



DEPARTMENT OF PHYSICS
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Search for tWZ production in the Full Run 2 ATLAS
dataset using events with four leptons

Jake Reich
Student Number: RCHJAK001
Supervisor: Dr. James Keaveney
Co-Supervisor: Dr. Sahal Yacoob

October 2021

Abstract

Declaration

I certify that this assignment/report is my own work, based on my personal study and/or research and that I have acknowledged all material and sources used in its preparation, whether they be books, articles, reports, lecture notes, and any other kind of document, electronic or personal communication. I also certify that this assignment/report has not previously been submitted for assessment in any other unit, except where specific permission has been granted from all unit coordinators involved, or at any other time in this unit, and that I have not copied in part or whole or otherwise plagiarised the work of other students and/or persons.

Acknowledgements

Contents

1	Introduction	7
2	Theory	8
2.1	Standard Model of Particle Physics	8
2.1.1	Electroweak Theory	8
2.1.2	Top Quark	8
2.2	tWZ	8
2.2.1	Tetra-lepton Channel	8
2.2.2	Comparison to Tri-lepton Channel	9
2.3	Effective Field Theory (EFT)	10
2.4	Machine Learning in the Context of Particle Physics Analyses	10
2.5	Statistical Techniques	10
2.5.1	Maximum Likelihood Fitting	10
2.5.2	p-value and χ^2	10
2.5.3	Significance	11
2.5.4	Limit Setting	11
3	The ATLAS Experiment and Detector	12
3.1	The ATLAS Experiment	12
3.1.1	Large Hadron Collider (LHC)	12
3.2	The ATLAS Detector	12
3.2.1	Coordinate System and Kinematics	12
3.2.2	Tracking Detectors	12
3.2.3	Calorimeter System	12
3.2.4	Muon Spectrometer	12
3.2.5	Trigger and Data Acquisition System	12
3.2.6	Particle Identification	12
4	Analysis Setup and Strategy	13
4.1	Data and Monte Carlo Simulation	13
4.1.1	Data Samples	13
4.1.2	Monte Carlo Samples	13
4.2	Objects	13
4.2.1	Leptons	14
4.2.2	Jets	14
4.2.3	b-tagging	14
4.3	Kinematic Pre-selection cuts	14
4.4	Regions and Event Selection	14
4.5	Machine Learning Techniques	14
4.6	Fake Lepton Estimation	14
4.7	Analysis Framework: TRExFitter (TRF)	14

5	Search for tWZ Production	15
5.1	Backgrounds	15
5.1.1	$t\bar{t}Z$	15
5.1.2	ZZ	15
5.1.3	other	15
5.2	Control Plots	15
5.3	Post-Fit Plots	15
5.4	Results	15
6	Conclusion and Outlook	16
A	Appendix	17

Chapter 1

Introduction

Write similar to what is in nrf application.

Talk about previous paper's (tWZ 3-lep) findings - <http://cds.cern.ch/record/2625170>

Explain that SM aims to describe fundamental physics, but fails in certain cases (DM, gravity etc.)

Possibly talk about EFT? Finding tWZ cross section - χ^2 global fit. FOR REFERENCE:

The production of a single top quark in association with a W and Z boson (tWZ) is sensitive to both the neutral and charged electroweak couplings of the top quark as the process involves the simultaneous production of a W boson and a Z boson in association with the top quark. Due to the very large coupling of the top quark to the Higgs boson, the electroweak couplings of the top quark are a theoretically well-motivated area to expect the first signs of new physics. The recent lack of signs of new physics from LHC data tells us that new physics is either very heavy, or is very weakly coupled to Standard Model particles, therefore we might only observe signs of new physics in anomalous rates of well-chosen processes. A prime example of such a process is tWZ . This has an extremely low production cross section (0.7 fb), meaning that it is an extremely rare process to observe and subsequently, it has never been observed by any particle physics experiment. However, the latest datasets recorded by ATLAS are sufficiently large to allow a potential observation of this rare, novel process.

We aim to use the Full Run 2 dataset recorded by the ATLAS experiment at CERN to search for the production of a top quark together with a W and Z boson for the channel with four leptons (two originating from the decay of the Z boson, one from the associated W boson and one from the W boson which decays from the top quark (together with a b quark)). The Standard Model of particle physics has been confirmed to an extraordinary degree of precision, however we know there are stark deficiencies therein. These include its incompatibility with the theory of gravity and an explanation of the matter-antimatter asymmetry in the universe. Especially relevant is the Standard Model's lack of an explanation for the vast differences in the strengths of the fundamental forces (The Hierarchy Problem), constraining the electroweak couplings of the top quark squarely addresses this fundamental scientific question.

Chapter 2

Theory

2.1 Standard Model of Particle Physics

What is SM (renormalisable qft), come from symmetry, brief description of group structure? Explain structure/properties (fermions, bosons, etc.), coupling constants. 'Particles carry color/electric charge, some particles can interact with others/themselves, some can't' Where does it how? How well does it work? Where doesn't it work?

2.1.1 Electroweak Theory

Properties of W and Z. Decay channels. Z as the standard candle (distinct OSSF lepton signal).

2.1.2 Top Quark

Properties. History (when/how was it discovered/theorised). Why interesting (large mass). Hierarchy problem. Decay channels. Extremely short lifetime (makes b's so important for top ID).

2.2 tWZ

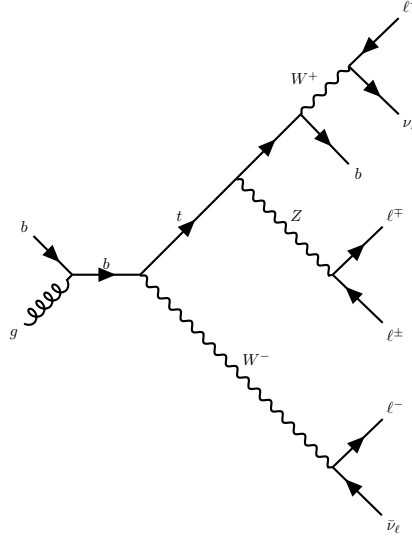
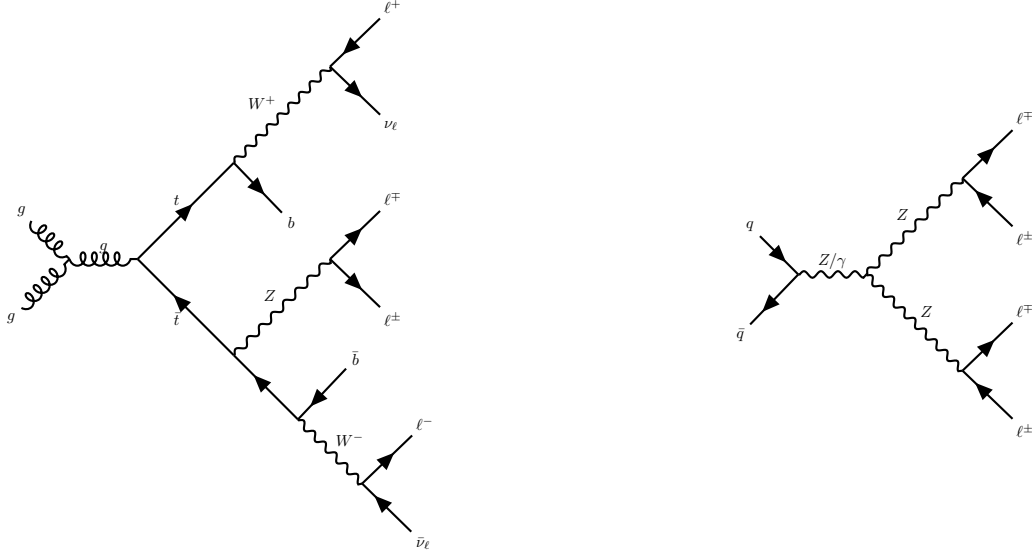
2.2.1 Tetra-lepton Channel

Provide Cross section.

The leading order Feynman diagram for tWZ in the tetra-lepton channel is shown below.

Backgrounds

The main backgrounds for tWZ (tetra-lepton channel) are the production of a two tops, both in the $\ell\nu b$ final state channel, together with a Z boson ($t\bar{t}Z$) and diboson production with fully leptonic final states (ZZ).


 Figure 2.1: Example Feynman diagram of tWZ production in the tetra-lepton channel.

 Figure 2.2: Example Feynman diagrams for $t\bar{t}Z$ (left) and ZZ (right) in the tetra-lepton channel.

2.2.2 Comparison to Tri-lepton Channel

Less backgrounds to deal with (in tetralepton). However lower stats (in tetra). Give cross sections (and feynman diagram). Maybe talk a bit about analysis related challenges (trilepton has a hadronically decaying W, does this make the analysis easier or more difficult?).

The most apparent difference between the tri and tetra-lepton channels is the amount of statistics present, with the tetra-lepton channel having far less events in its phase space than that of the tri-lepton channel. The lack of statistics in the tetra-lepton channel can be attributed to its low production cross section, $\sigma_{(tW\pm Z).Br(4\ell)}^{\text{NLO}} = 0.7 \text{ fb}[\text{twz'3'lep}]$.

The tri-lepton channel has a production cross section ($\sigma_{(tW\pm Z).Br(3\ell)}^{\text{NLO}} = 3.9 \text{ fb}[\text{twz'3'lep}]$) around a factor of 4 larger than that of the tetra-lepton channel. This difference between the production cross section of the two decay channels can be largely attributed to the difference in branching ratios ($\frac{\Gamma_i}{\Gamma}$) between a hadronically decaying W boson, $\frac{\Gamma_{W \rightarrow had}}{\Gamma_W} = (67.41 \pm 0.27)\%[\text{pdg}]$, present in the tri-lepton channel and a leptonically decaying W boson, $\frac{\Gamma_{W \rightarrow \ell\nu}}{\Gamma_W} = (10.86 \pm 0.09)\%[\text{pdg}]$, present in the tetra-lepton channel.

Despite the tetra-lepton channel's low statistics, it is not subject to the large WZ background present in

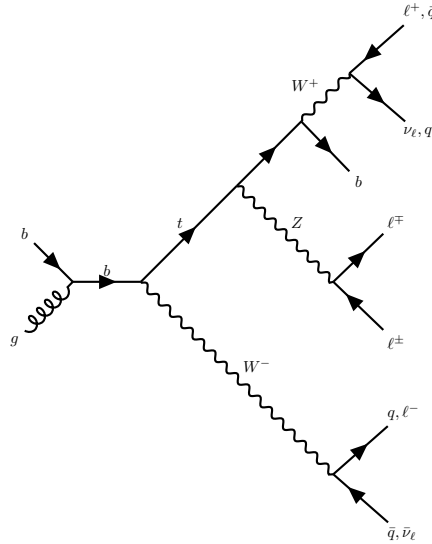


Figure 2.3: Example Feynman diagram of tWZ production in the tri-lepton channel.

the tri-lepton channel.

2.3 Effective Field Theory (EFT)

Brief overview of EFT. What is EFT? Why important to pp as a whole? Why important in twz (high sensitivity to wilson coefficients, expected to have a large impact on global fit)? Similar to what james says in INT note.

2.4 Machine Learning in the Context of Particle Physics Analyses

Brief overview of ML as a whole; History; increase in popularity in recent years (why increased in popularity \rightarrow novel techniques developed, increase in computing power for your buck). Where does it fit into pp (event selection, object reconstruction and ID (jet reco, b-tagging)). Explain concepts (vocabulary), overtraining, training, testing, classifier, classification. Why use x train/test ratio in pp (use some for analysis/use some for training)? Popular tools which are used today (scikit learn, TMVA, xgboost what is theano, what is keras).

Maybe a subsection on bdt (if end up using it) on the specific algorithm, and minimizing cost function (general). Explanation on ROC curve and why we can use it as a proxy to determine how well our bdt/nn is doing (and where it fails/can fail \rightarrow things to be aware of cautious of when straight up using ROC integral (e.g. overtraining)).

2.5 Statistical Techniques

Brief overview: frequentist and bayes approach in general in pp, why we use frequentist in this analysis

2.5.1 Maximum Likelihood Fitting

Go through theory. Varying histograms.

2.5.2 p-value and χ^2

Go through theory. Go through story of getting p-value, what it means, getting chi-squared, what it means, where the chi squared distribution comes in, degrees of freedom, what is a 'good' chi squared value and what is 'bad' and what do different values mean/infer/suggest.

2.5.3 Significance

Go through theory. What is means/interpreted as. 3 sigma observation, 5 sigma discovery.

2.5.4 Limit Setting

Go through theory.

Chapter 3

The ATLAS Experiment and Detector

3.1 The ATLAS Experiment

Brief overview of history of ATLAS and CERN in general. State accomplishments (higgs discovery - completed list of SM particles).

3.1.1 Large Hadron Collider (LHC)

Topology, parts of it.

3.2 The ATLAS Detector

3.2.1 Coordinate System and Kinematics

Physical dimensions, properties (mass). Coordinate system. Definition of ΔR , η .

In the next subsections:

Where are different particles detected? How are they detected at each part of the detector? How well do the different parts detect the different particles and how do they take advantage of different particle properties each part wishes to detect in an engineering/physics perspective (e.g. why we WANT some particles to get detected at some parts of the detector and how we do this (by use of correct materials which interact specifically with those particles which we wish to measure and NOT with other particles))

3.2.2 Tracking Detectors

3.2.3 Calorimeter System

Electromagnetic Calorimeter

Hadronic Calorimeter

3.2.4 Muon Spectrometer

3.2.5 Trigger and Data Acquisition System

3.2.6 Particle Identification

Chapter 4

Analysis Setup and Strategy

4.1 Data and Monte Carlo Simulation

4.1.1 Data Samples

Brief overview of Full Run 2 and where the data comes from and what time period.

Luminosity of full run 2.

4.1.2 Monte Carlo Samples

Signal and Background samples. Choice of cuts at ntuple level. Specifically why each background is chosen (and why others excluded) \rightarrow e.g. talk about branching fractions, cross sections, topology. How are these backgrounds passing our event selection (e.g. $t\bar{t}z \rightarrow b$ can be lost/untagged/mis-id'ed) \rightarrow provide an explanation for each background.

Details of each sample (event generator, parton shower).

The following background processes are considered:

- **$t\bar{t}Z$** : $t\bar{t}$ with an associated Z -boson, in the tetralepton final state. Therefore, both top-quarks decay leptonically (e.g. $t \rightarrow W^+b \rightarrow \ell^+\nu b$) and of these top-quarks emits a Z -boson which decays leptonically ($Z \rightarrow \ell^\pm \ell^\mp$ (OSSF lepton pair)). This results in a final state with 4 leptons and 2 b-quarks.
- **ZZ** : Diboson production with a tetralepton final state, therefore both Z -bosons decay leptonically ($Z \rightarrow \ell^\pm \ell^\mp$ (OSSF lepton pair)).
- **other**: Processes with a relatively minimal, but non-negligible background contribution in the tWZ signal region

- **Triboson**:

$$WZZ \rightarrow llll\nu$$

$$ZZZ \rightarrow llllll$$

$$ZZZ \rightarrow llll\nu\nu$$

$$WWZ \rightarrow llll\nu\nu$$

- **tZq** : Top quark in association with a Z -boson and another quark.

4.2 Objects

In the subsections below:

Explain why we applied each cut/selection.

4.2.1 Leptons

Tight/loose/med definitions, efficiency of electron and muons specifically at ATLAS. Why we don't consider taus.

4.2.2 Jets

What algorithm did we use and why

4.2.3 b-tagging

What algorithm/WP did we use and why

4.3 Kinematic Pre-selection cuts

Mass windows on Z (OSSF), sum charge = 0, explanations on all other non object level cuts/selections, OSSF < 10 GeV cut

4.4 Regions and Event Selection

Summary table of event selection. Why chose ZZb and ttz region.

4.5 Machine Learning Techniques

What tool did we use, how did we use it, parameters of bdt/nn, input variables/importance, conversion of event level bdt output (bdtscore) to variable for fitting. Used ROC curve integral as a proxy for how good bdt was doing.

4.6 Fake Lepton Estimation

Expected to be a small effect, why? Brief, general explanation/idea of methods used (full explanation/description of what we did and the results/plots/etc. later)

4.7 Analysis Framework: TRExFitter (TRF)

What is TRExFitter? What can it do? At which stage(s) in the analysis did we use it? Which version did we use? Binning method. Explain calculation of error bars in TRF.

Include general flow chart of analysis (not sure where)

Chapter 5

Search for tWZ Production

5.1 Backgrounds

Maybe remove subsections below...

Yields table, signal percentage, background percentage in each region (and each sample)

5.1.1 $t\bar{t}Z$

5.1.2 ZZ

5.1.3 other

5.2 Control Plots

Fit variables only. Why these variables were chosen (maybe provide separation plots or just the separation values of different variables). Comment on data/mc agreement.

5.3 Post-Fit Plots

Plots, yields.

Input into fitting procedure (systematics, normfactors, nuisance parameters , why these were chosen) → does this go in a previous section?

What did fit do? Explain results from push-pull (e.g. the fit constrained (had a constraining effect) x background uncertainty)

5.4 Results

Commentary on results, limit on $\sigma(tWZ)$, within uncertainties of SM?

Chapter 6

Conclusion and Outlook

Summary of study. Possible ways to improve/extend analysis. Better understand ttz background? Is there anything we can say about Z, t, b, W, whatever that we can take from the study (e.g. x is difficult to detect/discriminate in such analyses due to a and b, however it can be improved by doing y). Maybe talk about what preliminary studies showed, but couldn't fully implement that idea due to whatever restriction/limitation.

Appendix A

Appendix