Suche nach $t\bar{t}$ Resonanzen mit ATLAS Daten

Judith Gnade
judith.gnade@tu-dortmund.de

Hendrik Lauersdorf
hendrik.lauersdorf@tu-dortmund.de

John Wendel
john.wendel@tu-dortmund.de

July 25, 2022

TU Dortmund – Fakultät Physik

Contents

1 Introduction

In this lab course data from the ATLAS experiment at the LHC is used to search for new particles decaying to a top quark and an anti-top quark. As an extension of the Standard Model of particle physics, a resonance is predicted that might be such a new particle. In the Standard Model of particle physics the top quark is the most massive elementary particle known to date. It can be produced in $p\bar{p}$ collisions. In most cases a $t\bar{t}$ pair is produced via the strong interaction. The $t\bar{t}$ pair decays via the weak interaction to a W boson and a down-type quark, where the down-type quark usually is a bottom-quark b. The W boson can either decay into a lepton and a neutrino or into a pair of quarks: $W^+ \to l^+\nu_l, W^+ \to u\bar{d}$ or $W^+ \to c\bar{s}$.
The quarks in the decay hadronize and form jets of color neutral products. The channel in which both W bosons decay to a lepton and its corresponding neutrino is called the *dilepton channel*, whereas the channel in which both W bosons decay hadronically is called the *all-hadronic channel*. The channel in which one W boson decays leptonically and the other decays hadronically is called the *lepton+jets channel*.

graphics/decay.png

Figure 1: Feynman diagram of top quark production and the following decay to a *lepton+jets channel*.

Since there is much background in the *all-hadronic channel*, the *lepton+jets channel* is preferred for use in top-quark analyses. The *dilepton channel* has very little branching fractions. The decay particles can be registered in a detector and are reconstructed by algorithms which use the different signatures left by the particles that propagated through the detector. To distinguish events with top quarks in the final state from background events, the missing transverse momentum is used to determine whether a neutrino has been produced. This is the case for the *lepton+jets channel*. Jets originating from other quarks than a b-quark are distinguished by identifying a B-hadron within the b-jets.
First, to search for a hypothetical massive Z' boson decaying to top quarks, the size of the dataset is reduced. By event selection the signal-to-background ratio is increased. To model the background spectrum Monte Carlo simulations are used. MC simulations and data are then compared to check the validity of the simulations. A final discriminant is chosen and is used in a statistical analysis to determine the amount of signal on top

of the background distribution.

2 Analysis

2.1 Preselection

To reduce the large amount of data, a lose preselection was applied to the dataset beforehand. This preselection includes some requirements for each event. Some of the given variables in the data sample have more entries that there are entries in the sample itself. This is because events with more than one lepton or more than one jet have a seperate entry of variables connected to the leptons or jets. Some triggers are applied to the data as well. Since background leptons are difficult to seperate from signal leptons, a cut on the transversal momentum of the lepton is was made at 20 GeV to reduce combinatorial background. This can be shown in the `lep_pt` of one of the data samples in figure 2. The pseudorapidity shows two areas ($|\eta| \approx 1.4$) with a very low amount of events. In this area cables and other not detecting elements are placed so that the geometric acceptance of the detector is not at 100 %. The azimuth angle ϕ does not show any restrictions in the data set.

`plots/TB_lep_pt.png`

Figure 2: Distribution fo the transversal momentum of the lepton. This plot is a screenshot taken from the `TBrowser` in the `ROOT` enviroment.

2.2 Event selection

The chosen channel in this analysis is the `lepton+jets` channel where exactly one charged lepton is produced via W boson decays. A minimum transversal momentum of `lep_pt` > 50 GeV is recommended for this lepton. Events that include a lepton with

a smaller transversal momentum are being removed. This process also includes a neutrino which cannot be observed by the detector, so that a missing transversal energy of `met_et` $> 40\,\text{GeV}$ is required as well. Since both top quarks decay into a bottom quark and a W boson, at least a jet number of `jet_n` < 4 is required in the selection because of one W boson decaying into two quarks as well. Two of these jets have to be b-tagged. A `jet_pt` $> 40\,\text{GeV}$ is recommended for all jets and at least one with a `lep_pt` $> 100\,\text{GeV}$. Because of the detector acceptance, the pseudorapidity is set to a value of $|\eta| < 2.5$.
The coefficient $\epsilon \cdot A$ (efficiency times detector acceptance) is calculated as the ratio of events after each of the different requirements in the event selection to the total number of events in the sample.
The efficiencies after each step as well as the corresponding sample is found in table 1. Cuts on the pseudorapidity of the lepton and the jets are made after `lep_pt` and `jet_pt`. These do not show any significant change in the efficiency of any sample, thus they are not shown in the table.
After this event selection $t\bar{t}$ prodiction is the dominant background since it has a very similar final state as the signal channel.

2.3 Fundamental distributions

The dominant background process is the $t\bar{t}$ production. In the following, distributions of variables of simulated $t\bar{t}$ Monte Carlo events with muons are shown and compared to the expected distribution. Some of these distriutions are shown in figure **??**. These are in agreement with the expected distributions of these variables. In figure 3b, the number of b-tagged jets in each event is shown. Figure 1 shows the production of two bottom quarks so that most likely two jets are b-tagged. This expectation is in agreement with figure 3b. The remaining momentum that is not in the jets is evenly split on the muon and the neutrino. The distributions of the `lep_pt` and `met_et` is very similar. Both have their respective peak at the same location. The `met_et` is more smeared out since the indirect measured neutrinos have a worse resolution. Besides that these variables follow their expected distributions.

2.4 Derived quantities

After the event selection, some background processes like $t\bar{t}$ events are still not negligible. To have a better discrimination, some quantities are being computed that might increase the seperation power between Z' and $t\bar{t}$ events. The following quantities were computed and compared in Z' and $t\bar{t}$ events:

- the difference of the azimuth angle of the missing transversal energy and the lepton flight direction $\Delta\Phi$
- the invariant mass of the system formed by the three jets with the largest transversal momentum
- the invariant mass and the pseudorapidity of the system formed by the four jets with the largest momentum, the muon and the neutrino

Table 1: Calculated efficiencies after cutting on different variables.

Sample	lep_n	lep_pt	jet_n	jet_pt	btagged	met_et
ttbar.el	0.9316	0.5070	0.2440	0.1408	0.0606	0.0430
ttbar.mu	0.9200	0.4821	0.2328	0.1342	0.0580	0.0410
singletop.el	0.9806	0.4323	0.0532	0.0311	0.0103	0.0071
singletop.mu	0.9786	0.4104	0.0498	0.0288	0.0097	0.0068
diboson.el	0.9044	0.3813	0.0119	0.0048	0.0001	7.9691e-05
diboson.mu	0.8731	0.3510	0.0091	0.0038	9.99691e-05	7.38902e-05
wjets.el	0.9999	0.1709	0.0029	0.0017	1.99407e-05	1.39121e-05
wjets.mu	0.9999	0.1589	0.0027	0.0016	2.33554e-05	1.56505e-05
zjets.el	0.7167	0.1814	0.0067	0.0032	0.000135906	4.8692e-05
zjets.mu	0.5015	0.1256	0.0025	0.0014	7.37231e-05	3.18146e-05
zprime400.el	0.9308	0.4464	0.1640	0.0470	0.0192	0.0124
zprime400.mu	0.9201	0.4215	0.1597	0.0539	0.0203	0.0137
zprime500.el	0.9324	0.5616	0.2539	0.1501	0.0626	0.0418
zprime500.mu	0.9224	0.5353	0.2319	0.1358	0.0581	0.0389
zprime750.el	0.9237	0.6705	0.3881	0.3503	0.1573	0.1216
zprime750.mu	0.9133	0.6451	0.3710	0.3332	0.1526	0.1220
zprime1000.el	0.9339	0.7342	0.4599	0.4401	0.2088	0.1790
zprime1000.mu	0.9182	0.7055	0.4573	0.4335	0.2023	0.1682
zprime1250.el	0.9355	0.7475	0.4811	0.4667	0.2158	0.1900
zprime1250.mu	0.9246	0.7319	0.4867	0.4704	0.2218	0.1932
zprime1500.el	0.9401	0.7711	0.5014	0.4904	0.2121	0.1904
zprime1500.mu	0.9287	0.7483	0.5012	0.4906	0.2126	0.1894
zprime1750.el	0.9473	0.7758	0.5140	0.5038	0.2146	0.1932
zprime1750.mu	0.9383	0.7571	0.5036	0.4918	0.2091	0.1898
zprime2000.el	0.9451	0.7629	0.4988	0.4876	0.1914	0.1749
zprime2000.mu	0.9406	0.7611	0.5171	0.5069	0.1971	0.1802
zprime2250.el	0.9450	0.7451	0.4948	0.4841	0.1846	0.1688
zprime2250.mu	0.9448	0.7640	0.5183	0.5086	0.2012	0.1824
zprime2500.el	0.9480	0.7509	0.4958	0.4811	0.1762	0.1600
zprime2500.mu	0.9433	0.7595	0.5211	0.5080	0.1933	0.1765
zprime3000.el	0.9437	0.7151	0.4510	0.4268	0.1656	0.1463
zprime3000.mu	0.9398	0.7314	0.4833	0.4633	0.1781	0.1607

plots/ttbar.mu_selected_lep_pt.pdf

(a)

plots/btagged.png

(b)

plots/ttbar.mu_selected_jet_pt.pdf

(c)

plots/ttbar.mu_selected_met_et.pdf

(d)

Figure 3: Plots of different variables of the $t\bar{t}$ Monte Carlo simulations. The following variables are shown here: The transversal momentum of the lepton (a), the amount of b-tagged jets (b), the transversal momentum of the jets (c) and the missing transversal energy (d).

To calculate the invariant mass of the system including neutrinos, the class `TLorentzVektor` was used to compute the four-vector information. While calculating $\Delta\Phi$, only the smaller difference of $|\phi_1 - \phi_2|$ and $|2\pi - |\phi_1 - \phi_2|$ was used. These quantities were calculated for all data and Monte Carlo samples. To determine the quantity with the largest separation power, the dominant background ($t\bar{t}$) is compared with a possible signal ($Z'(1000)$). The best discriminate is the invariant mass of the full system. This quantity is shown in figure 4 for $t\bar{t}$ and $Z'(1000)$.

plots/ttbar.mu_selected_SysJetMass.pdf

(a)

plots/zprime1000.mu_selected_SysJetMass.pdf

(b)

Figure 4: Distribution of the chosen discrimination quantity, the invariant mass of the system formed by the four jets with the largest transversal momentum, the muon and the neutrino. This quantity is shown for $t\bar{t}$ simulations (a) and simulations of $Z'(1000)$ (b).

3 Agreement between data and simulation

After the selection is done, the agreement of data and simulation is investigated. For that, the expected amount of event is being calculated using the formula

$$N = \mathcal{L}\sigma(A\epsilon)\,. \tag{1}$$

Here $(A \cdot \epsilon)$ is the acceptance times the efficiency of the event selection that was calulated in section 2.2, σ is the cross section of the respective event and $\mathcal{L} = 1\,\mathrm{fb}^{-1}$ is the integrated luminosity of the dataset. Cross sections as well as expected amount of events in the chose data sample (`data.mu.2.root`) are shown in table 2. A total amount of 70.42 background events is expected in this sample. After selection, the total amount of events is $N_{tot} = 99$.

Table 2: Anzahl erwarteter Ereignisse für die einzelnen Prozesse im sample `data.mu.2.root`. Angegeben ist der zugehörige Wirkungsquerschnitt der für die Berechnung der einzelnen Werte nach Formel (1) benötigt wird.

Prozess	N_{expected}	σ / pb
`ttbar`	3.1893	252.82
`singletop`	3.5362	52.47
`diboson`	3.1561	29.41
`zjets`	6.6564	2516.20
`wjets`	53.8832	36214
`zprime400`	108.9000	1.1e2
`zprime500`	81.1800	8.2e1
`zprime750`	19.8000	2.0e1
`zprime1000`	5.4450	5.5
`zprime1250`	1.881	1.9
`zprime1500`	0.8217	8.3e-1
`zprime1750`	0.2970	3.0e-1
`zprime2000`	0.1386	1.4e-1
`zprime2250`	0.0663	6.7e-2
`zprime2500`	0.0346	3.5e-2
`zprime3000`	0.0119	1.2e-2

For these events, weights need to be applied since they are just refering to their respective sample. These weights normalize the MC samples to the data sample. The weights are calculated after

$$w = \frac{\mathcal{L}\sigma}{\mathrm{N}_{\mathrm{MC}}} . \tag{2}$$

N_{MC} is the number of MC events before the preselection. The sizes of the samples as well as the calculated weights are shown in table 3.

Table 3: Calculated weights for the different processes. The cross section used to calculate the weights is shown in table 2. The integrated luminosity is $\mathcal{L} = 1\,\text{fb}^{-1}$. The weights are calculated using (2).

process	N_{MC}	weights w
`ttbar`	7847944	0.03221
`singletop`	1468942	0.03572
`diboson`	922521	0.03188
`wjets`	66536222	0.54428
`zjets`	37422926	0.06724
`zprime400`	100000	1.10000
`zprime500`	100000	0.82000
`zprime750`	100000	0.20000
`zprime1000`	100000	0.55000
`zprime1250`	100000	0.19000
`zprime1500`	100000	0.08300
`zprime1750`	100000	0.03000
`zprime2000`	100000	0.01400
`zprime2250`	100000	0.00067
`zprime2500`	100000	0.00035
`zprime3000`	100000	0.00012

After applying these weights, stacked plots are being produced. For that the different background process distributions are stacked and the data events are added to compare them to the expected background. Four of these plots are shown in figure 5.

plots/stacked_lep_pt.pdf

(a)

plots/stacked_jet_eta.pdf

(b)

plots/stacked_jet_pt.pdf

(c)

plots/stacked_lep_E.pdf

(d)

Figure 5: Stacked background samples in comparison to the data sample. Shown is the transversal momentum of the muon (a), the jet pseudorapidity (b), the transversal momentum of the jets (c) and lepton energy (d).

All four plots show a mostly good agreement of expected background and data. Some data points are higher than the expected background which could be a hint on a Z' event. In this case there are also some data points lower than the expected background, indicating that these are just statistical fluctuations.
Figure 6 shows the discriminating variable, the invariant mass of the system.

plots/stacked_mass.pdf

Figure 6: Distrubution of expected background events and data of the invariant mass of the system.

The same properties as before can be observed in this plot. There are some data points higher than the expected background and some that are significantly lower. This again indicates that these peaks are just statistical fluctuations.

4 Statistical analysis

In the last chapter, the agreement between MC and data was qualitatively inspected. To judge the agreement between data and the background-only hypothesis, a χ^2 test will be performed using MC simulations as the background-only hypothesis. The value of χ^2 is a metric for the deviation of the data and MC sample. If the background-only hypothesis and data are in a good agreement, meaning no Z' exists, χ^2 will have a small value. In this case an exclusion limit will be set on the production cross section. For that a confidence level is needed that is calculated with

$$\mathrm{CL} = 1 - p \tag{3}$$

with p as the p-value for the background-only hypothesis.
The results of the χ^2 test are:

$$\chi^2 = 25.1862$$
$$p = 0.508458$$

Even though a high χ^2 value is achieved, the p-value is very high as well. This shows that the found peaks in the invariant mass is just a statistical fluctuation and that no observation of a Z' is found. Therefore an exclusion limit for the different Z' mass hypotheses will be calculated in a 95 % confidence level. To consider the weights, a scaler will be used that scales the weights until the confidence level is reached. These results are shown in table 4.

Table 4: Calculated parameters of the χ^2 test for the different Z' mass hypothises as well as the cross section exclusion limit in a 95 % confidence level.

process	degrees of freedom	p-value	scaler	σ/pb
$Z'(400)$	26	0.00549806	0.4	44
$Z'(500)$	26	0.00211708	0.2	16.4
$Z'(750)$	26	0.0150853	0.2	4
$Z'(1000)$	26	0.0472768	0.4	2.2
$Z'(1250)$	26	0.0394068	0.9	1.71
$Z'(1500)$	26	0.0433432	1.9	1.58
$Z'(1750)$	26	0.0462103	4.8	1.44
$Z'(2000)$	26	0.0499252	10.9	1.526
$Z'(2250)$	26	0.0495314	23.5	1.57
$Z'(2500)$	26	0.0498703	48.7	1.70
$Z'(3000)$	26	0.0499869	194.7	2.33

The resulting cross section at the 95 % confidence level are being compared against the expected cross sections in figure 7. Since the observed cross section is smaller than theoretically predicted, masses of $m_{Z'} < 1300\,\mathrm{GeV}$ can be excluded. A particle with a lower mass would have been observed in this analysis. Masses of $m_{Z'} > 1300\,\mathrm{GeV}$ could not be discriminated from the background so that these masses are not excluded in this analysis.

plots/limits.pdf

Figure 7: Calculated upper limits of the cross section in a 95 % confidence level for the different Z' mass hypothises. For comparison, the expected cross section is plotted as well.

5 Discussion

The selected data possesses high fluctuations when compared with the expected background. If the errors of the data is calculated poisson-like, these fluctuations are still in the standard deviation. This indicates that the difference between Monte Carlo and data are most likely statistical. Using the full data set instead of just one sample could lower these statistical fluctuations.
The statistical analysis seemed to worked fine and the lower mass limit of Z' is at 1300 GeV. Again, using the full dataset could increase this lower limit.
In terms of this lab course it was very helpful in understanding how a more or less complete data analysis at ATLAS is looking like. Some complications were appearing when facing the languages `C++` and `Root` since they were not that common to use. A small introduction or examples would be very helpful.