Top quarks in hot dense matter in the CMS detector

Luuk Vermunt
Supervised by: M. Verweij and P. Silva
August 25, 2016

Abstract

A feasibility study of the measurement of the top quark pair production cross section using the muon+jets channel in heavy-ion collisions is presented. Data, corresponding to a total integrated luminosity of 0.404 nb$^{-1}$ at 5.02 TeV/nucleon, accumulated by the CMS experiment is used. After establishing a robust event selection we evaluate the expected composition of the sample in data. Due to the still low $S/B$ ratio, no observation of this process has been made in our study. Further improvements on the baseline selection studied in this note are furthermore discussed.

1 Introduction

During my 9 week stay at CERN in the summer of 2016, I have been working on the measurement of the top quark in heavy-ion collisions. Top quarks, which are the heaviest detected quarks, have up to now only been observed in proton-proton collisions. Also the Standard Model Higgs boson and the $\tau$ are still not detected in heavy-ion collisions. Recently it was noted by d’Enterria et al. in [1] that top quarks, produced in pairs or singly, have to be clearly observable in p-Pb and Pb-Pb collisions at the LHC energies. This report is presenting the first study into such a measurement of top events using the dataset of 0.404 nb$^{-1}$ in the special PbPb run at 5.02 TeV/nucleon accumulated by the CMS experiment at the LHC.

There are several theoretical motivations for the measurement of the top quark in heavy-ion collisions. It can for example be used to put constraints on the nuclear parton distributions functions, study the heavy-quark energy loss dynamics or open new opportunities to study the dynamics of the Quark-Gluon Plasma (QGP). In this project we are mostly interested in this last motivation. Because of the short decay time of the top quark, it will decay before the thermalization of the medium. Therefore, its colour charged decay products will interact strongly with the QGP during its whole existence. This interaction will lead to a reduction of the energy of the jets in comparison with the same event in a collision without the accompanying QGP. This energy reduction phenomenon is called jet quenching [2]. The differences in PbPb and pp collisions of a measured variable affected by jet quenching, will therefore be an interesting probe to study the QGP.
In this study top quark pair production was explored. The top quark itself decays promptly and almost exclusively into a $W$ boson and a $b$ quark ($\text{BR}(t \rightarrow Wb) \approx 1$). The $b$ quarks will form jets and the $W$ bosons have two ways to decay, fully hadronic or fully leptonic. So top quark pair production will have three main final states:

1. $t\bar{t} \rightarrow WbW\bar{b} \rightarrow 2b\text{-jets} + 4\text{light-jets}$,
2. $t\bar{t} \rightarrow WbW\bar{b} \rightarrow 2b\text{-jets} + 2\text{light-jets} + l + \nu_l$,
3. $t\bar{t} \rightarrow WbW\bar{b} \rightarrow 2b\text{-jets} + 2l + 2\nu_l$.

Here $l$ can be an electron, muon or tau (or their anti-particles) and $\nu_l$ is the corresponding anti-neutrino (or neutrino). Because neutrino’s cannot be detected, they are usually observed by measuring the missing $E_T$ (MET). Due to the large underlying event, the performance of the MET estimator is degraded in PbPb collisions with respect to pp. Therefore, it is not used in this study.

In our study the lepton+jets final state is explored where the lepton is a muon. This process has a branching ratio of 16.9% when the feed down of the tau to the muon is included. The branching ratio is approximate ten times higher than the one of the dilepton final states, but this final state will be a lot harder to reconstruct. The main goal of this project is to identify the $W$ invariant mass peak of the two light jets. This peak is not expected to be at the normal value because of the jet quenching. The peak will be completely shifted towards lower masses, or it will peak at the original position but with a tail towards the left.

2 Experimental strategy

This study made use of a single non-isolated high $p_T$ muon trigger. The background in this sample is expected to be dominated by $W$+jets, DY and QCD multijets productions. To estimate the signal and background distributions, the following Monte-Carlo event generators have been used:

- Signal process: Powheg+Pythia8 (hvq model)
- Background process:
  - $W$+Jets: Pyquen ($p_T > 30$ GeV)
  - DY: Pythia 8 ($M_{\mu\mu} > 10$ GeV)

The simulations include the embedding of the hard process in a minimum-bias simulated PbPb collision with HYDJET. The background contribution from QCD multijets is estimated by reversing the muon isolation cut in the data sample and normalizing this using the ABCD method [3]. The cross section of these processes can be calculated by simply scaling the measured cross section in proton-proton collisions according to the number of nucleons $A = 208$. The results can be found in Table [4]. If no reconstruction efficiencies are taken into account, 204 $t\bar{t} \rightarrow \mu$+jets events are expected in the full CMS data sample.
### Table 1: Cross sections for pp and PbPb at 5.02 TeV for the different processes. The cross section in the PbPb collisions is calculated using the measured cross section in pp collisions.

<table>
<thead>
<tr>
<th>Process</th>
<th>pp [pb]</th>
<th>PbPb (=pp·A^2) [μb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>tt</td>
<td>69</td>
<td>2.985</td>
</tr>
<tr>
<td>W+Jets</td>
<td>21458</td>
<td>928.359</td>
</tr>
<tr>
<td>DY</td>
<td>2010</td>
<td>86.961</td>
</tr>
</tbody>
</table>

2.1 Selection cuts

To select the right signal events and to decrease the background, selection cuts have been applied on all samples. They can be summarized in three categories: Event selection, Muon selection and Jet selection. All the applied selection cuts can be found in section A.1. We require:

1. At least one global muon with a $p_T > 18$ GeV and $|\eta| < 2$. This muon needs to be isolated.

2. At least 4 jets with a $p_T > 30$ GeV, $|\eta| < 2$ and a distance larger than 0.3 to the selected muon.

The muon isolation cut is split up into the different centrality intervals. Because of the results of a recent study of H. Doan using the same dataset [4], the pfCS candidate instead of the pf candidate is chosen. For this new candidate, the signal selection efficiency is 6.2% better for the 0-10% centrality interval. The cut values and ROC-plots for both muon isolation cut candidates can be found in Table A.1 and Figure A.1.

2.2 b-tagging

A way to decrease the background even further is by requiring one or two $b$-jets. Jets can be tagged as a $b$-jet by using discriminators. For each discriminator a working point is defined and all jets that have a value higher than this working point are tagged as a $b$-jet. In this study the performance of four different discriminators were studied: the CSVv1, Jet probability, the $p_T$ of the Simple Secondary Vertex algorithm and the Track Counting High Efficiency. Further details can be found in [5]. The four distributions for MC $b$- and light-jets from the MC $t\bar{t}$ sample can be found in Figure A.2. As we can see there, the power to distinguish between a $b$- and light-jet is not optimal. Both distributions for all discriminators look quite similar, so by placing a working point to have a large light-jet rejection efficiency, a lot of real $b$-jets will also not be tagged. The combined ROC plot for these four distributions is shown in Figure 1. For all discriminators, a light-jet rejection efficiency larger than 90% comes along with a reduced $b$-jet tagging efficiency of 20-30%. Because in this area of the plot there is no significant difference between the four discriminators, we used the CSVv1 discriminator in our analysis because this one is normally used in heavy-ion physics analysis.

In Figure A.3 the CSVv1 ROC plots for different centrality intervals can be found, where we defined the working point at 0.75 (denoted by the stars). The $b$-tagging efficiency is strongly dependent on the centrality interval. For peripheral collisions, the efficiencies are
almost comparable with the $b$-tagging efficiencies for pp collisions [5]. But for central events (where most of the top quark pair production will happen), the efficiency is very low. Using our working point for the 0-10% centrality interval, we will reject 74% of the background events and only select 26.8% signal events.

### 2.3 QCD multijets background estimation method

One of the main backgrounds is the QCD multijets background. This background consists out of events where the high $p_T$ muon is coming from a jet instead of the $W$ boson. The distribution of this background can be found by reversing the muon isolation cut on the data sample. The normalization factor can be estimated by the ABCD method. In this method four regions in a two-dimensional phase space of two uncorrelated observables are used. Only one of these regions can be dominated by the signal, and the other three must consist mostly out of QCD events. To achieve this, the correlation between these regions must be minimal. The amount of background events in the signal region can then be estimated by the assumption that the ratio of the regions is equal (so $N_C = N_A \cdot N_D/N_B$). In our analysis the muon isolation and muon $p_T$ variables are used. An example of the ABCD plot for the 0-10% centrality events can be found in Figure B.1.

To account for the jet quenching, the $p_T$ of all the MC jets is shifted and smeared according to results found by the CMS Collaboration using the same dataset [6]. We take 85% of the $p_T$ of the jet and smear this using a Gaussian distribution with a width of 3%. The effect of this smearing procedure can be found in Figure B.2, where we see the $H_T$ distribution of all the selected jets. The stack of the smeared signal and background is giving a more realistic picture of the data. Note that the smearing implementation in this analysis is only used as a toy model, so there are no different smearing factors specified for the different centrality intervals.
3 Results

In Figure 2 three pre-fit control distributions for different variables in events with at least 4 jets and 1 tagged b-jet are shown. In Appendix C all of the control distributions for 4j1b events and all the invariant mass control distributions for 3j1b and 4j2b are shown. As we can see, the shapes of the data distributions are fairly reproduced by the stack of the signal and background distributions. Only for the muon $p_T$ control plot, the MC distributions describe the data poorly.

As can be seen in the logarithmic control plots in the Appendix, the signal is almost three orders of magnitude lower than the data for all different jet selections. The applied selection cuts are not enough to decrease the background significantly in comparison with the signal. In Table C.1 the percentages of selected events after each cut for the three MC samples are shown. Only the jet selection cuts decrease the background about a factor 3-4 more than the signal. At the end of the selection 5% of the events are left for all the MC samples and because the cross section of W+Jets and DY is significantly higher than the cross section of $t\bar{t}$ production, the sample is expected to be background dominated.

In Table 2 the expected number of events for the signal and background distributions for different jet-selection requirements is shown. As an example in the 4j1b (4j2b) category, $S/(S+B) = 0.1\% (0.2\%)$ yielding still a background dominated sample, even after pp-like requirements for $t\bar{t}$ lepton+jets selection. When we look at the last two rows, we see again that the total number of selected events from the signal and background distributions describe the number of selected data events quite well.

4 Conclusion

As we can see in the control distributions and in Table 2 there is still a lot of work ahead before a muon+jets top event can be measured in heavy-ion collisions. Too few signal events are selected and the background is still too high. Nevertheless, the control distributions show a good agreement between data and predictions. This indicates that we already have a good understanding about the events we select in the data. The low $b$-tagging efficiency is one of the reasons for this dominance of the background. Also the wide invariant mass peaks of the MC $t\bar{t}$ sample are a consequence of this. If the $b$-tagging purity would be better, the peaks
would be less broad. So to do a real measurement of the $t\bar{t}$ cross section in PbPb collisions, the $b$-tagging efficiency and $b$-tagging purity have to improve significantly.

The next PbPb run will be in 2018 and we expect an integrated luminosity of 1 nb$^{-1}$. By simply scaling the numbers in Table 2 to this luminosity, we expect 17 signal events on a total of approximate 14k for the 4j1b requirement. But if we make the assumption that the $b$-tagging efficiency has improved significantly to 90% (instead of the 30% it is now) we already have 51 $t\bar{t}$ events in this channel. The observed limits of the here presented and the hypothetical 2018 analysis can be calculated using the asymptotic CL$_S$ method:

\[ r < 15.2 \cdot \text{SM expectation (this analysis)}, \]
\[ r < 3.4 \cdot \text{SM expectation (hypothetical 2018 analysis)}, \]

So we need approximate 225 times the 0.404 nb$^{-1}$ luminosity for this analysis and approximate 11.6 nb$^{-1}$ for the 2018 analysis to be on top of the standard model prediction.

Table 2: Total number of expected events for the different signal and background distributions in the 0.404 nb$^{-1}$ PbPb dataset for different jet selection requirements.

<table>
<thead>
<tr>
<th></th>
<th>3j0b</th>
<th>3j1b</th>
<th>4j0b</th>
<th>4j1b</th>
<th>4j2b</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC $t\bar{t}$</td>
<td>44</td>
<td>13</td>
<td>20</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>MC $W+Jets$</td>
<td>17897</td>
<td>3228</td>
<td>12666</td>
<td>2813</td>
<td>479</td>
</tr>
<tr>
<td>MC $DY$</td>
<td>2867</td>
<td>467</td>
<td>1930</td>
<td>397</td>
<td>61</td>
</tr>
<tr>
<td>Multijets</td>
<td>10397</td>
<td>2954</td>
<td>8071</td>
<td>2617</td>
<td>711</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>31205</td>
<td>6662</td>
<td>22687</td>
<td>5834</td>
<td>1254</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>24884</td>
<td>6585</td>
<td>19310</td>
<td>5830</td>
<td>1372</td>
</tr>
</tbody>
</table>

Table 2: Total number of expected events for the different signal and background distributions in the 0.404 nb$^{-1}$ PbPb dataset for different jet selection requirements.
5 Bibliography

References


Appendices

A ROC plots

A.1 Section cuts

- Event selection: Primary vertex filter, $|vz| < 15$, cluster compatibility filter, HBHE noise filter and HF coincidence filter $3$ GeV.
- Tight muon selection: Global muon, $\chi^2/\text{ndf} < 10$, $\geq 6$ tracker layers, $\geq 1$ pixel hits, $\geq 1$ muon hits, $\geq 2$ matched stations, $|d_0| \leq 0.2$ cm and $|d_z| \leq 20$ cm
- Muons (at least 1) have $p_T \geq 18$ GeV, $|\eta| \leq 2.1$ and muon isolation is required.
- Jets (at least 4) have $p_T \geq 30$ GeV, $|\eta| \leq 2$ and $\text{drJetToMuon} \geq 0.3$. Jets are reconstructed using the anti-$k_T$ algorithm with cones of 0.2.

A.2 Muon-Isolation

<table>
<thead>
<tr>
<th>Centrality</th>
<th>isoCut pf</th>
<th>isoCut pfCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10 %</td>
<td>3.45</td>
<td>0.58</td>
</tr>
<tr>
<td>10-30 %</td>
<td>2.05</td>
<td>0.45</td>
</tr>
<tr>
<td>30-50 %</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>50-70 %</td>
<td>0.5</td>
<td>0.24</td>
</tr>
<tr>
<td>70-100 %</td>
<td>0.2</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table A.1: Muon isolation cut values for the different centrality intervals for the pf and pfCS candidate.

Figure A.1: ROC plots from [4] for the two different muon isolation cut candidates, pf candidate (left) and pfCS candidate (right), for 5.02 TeV/nucleon PbPb collisions at the CMS detector.
A.3 b-tagging

Figure A.2: Distributions for MC b- and light-jets for the four different studied discriminators: CSVv1 (top left), prob (top right), svtxpt (bottom left) and tcHighEff (bottom right).

Figure A.3: Centrality dependence of the b-tagging efficiency (left) and signal event selection efficiency (right) of the CSVv1 discriminator. The defined working point CSVv1 > 0.75 is denoted by the stars.
B Other methods

B.1 ABCD method

Figure B.1: The ABCD method plot for the selected events in the 0-10% centrality interval of the PbPb 5.02 TeV data sample. The boundaries of the four regions are placed at $p_T = 18$ GeV and muon isolation = 0.58 and 1.1. Region C is defined as the signal region, and the amount of QCD multijets background events can be estimated using the ratios of the regions.

B.2 Jet $p_T$ smearing

Figure B.2: The effect on the $H_T$ distribution of the MC jet $p_T$ smearing toy model. On the left no smearing is applied and on the right all MC jets have been shifted and smeared using a Gaussian function with a mean of 0.85 and a width of 0.03. These values are taken from [6].
C Control distributions

C.1 4j1b

Figure C.1: Control distributions for muon $p_T$ (left) and muon $|\eta|$ (right) in normal and logarithmic scale for the 4j1b requirement.

Figure C.2: Control distributions for b-jet $p_T$ (left), $\Delta \phi$ between muon and leading-CSVv1-jet (middle) and $\Delta \phi$ closest to $\pi$ between muon and tagged b-jet (right) in normal and logarithmic scale for the 4j1b requirement.
Figure C.3: Control distributions for $M_{lb}$ (left), $M_{qq}$ (middle) and $M_{ttbar}$ (right) in normal and logarithmic scale for the 4j1b requirement. The wiggle between 50 and 100 GeV in the $M_{qq}$ plot is probably because of a failure in the jet reconstruction in the endcap.

C.2 3j1b

Figure C.4: Control distributions for $M_{lb}$ (left), $M_{qq}$ (middle) and $M_{ttbar}$ (right) in normal and logarithmic scale for the 3j1b requirement.
C.3 4j2b

Figure C.5: Control distributions for $M_{lb}$ (left), $M_{qq}$ (middle) and $M_{ttbar}$ (right) in normal and logarithmic scale for the 4j2b requirement.

C.4 Percentage decrease after selection cuts

<table>
<thead>
<tr>
<th>Sample</th>
<th>1Mu4j</th>
<th>..+MuonCuts</th>
<th>..+MuIsoCut</th>
<th>..+JetCuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC $tt$</td>
<td>55.5</td>
<td>29.8</td>
<td>25.3</td>
<td>5.6</td>
</tr>
<tr>
<td>MC W+Jets</td>
<td>86.2</td>
<td>65.3</td>
<td>62.3</td>
<td>4.2</td>
</tr>
<tr>
<td>MC DY</td>
<td>96.4</td>
<td>89.1</td>
<td>87.6</td>
<td>6.7</td>
</tr>
</tbody>
</table>

Table C.1: Percentage of MC events left after the different selection cuts (specified in section 2.1). Note that for the MC $tt$ sample also electrons and taus were simulated, so when a muon is required a lot of events are lost. The number between brackets is the percentage of events where the muon is coming from a $W$ boson decay, which corresponds with the simulated 33%.