 (2012) [arXiv:1201.2768 [hep-ph]].

[7] P. Baldi, P. Sadowski and D. Whiteson, Nature Commun. **5**, 4308 (2014) [arXiv:1402.4735 [hep-ph]].

[8] P. Baldi, P. Sadowski and D. Whiteson, Phys. Rev. Lett. **114**, no. 11, 111801 (2015) [arXiv:1410.3469 [hep-ph]].

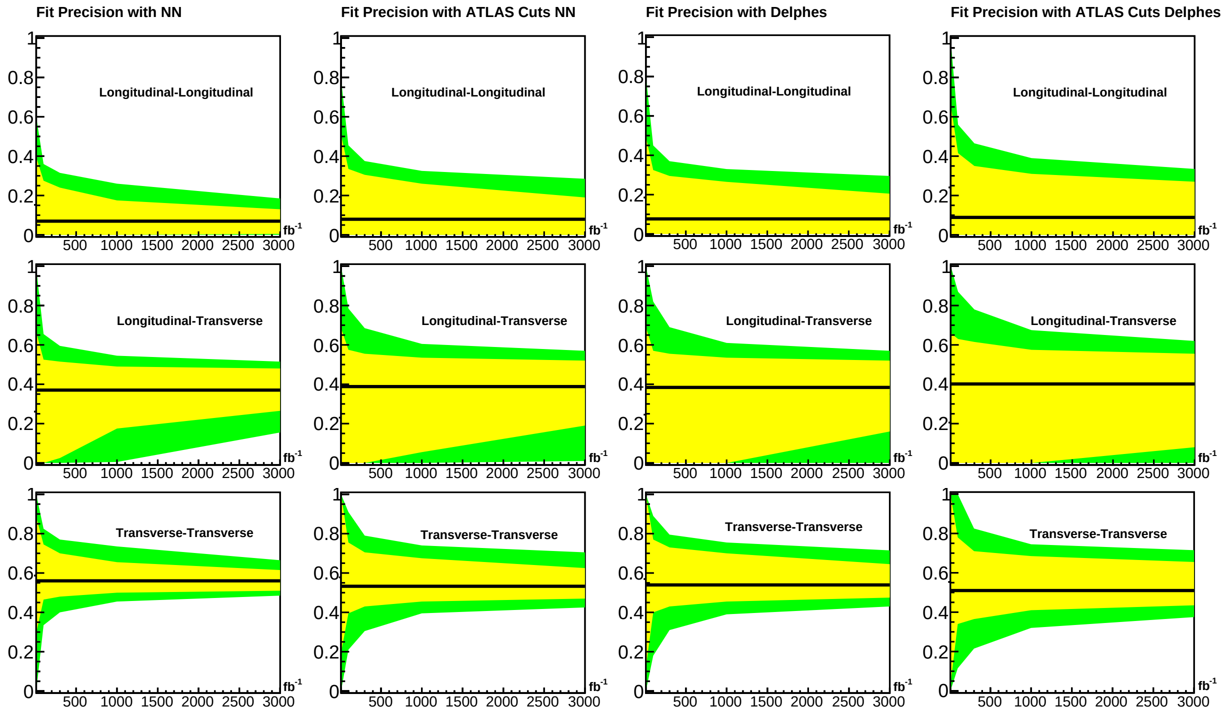

FIG. 3. Precisions for three polarization fractions as a function of the integrated luminosity for four scenarios discussed in the text.