A search for tWZ production with the ATLAS detector using the three and four lepton final states in proton-proton collisions at $\sqrt{s} = 13\text{TeV}$

Alexander Veltman, James Keaveney

Department of Physics, University of Cape Town, South Africa, 7701

E-mail: `vltale001@myuct.ac.za`

Abstract. The production of a single top quark with an associated W and Z boson (tWZ) is a rare Standard Model process which has never before been detected. This process is sensitive to the top quark electroweak coupling found in some Beyond Standard Model theories such as Standard Model Effective Field theory and may hold information for constraining these theories. A previous search has been performed for tWZ production using 139 fb^{-1} of proton-proton collision data at a centre of mass energy of 13 TeV recorded at the ATLAS detector. The search was performed across the tetralepton and trilepton final states and have been combined to further increase the sensitivity of the analysis. This analysis was expanded to include a comprehensive set of systematic uncertainties. The work presented will include new preliminary blinded results for the cross section of tWZ production.

1. Introduction

The production of a single top quark with an associated W and Z boson (tWZ) is a rare Standard Model process which has never before been measured. Due to the electroweak vertices involved in the radiation of the W and Z boson, the rate of production of the tWZ process is dependent on how the top quark couples to the electroweak force. The top-electroweak coupling has been identified [1] as a possible area for constraining new beyond Standard Model theories. A precise measurement of the cross-section of the tWZ process will provide more information regarding the coupling.

However, the measurement of the cross section of the tWZ process is difficult. The required experimental signatures used to measure tWZ production are particularly rare due to the prevalence of electroweak interactions. The rarity of the production causes a limiting of available event statistics even when considering a large dataset such as the ATLAS Run 2 dataset. The more damning factor in the measurement is the similarities between the diagrams of tWZ and $t\bar{t}Z$ production. When the next-to-leading order diagrams are considered, the tWZ diagram has major overlaps with that of $t\bar{t}Z$ in which the diagrams differ by only a single lepton originating from a top quark. This issue results in $t\bar{t}Z$ being a major background in the analysis as well as requiring additional steps to perform diagram removal in order to differentiate these processes.

This analysis of the tWZ process uses two possible experimental signatures which are the trilepton and tetralepton channels. Both channels require that the Z boson in the event decays leptonically to pairs of electrons or muons. These experimental signatures have previously been analysed by Masters students at the University of Cape Town. The trilepton channel was investigated in a Masters thesis by Benjamin Warren [2] which requires one of the W bosons (either the radiated W or the W produced by the top decay) to decay hadronically with the other W decaying leptonically. The tetralepton channel was investigated in a Masters thesis by Jake Reich [3] where both W bosons decay leptonically. These channels are joined in a combined fit in order to produce a combined result for the tWZ cross section. The analysis in the tetralepton channel has been extended to account for a wider collection of systematic uncertainties as well as updated simulated samples. The details of the tetralepton analysis and its results will be presented in this proceedings. The analysis of the trilepton channel is in the process of a similar extension but full results are not available at this time.

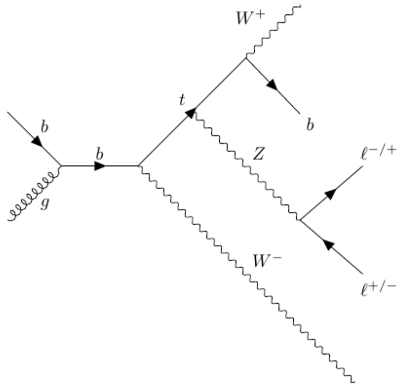

Figure 1. An example of a leading order Feynman diagram of tWZ production where the Z boson decays leptonically.

2. Datasets and Event Selection

This analysis utilises the full $139\,\mathrm{fb}^{-1}$ Run 2 proton-proton collision dataset taken using the ATLAS detector at the Large Hadron Collider between the years of 2015 and 2018. A comparable set of simulated samples were generated for the tWZ process as well as any relevant background processes in order to compare the ATLAS data to Standard Model predictions.

A summary of the baseline requirements for lepton and jet objects in each event is present in Table 1. Any leptons which meet these requirements are considered loose leptons. A loose lepton which passes an extra set of isolation and identification criteria is considered to be a tight lepton. A fake lepton is an object which has incorrectly been misidentified as the expected lepton or a lepton from a process which is not the process of interest. The classification of jets containing b-hadrons, known as b-jets, is done using a machine learning algorithm DLr1 [4] which uses various working points to describe the likelihood for a jet to contain a b-hadron. In the analysis, a jet is defined as tight if it passes the 77% DLr1 working point while it is defined as loose if it passes the 85% DLr1 working point but not the 77% DLr1 working point. A Z boson candidate is defined as any pair of oppositely signed same flavoured (OSSF) leptons in an event which have a reconstructed mass which is within $30\,\mathrm{GeV}$ of the Z boson mass.

The definitions of the signal regions (SR) and control regions (CR) for the analysis are shown in Table 1. Signal regions are constructed to maximise the presence of the signal process with respect to the background processes. The control regions are constructed to provide estimations for the contributions of the various backgrounds so as to control for them when performing the fitting procedure. The analysis uses two signal regions, tWZ OF SR and tWZ SF SR, which differ only by a requirement on the two leptons which did not originate from a Z boson. The region in which the two non-Z leptons are the same flavour, tWZ SF SR, is designed to try isolate contributions from the ZZ background process in the signal regions. Three control regions, $t\bar{t}Z$ CR, ZZb CR and $(tWZ)_{\text{fake}}$, are used in the analysis. The $t\bar{t}Z$ CR and ZZb CR regions are constructed to control for the background contributions from the $t\bar{t}Z$ and ZZ processes respectively. The $(tWZ)_{\text{fake}}$ region requires 3 tight leptons and an additional loose lepton where the region is designed to estimate the contribution of $t\bar{t}Z$ events which contain a fake lepton.

<table>
<tr><th colspan="5">Baseline selections</th></tr>
<tr><td colspan="5">$\text{N}_\ell = 4$
$p_T(\ell_1, \ell_2, \ell_3, \ell_4) > (28, 18, 10, 10)$ GeV
$p_T(\text{jet}) > 20\,\text{GeV}$, $|\eta(\text{jet})| < 4.5$, jvt > 0.5
$|\eta(\ell_e)| < 2.47$ excluding $1.37 < |\eta(\ell_e)| < 1.52$
$|\eta(\ell_\mu)| < 2.5$
$\sum_{i=1}^{4} \text{charge}(\ell_i) = 0$
All OSSF lepton pairs require $m_{\text{OSSF}} > 10\,\text{GeV}$</td></tr>
<tr><th colspan="5">Regions</th></tr>
<tr><th>tWZ OF SR</th><th>tWZ SF SR</th><th>$t\bar{t}Z$ CR</th><th>ZZb CR</th><th>$(tWZ)_{\text{fake}}$ CR</th></tr>
<tr><td>$\text{N}_\ell(\text{tight}) = 4$</td><td>$\text{N}_\ell(\text{tight}) = 4$</td><td>$\text{N}_\ell(\text{tight}) = 4$</td><td>$\text{N}_\ell(\text{tight}) = 4$</td><td>$\text{N}_\ell(\text{tight}) = 3$
$\text{N}_\ell(\text{loose and NOT tight}) = 1$</td></tr>
<tr><td>$\text{N}_{\text{Z candidate}} = 1$</td><td>$\text{N}_{\text{Z candidate}} = 1$</td><td>$\text{N}_{\text{Z candidate}} = 1$</td><td>$\text{N}_{\text{Z candidate}} = 2$</td><td>$\text{N}_{\text{Z candidate}} = 1$</td></tr>
<tr><td>$\text{N}_{\text{jet}} \geq 1$</td><td>$\text{N}_{\text{jet}} \geq 1$</td><td>$\text{N}_{\text{jet}} \geq 2$</td><td>$\text{N}_{\text{jet}} \geq 1$</td><td>$\text{N}_{\text{jet}} \geq 1$</td></tr>
<tr><td>$\text{N}_{\text{b-jet}}(\text{tight}) = 1$</td><td>$\text{N}_{\text{b-jet}}(\text{tight}) = 1$</td><td>$\text{N}_{\text{b-jet}}(\text{tight}) \geq 1$
$\text{N}_{\text{b-jet}}(\text{loose}) \geq 0$
$\text{N}_{\text{b-jet}}(\text{tight}) + \text{N}_{\text{b-jet}}(\text{loose}) = 2$</td><td>$\text{N}_{\text{b-jet}}(\text{tight}) = 1$</td><td>$\text{N}_{\text{b-jet}}(\text{tight}) = 1$</td></tr>
<tr><td>Opp. Flavour Non-Z leptons</td><td>Same Flavour Non-Z leptons</td><td>-</td><td>-</td><td>-</td></tr>
</table>

Table 1. A summary of the baseline selections for objects definitions and the requirements for the signal and control regions.

3. Background Discrimination

In order to differentiate the signal and background within the signal regions, a discriminating variable is determined with a focus on separating tWZ events from major contributing background processes, particularly $t\bar{t}Z$ events. This variable is calculated per event and is produced by using the output of a scanning algorithm, the Two Neutrino Scanning Method ($2\nu SM$), as an input into a broader machine learning model, a boosted decision tree (BDT), along with other event properties.

3.1. Two Neutrino Scanning Method

The Two Neutrino Scanning Method ($2\nu SM$) algorithm is constructed to produce a weight between 0 and 1 which indicates the likelihood that an event contains a $t\bar{t}$ pair. The $2\nu SM$ score helps discriminate between the signal tWZ events and background $t\bar{t}Z$ since only the $t\bar{t}Z$ events will have a $t\bar{t}$ system. The algorithm tries to reconstruct the $t\bar{t}$ system by considering different neutrino kinematics.

The algorithm relies on knowing probability distributions for kinematic observables of the top quarks and their neutrino decay products. A sample of simulated $t\bar{t}$ events is used to construct these distributions using generator-level information where the following variables are calculated:

$$\begin{aligned} m_{t_1}^2 &= \ell_1^2 + b_1^2 + \nu_1^2 & \qquad \Delta E_x &= (p_{T,\nu_1})_x + (p_{T,\nu_2})_x - (E_T^{\text{miss}})_x \\ m_{t_2}^2 &= \ell_2^2 + b_2^2 + \nu_2^2 & \qquad \Delta E_y &= (p_{T,\nu_1})_y + (p_{T,\nu_2})_y - (E_T^{\text{miss}})_y \end{aligned} \tag{1}$$

where ℓ_i are the 2 non-Z leptons and the b_i are the 2 highest weighted jets according to the DL1r algorithm. The product of these distributions are then used to define the $2\nu SM$ weight:

$$\omega_{2\nu SM} = Pr[m_{t_1}]Pr[m_{t_2}]Pr[\Delta E_x]Pr[\Delta E_y] \tag{2}$$

where $Pr[X]$ is the probability distribution function of X which results in a value between 0 and 1. A higher score indicates a more suitably reconstructed $t\bar{t}$ system. In order to calculate the maximum $\omega_{2\nu SM}$ for some event, different possible configurations of neutrino pairs are considered. Different values of angular components of the neutrino 4-vectors are scanned over with the m_{t_1}, m_{t_2}, ΔE_x and ΔE_y being calculated for each configuration. These kinematic variables are then evaluated on the generated distributions and the $\omega_{2\nu SM}$ weight is found using equation 2. Only the maximum $\omega_{2\nu SM}$ weight is used to label the likelihood of a $t\bar{t}$ system being in an event.

3.2. Event-Level BDT

A binary classification boosted decision tree (BDT) was created to provide the required discriminating variable. The input features of the model include the scalar sum of the transverse momenta of the leptons, jets and b-jets, the angular differences between the two non-Z leptons and the maximum $\omega_{2\nu SM}$ score of the event. Since the objective of the model is binary classification, the output of the model is a number between 0 and 1 where 0 is labelled a background event and 1 is labelled a signal event. The model was trained using simulated samples of tWZ, $t\bar{t}Z$ and ZZ where tWZ is labelled as examples of signal events and $t\bar{t}Z$ and ZZ are labelled as examples of background events. An investigation of the feature importance of the model found that the maximum $\omega_{2\nu SM}$ score had the highest importance in the model.

4. Extraction Method

The rate of tWZ production is determined by the signal strength $\mu(tWZ)$ which is defined as the ratio between the measured tWZ cross section and the Standard Model prediction for the tWZ cross section. A signal strength of $\mu(tWZ) = 0$ represents no detection of any tWZ production while $\mu(tWZ) = 1$ represents the Standard Model prediction for tWZ production. The signal strength is determined using a binned profile likelihood fit [5]. The signal regions and the $t\bar{t}Z$ control region are treated as histograms of the BDT discriminating variable while the ZZb control region uses the scalar sum of all jet and lepton p_T and E_T^{miss}. The $(tWZ)_{\text{fake}}$ region uses the p_T of the loose lepton in the event. During the fitting procedure, $\mu(tWZ)$ is considered a parameter of interest while the weights of each histogram bin and each systematic uncertainty are treated as nuisance parameters and are allowed to vary.

5. Extracting signal strength in tetralepton channel

The analysis was first conducted using a fully blinded Asimov dataset which assumes that the data within all regions appears exactly like the simulated samples. This toy dataset is useful for making comparisons between different analysis strategies without exposing the analysis to possible biases introduced by using the ATLAS dataset. The fully blinded fit result for the tWZ signal strength $\mu^{\text{stat}}_{Asimov}(tWZ)$ which allowed only statistical nuisance parameters to vary was 1.00 ± 0.93. When the systematic uncertainties were included, the fit result for $\mu_{Asimov}(tWZ)$ was $1.00^{+1.25}_{-1.02}$. The comparison between the fit with only statistical uncertainties and the fit with both statistical and systematic uncertainties shows that the analysis is mainly dominated by statistical uncertainties. This is unsurprising due to the limited number of tWZ events available in the tetralepton channel. However, the introduction of a wider set of systematic uncertainties related to new calibrations increased the contributions of systematic uncertainties more than expected. This may indicate that some precision may be gained by attempting to mitigate these systematic contributions.

A fit was also performed using a mixed dataset where blinded Asimov data is only used in the signal regions while the actual ATLAS data is used in the control regions. The fit result for the tWZ signal strength $\mu(tWZ)$ using a mixed dataset was $2.00^{+1.61}_{-1.20}$ which agrees with the Standard Model prediction. This result has an associated expected significance of $Z_{\mu}^{exp} = 0.95\sigma$ with respect to the Standard Model prediction. The measurement shows no significant deviation from the Standard Model prediction.

The major contributing systematic uncertainties can be seen in the ranking plot in Figure 2. When considering each of these systematic uncertainties, the uncertainty with the largest impact on the fit result is that associated with the $t\bar{t}Z$ cross section. This is expected since $t\bar{t}Z$ is the largest background process in the analysis and the ability to constrain its contribution should greatly affect the fit result. Other notable systematic variables are those associated with the tracking and reconstruction of jets and parton distribution functions (PDF) calculations. These are relevant due to the number of jets present in the various regions.

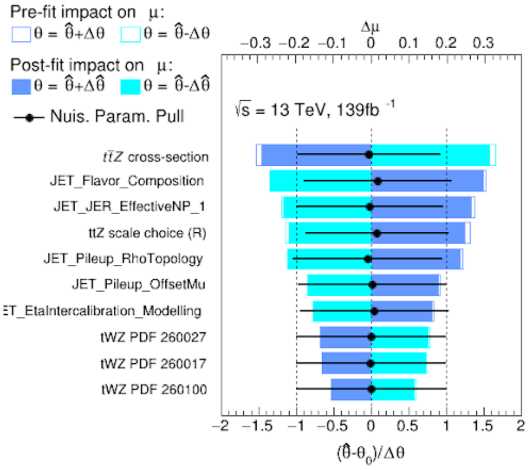

Figure 2. Ranking plot for the variations of the systematic uncertainties involved in the mixed data profile likelihood fit. The black points (bottom axis) represent the pull associated with each nuisance parameter. The impact of each nuisance parameter associated with the systematic uncertainties (top axis) is shown using the blue bars. Only the nuisance parameters with the ten highest impact values are shown.

6. Conclusion

The single top quark production with an associated W and Z boson is a rare Standard Model process which has not been detected. A search for tWZ production was performed which investigated the tetralepton and trilepton final states where only the blinded results for the tetralepton channel were presented. The best fit value for the signal strength of tWZ production in the tetralepton channel with respect to the Standard Model prediction was $\mu(tWZ) = 2.00^{+1.61}_{-1.20}$ where Asimov data was used in the signal regions of the analysis. The result had an associated expected significance of $Z_{\mu}^{exp} = 0.95\sigma$ with respect to the Standard Model prediction. No deviation from the Standard Model prediction was found. This result will be extended through combination with the measurement in the tWZ trilepton channel.

References

[1] Maltoni F, Mantani L and Mimasu K 2019 *Journal of High Energy Physics* **2019** URL `https://doi.org/10.1007/jhep10/82019/9004`

[2] Warren B 2021 A search for twz production in the trilepton channel using run 2 data from the atlas experiment URL `http://hdl.handle.net/11427/36228`

[3] Reich J 2021 A search for twz production in the full run 2 atlas dataset using events with four leptons URL `http://hdl.handle.net/11427/36050`

[4] Aad G *et al.* 2019 *The European Physical Journal C* **79** URL `https://doi.org/10.1140/epjc/s10052-019-7450-8`

[5] Cowan G, Cranmer K, Gross E and Vitells O 2011 *The European Physical Journal C* **71** URL `https://doi.org/10.1140/epjc/s10052-011-1554-0`