



Using $t\bar{t}$ and missing energy to search for dark matter

Martha Elliott, University of Manchester, United Kingdom

September 7, 2017

Abstract

The search for decays of $t\bar{t}$ into invisible particles was performed by analysing the signals from $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ decays and by selecting an optimised signal region in which to do so (SRC). The Monte Carlo simulation used was validated by comparison to real data at preselection level. $t\bar{t}Z(\rightarrow l\bar{l})$ was also studied in SRC with the goal of computing a ratio of $\sigma(t\bar{t}Z(\rightarrow \nu\bar{\nu}))/\sigma(t\bar{t}Z(\rightarrow l\bar{l}))$ to be compared to theoretical data to indicate whether invisible particles are being produced. The predicted number of $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ signals was 7.4 with an uncertainty of 92%. This method of searching for new phenomena was judged to be feasible but only with significant improvements to sample purity.

Contents

1	Introduction	3
2	Theory	4
2.1	Production of dark matter	4
2.2	Using $t\bar{t}(Z \rightarrow \nu\bar{\nu})$ and $t\bar{t}(Z \rightarrow l\bar{l})$ to find corrections to the Standard Model	4
3	Variables and decay channels of interest	5
4	Experimental procedure	6
4.1	Creating histograms for analysis	6
4.2	Calculating purity and number of events	6
4.3	Determining the uncertainty	6
5	Investigating $t\bar{t}Z(\rightarrow \nu\nu)$ and background decays	6
5.1	Preselection	6
5.2	DM high and DM low	8
5.3	$t\bar{t}Z(\rightarrow \nu\bar{\nu})$ signal region	8
6	Investigating $t\bar{t}Z(\rightarrow l\bar{l})$ and background decays	11
6.1	Preselection	11
6.2	$t\bar{t}Z(\rightarrow \nu\bar{\nu})$ signal region	11
7	The future	11
8	Conclusion	12

1 Introduction

Though many theories attempt to predict the processes by which dark matter is created, experiment so far has failed to make a successful measurement. It is for this reason that it is useful to assess the feasibility of performing a general search through the analysis of specific decays. As the heaviest Standard Model particle, the top quark is useful to study when searching for new particles with an unknown mass. This is because it couples most strongly to scalar particles such as the Higgs boson, and is thus expected to also couple most strongly to the boson mediating the production of dark matter.

The method used to find new particles was, in part, inspired by previous work conducted within ATLAS[3] to infer the occurrence of dark matter production by comparing the measured cross section of a particular process producing missing transverse energy to the known branching ratio of the process. The errors that arise from this method are significantly reduced through the normalisation of the cross section and branching ratio with that of a visible process. Figure 1 shows the predictions of the cross section ratio, $\sigma(jets + Z(\rightarrow \nu\bar{\nu}))/\sigma(jets + Z(\rightarrow l\bar{l}))$, labelled as R^{miss} , by different theories and the final comparison to data. In this case the data agrees with the Standard Model, and this provided the motivation to perform this analysis with alternative decays.

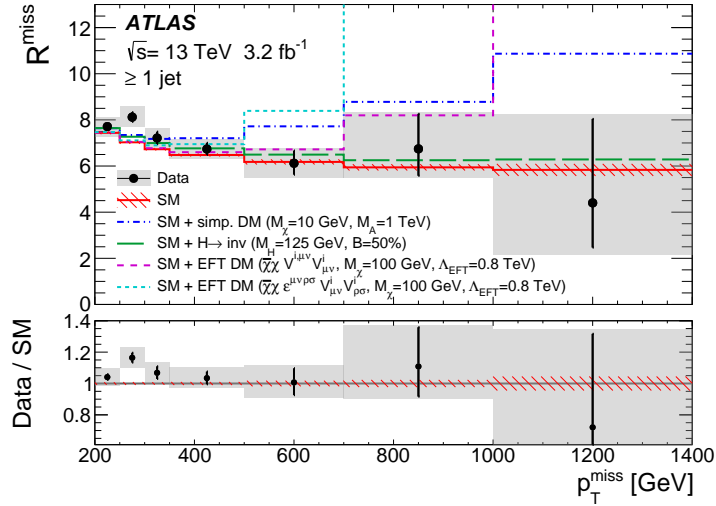


Figure 1: Plots showing the comparison of data to prediction of theoretical models. R_{miss} is the ratio of cross sections.

This project attempted to analyse the $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ decay process to perform a general search for invisible decays and assess the feasibility of doing this using data from the Large Hadron Collider (LHC).

2 Theory

2.1 Production of dark matter

The large mass of the top quark means it has the potential to decay into a variety of different particles. It also means it has a strong coupling to the Higgs boson and, potentially, other new scalar particles such as the mediator, m_ϕ , of dark matter fermions, χ . It is for this reason that the decay as shown in figure 2 is hypothesised.[2]

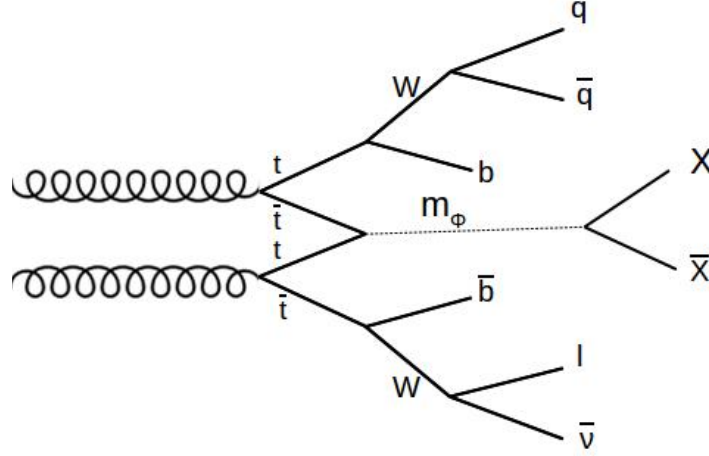


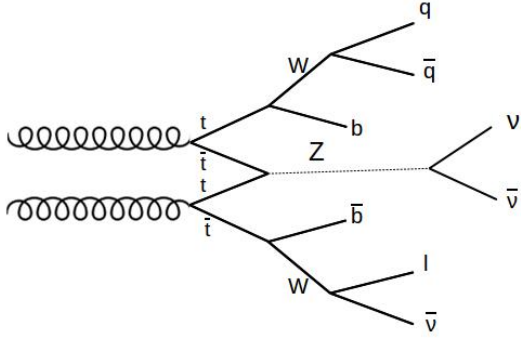
Figure 2: Feynman diagram displaying a process leading to the production of dark matter particles. q is any quark and l is an electron or muon and the neutrino ν is of the corresponding flavour. χ and $\bar{\chi}$ are a pair of dark matter fermions and m_ϕ is a scalar or vector boson.[3]

2.2 Using $t\bar{t}(Z \rightarrow \nu\bar{\nu})$ and $t\bar{t}(Z \rightarrow l\bar{l})$ to find corrections to the Standard Model

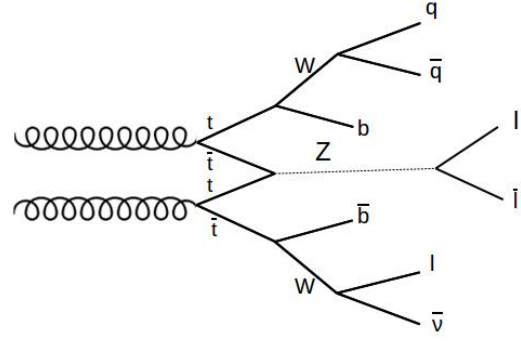
The lack of interactivity of dark matter particles means that any production of these particles in detectors is likely to be assumed to be the production of neutrinos, since the signatures of both of these processes are almost identical in the ATLAS detector. This is because both dark matter and neutrinos have low interactivity so their production is signified by missing transverse energy, missing transverse momentum and the lack of tracks in calorimeters. Due to this, decays into neutrinos are a good area in which to search for dark matter production. Figure 3a shows the $t\bar{t}(Z \rightarrow \nu\bar{\nu})$ process, which can be directly compared to figure 2 to demonstrate the clear correspondence between signatures.

The ratio of branching ratios for $t\bar{t}(Z \rightarrow \nu\bar{\nu})$ and $t\bar{t}(Z \rightarrow l\bar{l})$ is known to be approximately 6, which allows the Standard Model to be tested experimentally. The signature of $t\bar{t}(Z \rightarrow l\bar{l})$ is tracks in the tracker and the muon system in the case that l is a muon, and tracks in the tracker and the electromagnetic calorimeter in the case that l is an electron. By making measurements of the cross sections, σ ,

$$\frac{\sigma(t\bar{t}Z(\rightarrow \nu\bar{\nu}))}{\sigma(t\bar{t}Z(\rightarrow l\bar{l}))} \quad (1)$$



(a) The production of a $\nu\bar{\nu}$ pair mediated by a Z boson.



(b) The production of a $l\bar{l}$ pair mediated by a Z boson.

Figure 3: Feynman diagrams displaying decay processes of interest for $t\bar{t}$. q is any quark, l is an electron or muon and ν is a neutrino.

can be calculated. If the value of equation 1 is ~ 6 , this suggests that decay to neutrinos makes up the invisible portion of decays. If it does not, this could indicate that decay to neutrinos is not the only invisible process occurring, thereby proving the Standard Model incomplete. Studying the ratio reduces the errors because the uncertainty for each process is correlated. It was in order to study this ratio that the $t\bar{t}Z(\rightarrow l\bar{l})$ process, as demonstrated in figure 3b, was analysed.

3 Variables and decay channels of interest

Before attempting to identify the production of dark matter from a Z boson, the appropriate background processes had to be identified and analysed. The main background process are single top, $t\bar{t}$, W +jets, Z +jets, diboson, $t\bar{t} + Z(\rightarrow \nu\bar{\nu})$ and $t\bar{t} \rightarrow V$, where V is a W or Z boson. For each of these processes the variables of interest were the missing transverse energy (E_T^{miss}), the transverse momentum of each jet (p_T), the number of bottom jets (N_b), the transverse mass of the event (m_T), a generalisation of transverse mass applied to dileptonic events where one lepton is not reconstructed (m_{T2}) and the angular separation between the b jet and leptons ($\Delta R(b, l)$). $H_{T, \text{sig}}^{\text{miss}}$, as defined by

$$H_{T, \text{sig}}^{\text{miss}} = \frac{|\vec{H}_T^{\text{miss}}| - M}{\sigma_{|\vec{H}_T^{\text{miss}}|}}, \quad (2)$$

and the angular variable $\Delta\phi p_T^{\text{miss}}$ for the lepton with jets one and two, as defined by

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \quad (3)$$

where \vec{H}_T^{miss} is the negative vectorial sum of the momenta from the leptons and the jets, M is the characteristic scale of the background processes, $\sigma_{|\vec{H}_T^{\text{miss}}|}$ is calculated using jet energy uncertainties and $\Delta\eta$ is the pseudorapidity, were also of interest.

The background processes for the $t\bar{t}Z(\rightarrow l\bar{l})$ decay were single top, $t\bar{t}$, diboson, tZ and Z +jets. The same variables were studied as for $t\bar{t}Z(\rightarrow l\bar{l})$ with the addition of the dileptonic mass and the dileptonic p_T .

4 Experimental procedure

4.1 Creating histograms for analysis

PyROOT was used to loop over a chosen number of events in the sample, identify the ones of interest and add them to a histogram. This produced a separate histogram for each variable and for each background decay. In order for comparisons to be drawn easily between these diagrams, the histograms were stacked and overlaid. This was done using a code that looped round each histogram and, for the input option 'overlay', plotted one on top of the other whereas, for the input option 'stack', the histograms were summed and plotted above each other. Another input option 'data' was added which stacked the the histograms and then superimposed a real data line on top. For the latter two options, a ROOT package, 'THStack', was utilised. Overlaid plots provide a useful method of comparing the shapes and distributions of the variables of interest for each background decay. To compare shapes more clearly the histograms were normalised before being overlaid. In contrast, stacked plots display the proportions of each background decay to the overall signal.

4.2 Calculating purity and number of events

Another code was designed to loop through the histograms created for each background decay for each variable and compute the integral of each, the result of this being equal to the weighted number of events passing the selections. Variations in bin size caused small fluctuations between results from each variable but the numbers and proportions were generally in agreement.

4.3 Determining the uncertainty

Writing the number of signal events as the difference between total events and the amount of background, allowed the errors to be added in quadrature. The systematic uncertainty was assumed to be approximately 10% and the error on the total number of events was assumed to be roughly equal to its square root. These assumptions allowed the fractional uncertainty to be calculated as being

$$\frac{\sigma_{t\bar{t}Z(\rightarrow \nu\bar{\nu})}}{N_{t\bar{t}Z(\rightarrow \nu\bar{\nu})}} = \frac{\sqrt{N_{total} + (0.1N_{background})^2}}{N_{total} - N_{background}}. \quad (4)$$

5 Investigating $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ and background decays

5.1 Preselection

In order to isolate the $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ process, some appropriate cuts were applied to the data samples. This was done by restricting which events would be added to the histograms based on their properties. The first cuts applied ensured the total number of jets was at least four, the number of b jets was at least one and that exactly one lepton was produced in each process. This combination ensured that one of the W bosons would decay leptonically and the other would decay hadronically. Asserting this combination causes the signal to be clearer, as too many jets cause confusion and increase the uncertainties.

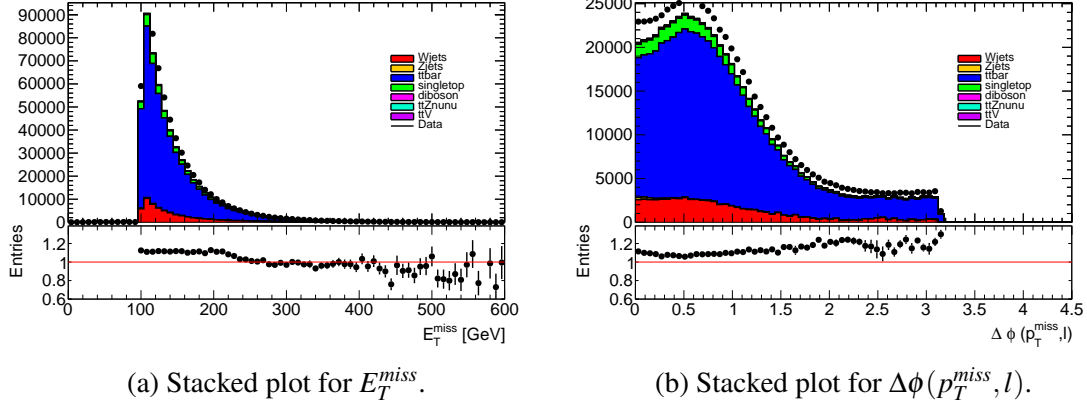


Figure 4: Stacked plots showing the contributions of background events to the overall signal for $t\bar{t}(Z \rightarrow \nu\bar{\nu})$ at preselection level.

Figure 4 shows the stacked background data for this signal region with the real data superimposed on top. This allows for easy judgement of the accuracy of the Monte Carlo (MC) simulation. The ratio panel beneath shows, with more clarity, the spread of data points around the MC signal line. The shape is in good agreement and this validates the usage of the MC simulation throughout the project. Despite this, the data line for $\Delta\phi(p_T^{miss}, l)$ remains approximately 10% higher than the predicted signal line, suggesting that the scaling of certain background events in separate control regions is inaccurate. However, both the shape of the plot and the 10% disagreement with data, correspond to an identical plot in [2].

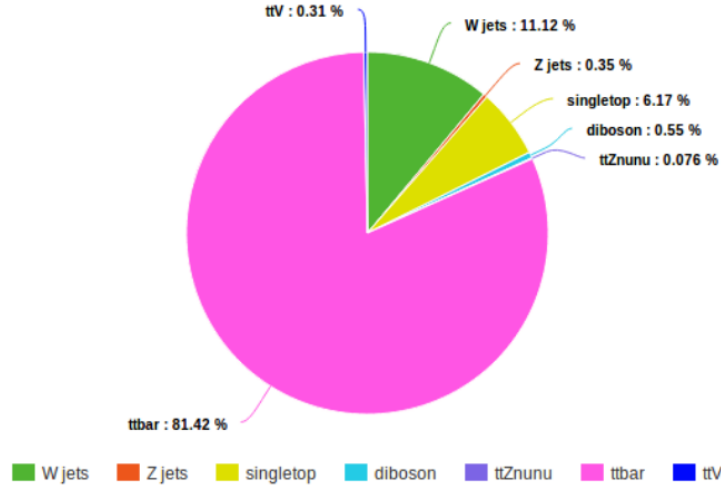


Figure 5: Pie chart showing the proportions of background events at preselection level.

From figure 5 it is clear that a different signal region is needed to study $t\bar{t}(Z \rightarrow \nu\bar{\nu})$ since the purity of this process at preselection level is only 0.076%.

5.2 DM high and DM low

After analysing the plots produced using the preselection, new signal regions (DM low and DM high) were studied. This involved applying additional cuts as well as the preselection. These cuts are demonstrated in table 1. These cuts were adapted from those used in the SUSY search[1] due to the increased purity of $t\bar{t}(Z \rightarrow \nu\bar{\nu})$ that they produce, as can be seen in figure 6.

Table 1: Selections defining signal regions DM low and DM high.[1]

Variable	DM low	DM high
Jet p_T [GeV]	$> (120, 85, 65, 25)$	$> (125, 75, 65, 25)$
b-jet p_T [GeV]	> 60	> 25
E_T^{miss} [GeV]	> 320	> 380
m_T [GeV]	> 170	> 225
am_{T_2} [GeV]	> 160	> 190
$\Delta\phi(p_T^{miss}, l)$	> 1.2	> 1.2
$\Delta\phi(jet_i, p_T^{miss})$	> 1.0	> 1.0

Under both the DM low and the DM high cuts, the $t\bar{t}$ background dominated but less so than had been the case at preselection level. Both of these regions increased the purity of $t\bar{t}(Z \rightarrow \nu\bar{\nu})$ significantly as can be seen in figure 5, however the number of total events dropped dramatically from $\sim 560,000$ to ~ 50 .

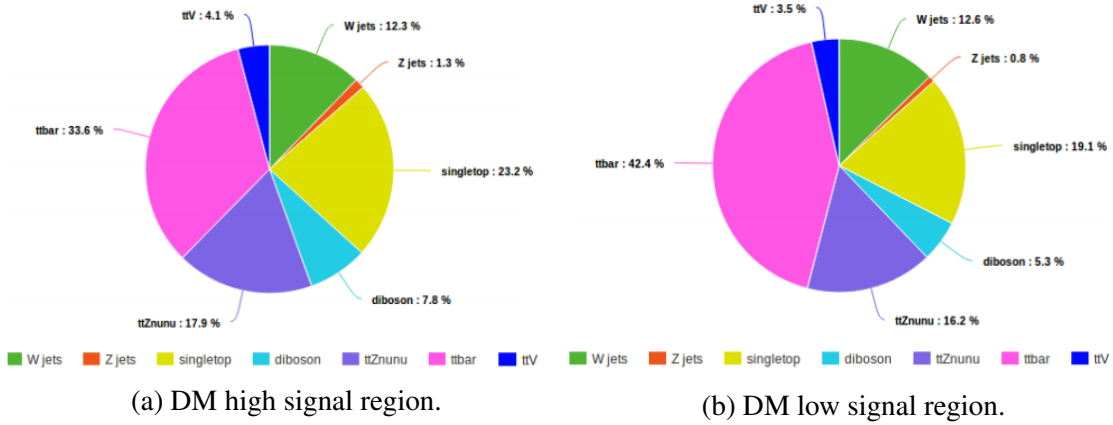
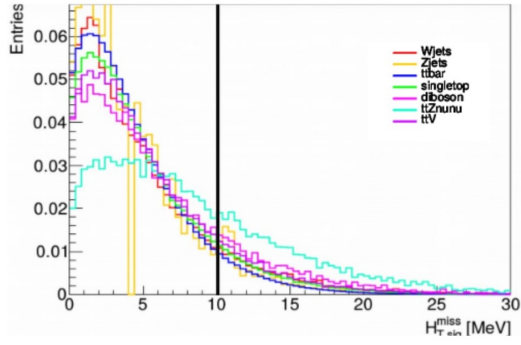


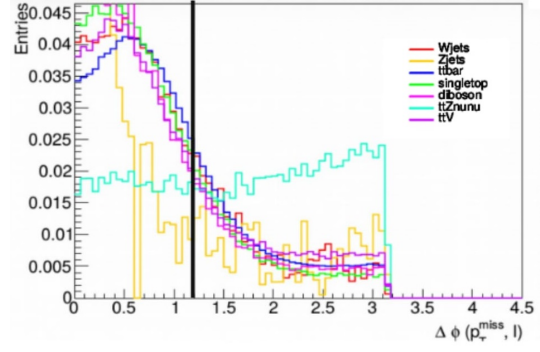
Figure 6: Pie chart showing the proportions of processes occurring in different signal regions.

5.3 $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ signal region

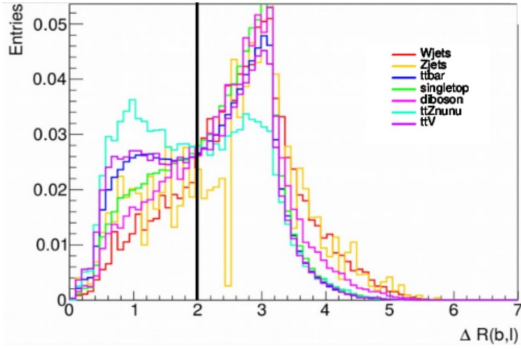
In order to study the $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ background process more clearly, a signal region that optimised its purity had to be defined by making appropriate cuts. The overlaid graphs from the preselection were studied by eye to determine which cuts would increase the proportion of this decay. Figure 7 shows some examples of the selections that were made and demonstrates why. Other selections that were made are summarised in table 2.



(a) Distributions for $H_{T,sig}^{miss}$. Selection made such that $H_{T,sig}^{miss} > 10$ was required.



(b) Distributions for $\Delta\phi(p_T^{miss}, l)$. Selection made such that $\Delta\phi(p_T^{miss}, l) > 1.2$ was required.



(c) Distributions for $\Delta R(b - jet, l)$. Selection made such that $\Delta R(b - jet, l) < 2$ was required.

Figure 7: Overlaid plots for comparing distributions of different background processes for different variables. The black line indicates where a selection has been made based on $t\bar{t}(\rightarrow \nu\nu)$ behaviour.

Many of the trial signal regions reduced the total number of events (including the events of interest) too extremely, so a compromise had to be made. Three signal regions were fine tuned and labelled A, B and C. The purities of $t\bar{t}(\rightarrow \nu\nu)$ events in these regions were 4.4%, 5.7% and 19.7% and the number of $t\bar{t}(\rightarrow \nu\nu)$ events in these regions were 61, 43 and 7.4 respectively. Table 2 shows these signal regions in more detail.

Table 2: Selections defining signal regions for $t\bar{t}(\rightarrow \nu\bar{\nu})$.

Variable	A	B	C
E_T^{miss} [GeV]	> 250	> 280	> 380
m_T [GeV]	> 100	> 125	> 225
am_{T_2} [GeV]	-	-	> 190
$H_{T,sig}^{miss}$	> 10	> 10	> 10
$\Delta\phi(p_T^{miss}, l)$	> 0.9	> 1.0	> 1.2
$\Delta\phi(jet_i, p_T^{miss})$	> 0.6	> 0.7	> 1.0
$\Delta R(b - jet, l)$	-	-	< 2

In order to assess which of these regions was optimal, the percentage uncertainties had to be calculated. This was done using equation 3. Despite having much fewer events, signal region C's increased purity meant it had lower percentage uncertainties than regions A and B. It was for this reason that this region was selected.

Despite the significantly improved purity, it was still not enough to reduce the uncertainties to a satisfactory level. With 10% systematic uncertainty, the 19.7% purity meant that the percentage uncertainty was still 92%. Figure 8 shows the variation of percentage uncertainty with changing parameters. It is clear from this plot that the most significant improvement in order for this process to be useful is the purity rather than the systematic uncertainty.

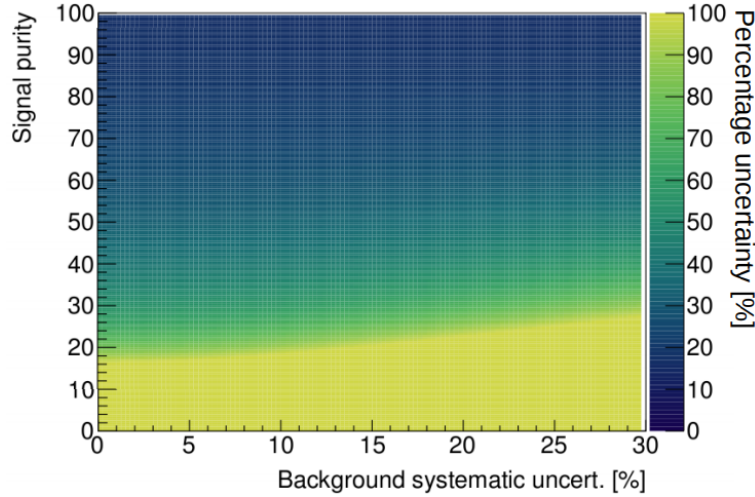


Figure 8: Plot showing how the percentage uncertainty on the number of $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ events depends on the systematic uncertainty and the purity of the signal region.

6 Investigating $t\bar{t}Z(\rightarrow l\bar{l})$ and background decays

6.1 Preselection

By modifying the preselection conditions so that three final state leptons were required rather than one, it was possible to analyse the data for the $t\bar{t}Z(\rightarrow l\bar{l})$ process. Two new variables were also studied, the dileptonic mass and the dileptonic transverse momentum. Figure 9 shows the stacked plots for E_T^{miss} and dilepton mass. The purity of this sample at preselection is significantly better than for $t\bar{t}Z(\rightarrow \nu\bar{\nu})$.

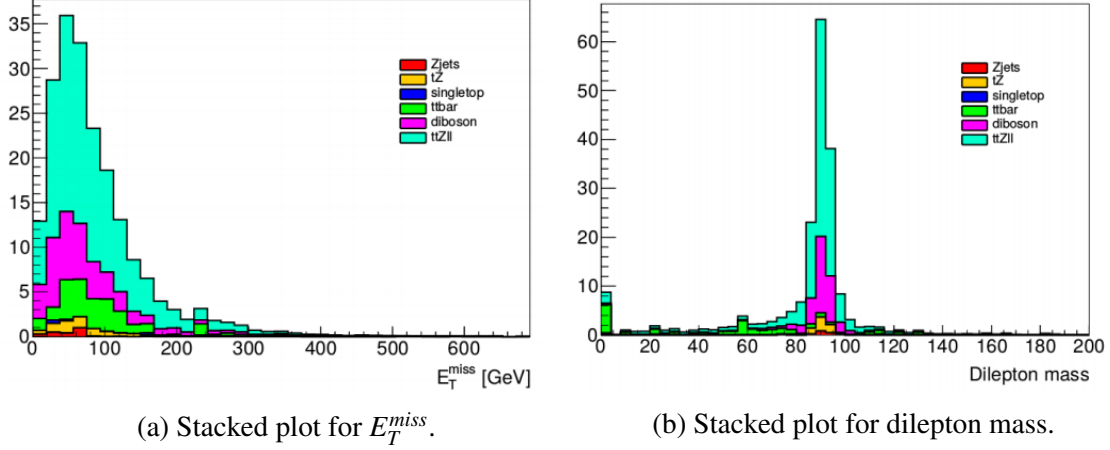


Figure 9: Stacked plots demonstrating the contributions of background processes of $t\bar{t}Z(\rightarrow l\bar{l})$ to the overall signal.

6.2 $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ signal region

In order for the ratio in equation 1 to be computed, the data from the $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ and $t\bar{t}Z(\rightarrow l\bar{l})$ processes must be from the same signal region. Since $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ is the process of interest it is natural to choose region C. The cuts corresponding to this region were applied to the events for $t\bar{t}Z(\rightarrow l\bar{l})$ and its background processes. Unfortunately, due to the severity of the cuts in this region, the events were reduced so significantly that the percentage uncertainty on the number of $t\bar{t}Z(\rightarrow l\bar{l})$ signals was 184%. It was for this reason that a calculation of the ratio in equation 1 was not yet useful.

7 The future

The current luminosity of the LHC is $36.5 fb^{-1}$. This is from data collected in 2015 and 2016. By the end of Run 2 in 2018, this luminosity will have increased to approximately $100 fb^{-1}$, and there are plans to begin the new High Luminosity LHC in 2025 which will have a luminosity of $3000 fb^{-1}$. With increased luminosity comes increased number of events, so naturally this means that the percentage uncertainties will be reduced. Tables 3 and 4 shows the extrapolation of the known data to summarise how this increased luminosity will improve the uncertainties for $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ and $t\bar{t}Z(\rightarrow l\bar{l})$ respectively.

Table 3: How percentage uncertainty scales with luminosity for $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ in signal region C.

Luminosity (fb^{-1})	Total expected number of events	Number of signals	Number of background events	Percentage uncertainty
36.5	37.3	7.4	29.9	92%
100	102.2	20.3	81.9	64%
3000	3065.8	608.2	2457	41%

Table 4: How percentage uncertainty scales with luminosity for $t\bar{t}Z(\rightarrow l\bar{l})$ in signal region C.

Luminosity (fb^{-1})	Total expected number of events	Number of signals	Number of background events	Percentage uncertainty
36.5	1.65	0.70	0.95	184%
100	4.52	1.9	2.6	113%
3000	136	58	78	24%

It is clear from the percentage uncertainties in tables 3 and 4 that increasing the luminosity alone is not sufficient for reducing the error, just as it is clear from figure 8 that reducing the systematic uncertainty is not sufficient either. In order to have the ability to compute the ratio of cross sections and draw conclusions from it, a signal region with much higher purity needs to be found. A possible method of achieving this would be to use a boosted decision tree (BDT) to improve SRC as opposed to the method of observation and trial and error implemented in this study.

Another possible method of extending this search for dark matter would be to analyse alternate $t\bar{t}$ decay channels to neutrinos using the same method, as perhaps this would lead to more events. What is likely to be more effective however, is to investigate the possibility of finding new variables to measure in order to find a more optimal signal region in which to study $t\bar{t}Z(\rightarrow \nu\bar{\nu})$.

8 Conclusion

Data from $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ decays were analysed due to the similarity of this process to the $t\bar{t}$ decay into dark matter via a spin 0 or spin 1 mediator boson, m_ϕ . It was assumed that any occurrence of dark matter production would be disguised as the production of $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ due to the correspondence between the signatures of these processes. $t\bar{t}Z(\rightarrow l\bar{l})$ was also investigated as a normalisation upon which to reduce the uncertainty on the cross section and provide an experimental method of comparing this to the accepted branching ratios.

Despite the optimisation of a control region, the separation of $t\bar{t}Z(\rightarrow \nu\bar{\nu})$ from its background signals proved too difficult to provide values with sensible uncertainties, and in order for the cross section ratio, $\sigma(t\bar{t}Z(\rightarrow \nu\bar{\nu}))/\sigma(t\bar{t}Z(\rightarrow l\bar{l}))$, to be useful, measures need to be taken to significantly improve the purity of the signal. A possible method of achieving this is to use a boosted decision tree to select a signal region more efficiently. Another alternative would be to find new variables with which to optimise the signal region further.

The feasibility of using specific decays to perform a general search for invisible particles was judged to be possible, but only with processes with limited background signals. It is hoped that with improved sample purity, increased luminosity and reduced systematic uncertainty, this method could still lead to the discovery of new phenomena in particle physics.

References

- [1] Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 13\text{TeV}$ p p collisions using 36.5 fb^{-1} of ATLAS data. ATL-COM-PHYS-2016-1623. *D. Brner, T. Eifert, J. Gonski. et al.*
- [2] Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 13\text{TeV}$ p p collisions using 36.5 fb^{-1} of ATLAS data. ATLAS CONF Note, ATLAS-CONF-2017-037. *The ATLAS Collaboration.*
- [3] Measurement of detector-corrected observables sensitive to the anomalous production of events with jets and large missing transverse momentum in pp collisions at $\sqrt{s} = 13\text{TeV}$ using the ATLAS detector. CERN-EP-2017-116. *The ATLAS Collaboration.*