Abstract

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, ATLAS and CMS presented the first studies of VBS in events with two leptonically decaying same sign W bosons produced in association with two jets. This VBS channel has the advantage of having very small backgrounds compared to other VBS channels. However, the two neutrinos in the final state make full kinematic event reconstruction and hence evaluation the longitudinal scattering fraction difficult. The angular distributions of the leptons in the W boson rest frame, which are commonly used to fit polarization fractions, are not readily available due to the missing information resulting from the unmeasured neutrinos. In this paper we present a method to circumvent this problem by using deep machine learning to recover the angular distributions from measurable event kinematics, and compare sensitivities to longitudinal vector boson scattering with more traditional methods.

1 Introduction

The discovery of the Higgs-like boson at the LHC [1, 2], was the first step toward determining the properties of electroweak symmetry breaking (EWSB). One important, and unverified, prediction of the Standard Model is that the scattering amplitude of longitudinal vector bosons ($V_LV_L \rightarrow V_LV_L$) is unitarized by the Higgs boson. Measuring VBS processes at a hadron collider, however, is experimentally challenging. Currently, the ATLAS and CMS collaborations have only been able study the W+,W+,jj final state [3, 4]. This final state with two same sign leptons has the advantage of relatively small background from other standard model processes including "strong" production which dominates other VBS/VBF process. While an ideal candidate for observing VBS, measuring the longitudinal fraction of these events is not straight froward. In general the polarization of a gauge boson can be determined from the angular distribution of its decay products [**?**]. For a leptonic W the differential cross section can be written in terms of polarization fractions as

$$\frac{1}{\sigma}\frac{d\sigma}{dcos(\theta^*)} = \frac{3}{8}f_L(1\mp cos(\theta^*))^2 + \frac{3}{8}f_R(1\pm cos(\theta^*))^2 + \frac{3}{4}f_0(1-cos^2(\theta^*)) \quad (1)$$

Where θ^* is the angle in the W's rest frame between the charged lepton and the W's direction in the lab frame. However, because of two neutrinos in the final state the W's rest frame cannot be directly measured. Do to these issues many proposals have been made to measure Longitudinal fractions in different final states, such as semileptonic WW [5], or ZZ (le houche report), however, these channel suffer from large backgrounds not present in the W+W+jj channel. However, due to advantages of the same sign WW state attempts have been made to gain sensitivity through other variables than θ^*. Doroba et.al. has show that the variable $R_{pt} = \frac{Ptl1*ptl2}{ptj1,ptj2}$ is sensitive to the longitudinal fraction in the W+W+jj final state [6]. It is natural to assume that all of the sensitivity to longitudinal scattering is not encompassed in a single variable, and that better discrimination could be obtained by combining all the event information with a machine learning technique. Thus the authors have developed a method using a neural network to map measurable quantities to the $cos(\theta^*)$ distribution that contains the polarization information of the W boson.

2 Machine learning model

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.

For our purpose the goal of the multi-variate technique is to find the best mapping from the 14 measurable quantities (2 Leptons and 2 Jets each with $p_T, \eta, \phi,$ and $\not{E}, \phi, E_t$) to the two values of $cos(\theta^*)$ (one for each W) present in each event. Do to recent

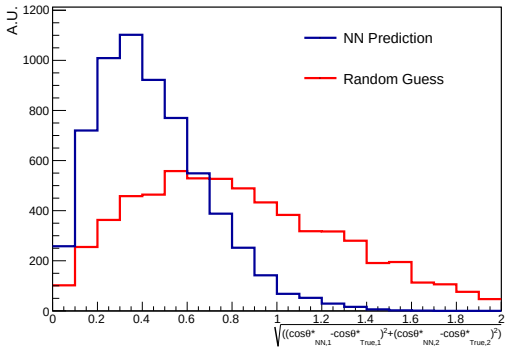

Figure 1: Distance of prediction from the true value of ($cos(\theta^*(1)$,$cos(\theta^*(2))$)

success with deep learn in other areas of HEP [8, 8]. We choose a multi-layer neural network with a final output layer with linear activation. The neural network was implemented with the Theano software packages. Hyperparameters were roughly tuned by hand, but undoubtedly could be improved. The cost function was defined as the mean error squared and stochastic gradient descent was used to train the sample.

$$F = -\sum_{i=1}^{n}[(cos(\theta^*_{1,i}) - N_{1,i})^2 + (cos(\theta^*_{2,i}) - N_{2,i})^2]/n \quad (2)$$

Where $cos(\theta^*_{1/2,i})$ is true value of $cos(\theta^*)$ for each W boson with random ordering for i-th event, and $N_{1,2/i}$ is the value of the two output layers.

Training events were split into three categorizes. 1/4 of the events where used in a training sample, 1/4 were used as a validation test against over training, and the remaining 1/2. were used to build templates and do sensitivities studies.

3 Signal Model

As our default sample, leading order electroweak-Same sign WW events were produced at parton level with madgraph [] using nn23l01 [] pdf, and the following generator level cuts.

- Parton $p_T > 20$ GeV, $|\eta| < 5$
- Lepton $p_T > 10$ GeV, $|\eta| < 2.5$
- $\Delta R(j,j), \Delta R(l,l), \Delta R(l,j) > 0.4$
- Invariant mass of the two parton system $M(j,j) > 150$ GeV

The resulting cross section at 13 TeV is 8.4 fb^{-1} which is used to normalized the expected number of events.

we will first demonstrate the usefulness of deep learning networks with this general sample, then discuss the effects of cuts to reject other backgrounds as well as detector modeling effects.

Polarization fractions were obtained on the default sample by fitting the $cos(\theta^*)$ variable determined at truth level. 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. With weights W_i given by

$$W_i = \frac{F_i}{Norm(\theta^*_1, \theta^*_2)}$$

$$Norm(\theta^*_1, \theta^*_2) = [\frac{3}{8}f_L(1 \mp cos(\theta^*_1))^2 + \frac{3}{8}f_R(1 \pm cos(\theta^*_1))^2 + \frac{3}{4}f_0(1 - cos^2(\theta^*)_1))][(cos(\theta^*_1) \to cos(\theta^*_2))] \quad (3)$$

where F_i represents the six possible polarization states for the two W's: Left-Left (LL), Left-Right (LR), Right-Right(RR), Right-Longitudinal (RO), Left-Longitudinal (LO), or Longitudinal-Longitudinal (LO).

$$F_i \in \begin{pmatrix} OO = F_0^2[(1 - cos(\theta^*_1))(1 - cos(\theta^*_2))], \\ LL = F_L^2[(1 \mp cos(\theta^*_1))(1 \mp cos(\theta^*_2))], \\ RR = F_R^2[(1 \pm cos(\theta^*_1))(1 \pm cos(\theta^*_2))], \\ LR = F_L F_R[(1 \mp cos(\theta^*_1))(1 \pm cos(\theta^*_2)) \\ +(1 \mp cos(\theta^*_2))(1 \pm cos(\theta^*_1))], \\ LO = F_L F_O[(1 \mp cos(\theta^*_1))(1 - cos(\theta^*_2)) \\ +(1 \mp cos(\theta^*_2))(1 - cos(\theta^*_1))], \\ RO = F_R F_O[(1 \pm cos(\theta^*_1))(1 - cos(\theta^*_2)) \\ +(1 \pm cos(\theta^*_2))(1 - cos(\theta^*_1))] \end{pmatrix} \quad (4)$$

Where, since no ordering is applied to the W bosons we require that the individual polarization fractions

F_L,F_R,F_0 are equal for both W's. For reweighting the original sample F_L,F_R,F_0 are take as a function of M(W,W). Weights are calculated before any additional event level cuts are made, and the resulting templates are remade for each set of cuts explored.

While there are six possible polarization combinations not all polarization states are as interesting from the the stand point of new physics, and better measurements can be preformed if assumptions can be made. To Illustrate this we define 4 more templates L,R,O,$\phi\phi$. L,R,O fit for the number of Left,Right, and Longitudinal W's as would commonly be done in $cos(\theta^*)$ distributions in events with two W's such as in semi-leptonic $t\bar{t}$ events []. $\phi\phi$ assumes all new physics would be encoded in the longitudinal-longitudinal polarization final state and the only free parameter is the longitudinal-longitudinal fraction.

$$\begin{aligned} TT &= LL + LR + RR \\ OT &= OL + LT \\ \phi\phi &= LL + RR + LR + LO + RO \end{aligned} \tag{5}$$

In addition the templates are checked by requiring that the fitted values for each fraction match the values obtained a truth level. In addition these fractions and kinematics were checked explicitly using Madspin. In order to compare against other sensitive variables we also apply this reweighting to R_{pt}.

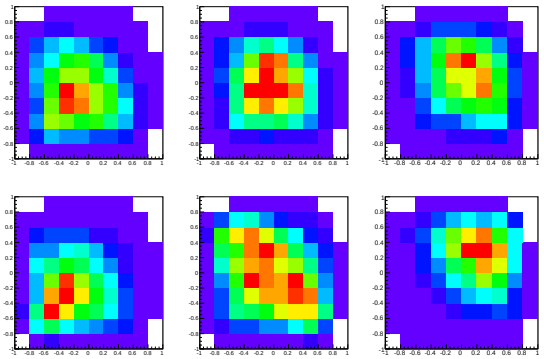

Figure 2: Two dimensional templates after reweighting to pure polarization states which clockwise from the upper left LO,OO,RO,LL,LR,RR. Distance of prediction from the true value of ($cos(\theta^*(1))$,$cos(\theta^*(2))$)

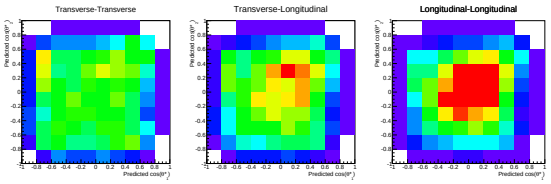

Figure 3: Templates for Neural Network output

4 Predictions for the LHC

Armed with templates for each polarization state, and a sensitive distribution all we have to do is fit the resulting 2D-Data for each polarization fraction. In actually data analysis, this would involve first selecting events to remove background, and then subtracting predicted background from data. The effect of additional cuts and backgrounds is covered below, but for the case of this section we assume that Madgraph selection can be fit directly to obtain sensitivities. The maximum likelihood fit is preformed with the RooFit [] frame work. Fit errors were determined by randomly fluctuating data expectations within their Poisson errors and repeating the fit, and confidence intervals were derived from these toy experiments. The precision for various configurations of

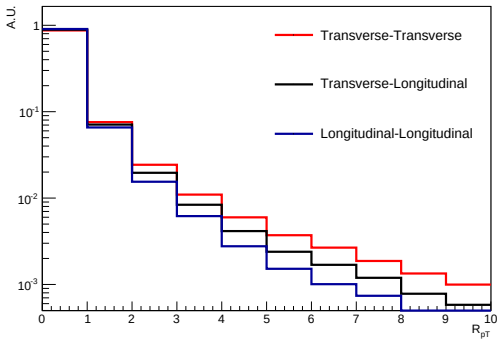

Figure 4: Templates for R_{pT}

Example fit

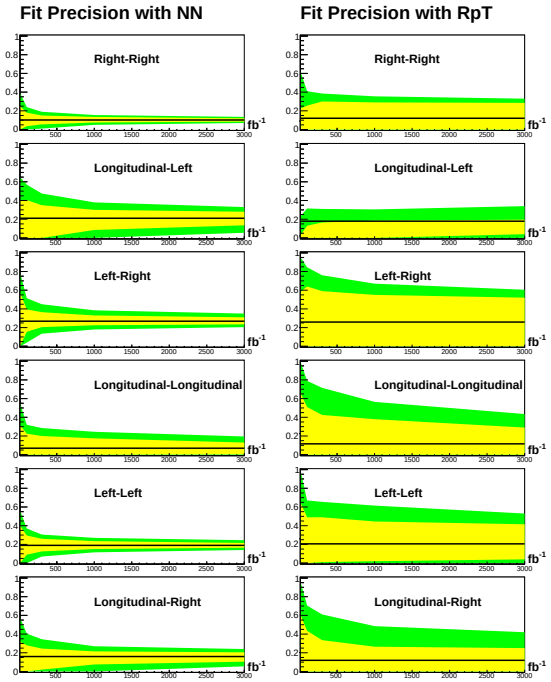

Figure 5: 68% and 95% confidence intervals for fits with six templates. Fits of the neural network on the right and fits of R_{pT} the on the left

templates can be seen in Figures **??**. It can be seen that as expected better precision can be obtained by reducing the number of it templates by fixing groups of polarization to the SM prediction. In **??** that the the Transverse components can be measured we great precision, whereas separating pure longitudinal-longitudinal scattering from longitudinal-transverse scattering is challenging. However, in 7 shows that fixing the the Transverse-Longitudinal fraction to the SM prediction allows for an excellent extraction of the Longitudinal-Longitudinal fraction from the combination of all others. In all cases the neural network output outperforms R_{pT}.

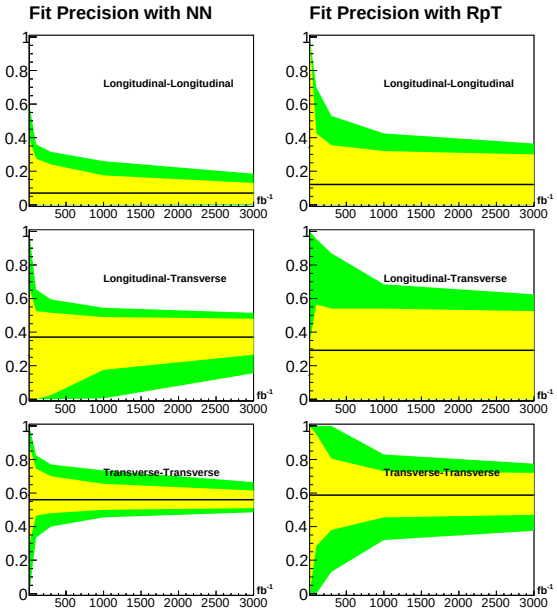

Figure 6: 68% and 95% confidence intervals for fits with three templates. Fits of the neural network on the right and fits of R_{pT} the on the left

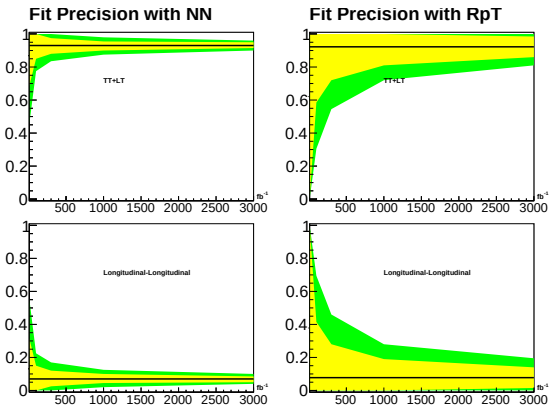

Figure 7: 68% and 95% confidence intervals for fits with two templates. Fits of the neural network on the right and fits of R_{pT} the on the left

5 Detector Studies

While the success of this neural network at simple parton level is encouraging, it is important to check if this procedure will stand up to experimental realities of finite detector resolution, and non $W^{\pm}W^{\pm}jj$ background. To remove background the ATLAS collaboration applied the following cuts to a sample of two same sign leptons with two jets.

- Jet $p_T > 30$ GeV
- Lepton $p_T > 25$ GeV
- Lepton $\not{E}_T > 40$ GeV
- $M(j,j)$ >500 GeV
- $\Delta Y(j,j) < 2.4$

After these cuts the dominate background results WZ production. Table 1 shows the effect of applying these cuts and preforming the same fits at parton level. Additional Table **??**, shows the effect of passing the parton level events through pythia6 for hadronization, and then through Delphes (using the CMS simulation card.) to simulate generic detector smearing. Each additional cut or smearing decreases the sensitivity to the various polarization components as expected, but the possibility to strongly constrain the VBS polarization remain. Figure 8 shows for comparison the shape produced by the dominant background to ATLAS's $W^{\pm}W^{\pm}jj$ study. I can be seen that this background closely resembles the RR template, and it will be important for the WZ to be correctly subtracted before the $cos(\theta^*)$ is fitted for polarization fractions.

6 Conclusions

We've demonstrated that same sign VBS studies can be used to directly study Longitudinal-Longitudinal bosons scattering in spite of the missing information resulting for neutrinos. The use of a deep neural network significantly out preforms the benchmark variable of R_pT. Cuts to reject backgrounds as well as detector smearing reduces the sensitivity as expected,

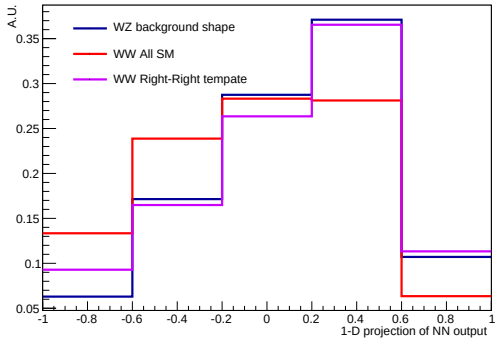

Figure 8: Signal and background shapes

but the method remains a useful tool for the study of VBS.

Templates	!OO		OO	
	LL	UL	LL	UL
Parton	0.92	0.95	0.05	0.09
ATLAS Cuts Parton	0.89	0.95	0.05	0.11
Delphes	0.9	0.95	0.05	0.11
ATLAS Cuts Delphes	0.87	0.95	0.05	0.13

Table 2: Lower Limits (LL) and Upper Limits (UL) for various different simulations.

7 conclusions

References

[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]].

Templates	TT		OT		OO	
	LL	UL	LL	UL	LL	UL
Parton	0.51	0.62	0.27	0.48	0.01	0.13
ATLAS Cuts Parton	0.47	0.63	0.19	0.52	0.0	0.19
Delphes	0.48	0.65	0.16	0.52	0.0	0.21
ATLAS Cuts Delphes	0.44	0.66	0.08	0.56	0.0	0.27

Table 1:) Lower Limits (LL) and Upper Limits (UL) for various different simulations.