Radiating top quarks

Gosselink, M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
In this chapter, a study of the top quark pair production cross section with the ATLAS detector in the commissioning phase is presented. The proposed measurement aims at a ‘rediscovery’ of the top quark, using a small amount of early collision data containing semi-leptonic \( t\bar{t} \). Since it was not expected that the \( b \)-tagging performance would be sufficiently known at this stage, \( b \)-tagging is not used in the analysis. The work is based on previous studies [158, 159] and is part of a larger effort of the ATLAS collaboration to document physics analyses and detector performance studies using simulated data [157], referred to as Computing System Commissioning (CSC).

Although the initial running of the LHC is actually at a centre-of-mass energy of 7 TeV, for this Monte Carlo study the assumption was 14 TeV. As a benchmark, the integrated luminosities of the simulated data samples are taken to be 100 pb\(^{-1}\) (equivalent to roughly ten days of data taking at a luminosity of \( 10^{32} \text{ cm}^{-2}\text{s}^{-1} \)). For a realistic simulation of the full detector response, a modified detector geometry is used with respect to the ‘as-built’ geometry\(^1\). The reconstruction is performed with release 12 of the ATLAS software framework.

First, the event topology of \( t\bar{t} \) will be studied in Section 5.1. In Section 5.2 the physics objects, which are the building blocks of the top cross section measurement, will be introduced. In Section 5.3 the criteria and efficiencies to select events with semi-leptonic \( t\bar{t} \) candidates will be given. The reconstruction of top quarks from the selected events is explained in Section 5.4. The actual \( t\bar{t} \) cross section extraction using a maximum likelihood fit is performed in Section 5.5. In Section 5.6, the systematic uncertainties associated with this method are mentioned and, finally, discussion on the measurement results takes place in Section 5.7.

\(^1\)In particular, misalignments are introduced for the inner detector and extra material is added in the inner detector and in front of the calorimeters. In addition, distorted magnetic field configurations are introduced, where the symmetry axis of the field does not coincide with the beam axis.
Chapter 5. Cross section at \( \sqrt{s} = 14 \) TeV

5.1 Event topology

The decay of a top quark pair can be classified according to the subsequent decay of the two \( W^{\pm} \) bosons in ‘fully hadronic’ \((4/9^{\text{th}})\), ‘semi-leptonic’ \((4/9^{\text{th}})\), and ‘di-leptonic’ \((1/9^{\text{th}})\). The fully hadronic decay results typically in six or more jets of which on average two \( b \)-jets. Semi-leptonic decay is illustrated in Figure 5.1 and gives at least four jets of which two \( b \)-jets, an isolated lepton, and missing transverse energy due to an escaping neutrino. The di-leptonic decay leads to two \( b \)-jets, possibly with additional light jets, two isolated and oppositely charged leptons, and missing transverse energy.

\[
\nu_\ell \rightarrow E_T \quad \ell^+ \\
W^+ \\
b \\
t \quad \tilde{t} \\
\tilde{b} \\
w \\
q \quad \bar{q}' \\
jet \\
jet \\
jet \\
jet
\]

Figure 5.1: Event topology of semi-leptonically decaying top quark pair: \( t\bar{t} \rightarrow b\ell^+\nu_\ell\tilde{b}q\bar{q}' \). The typical signatures of these kind of events are four jets of which two \( b \)-quark jets, missing transverse energy from the neutrino, and a single isolated lepton.

Signal

Because the total cross section for QCD multi-jet production, \( \mathcal{O}(55 \text{ mb}) \), is many orders of magnitude larger than for \( t\bar{t} \) production, \( \mathcal{O}(900 \text{ pb}) \), the fully hadronic \( t\bar{t} \) signal is practically indistinguishable from this background. The semi- and di-leptonic channels benefit from the presence of isolated leptons and a considerable amount of missing transverse energy. The semi-leptonic channel has two advantages with respect to the di-leptonic channel. First, only one (high-\( p_T \)) neutrino appears in semi-leptonic \( t\bar{t} \) decay and therefore it is possible to kinematically reconstruct the individual top quarks in an event. And second, the semi-leptonic \( t\bar{t} \) branching ratio is four times larger, hence a greater amount of events will decay via this channel.

It should be noted however that \( \tau \) leptons from \( W \rightarrow \tau \nu_\tau \) decay (1/3\(^{\text{rd}}\) of the leptonic \( W \) boson decays) do not always produce clear signals in the detector like electrons and
5.1. Event topology

muons do. The reason for this is that, with an average lifetime of $290.6 \pm 1.0$ fs ($c\tau \approx 87 \mu$m), $\tau$ leptons decay before they reach the detector. This decay is mainly hadronically and, without explicit $\tau$-identification, the $\tau$ channel is therefore not considered as part of the semi-leptonic $t\bar{t}$ signal in this analysis. Indirectly, the $\tau$ lepton still contributes to the semi-leptonic $t\bar{t}$ signal, since in $17.85 \pm 0.05\%$ of the cases it decays to an electron and in $17.36 \pm 0.05\%$ of the cases to a muon [33].

Background

Without using $b$-tagging capability, the main background to the semi-leptonic $t\bar{t}$ signal is $W$ boson production in association with extra jets from gluon radiation ($W +$ jets). As shown in Figure 5.2, the $W + 4$ jets process, with the $W$ boson decaying leptonically, has the same event topology as semi-leptonic $t\bar{t}$ and is therefore called irreducible. The cross section for this process is $\mathcal{O}(250 \text{ pb})$ and is thus comparable in size with that of the $t\bar{t}$ cross section.

![Figure 5.2: Examples of irreducible background to the semi-leptonic $t\bar{t}$ signal: $W +$ jets (left) and QCD multi-jet (right)](image)

Although QCD multi-jet events do not contain leptons and neutrinos in the hard scattering, they still contribute to the background of the semi-leptonic $t\bar{t}$ analysis. For example, as shown in the right illustration of Figure 5.2, a jet might be misidentified as an electron. This misidentification results in a ‘fake’ lepton and possibly also in missing transverse energy, giving the same signature as semi-leptonic $t\bar{t}$. The probability of this misidentification is small, but the cross section for QCD multi-jet production is large. Careful evaluation of this background using simulation, including full detector response, is however problematic due to the limitation in computing resources.

Estimating QCD multi-jet background

In the absence of a sample of Monte Carlo simulated QCD multi-jet events including full detector response simulation, the rate of extra electrons and muons has been determined.

$^2$The cross section for $W(\to e, \mu) + 4$ jets is calculated with ALPGEN at tree level and depends heavily on the scale choice (here: $\mu_F^2 = \mu_R^2 = m_W^2 + \sum_{\text{jets}} p_T^2$) and the jet definition (here: $p_T > 20$ GeV, $|\eta| < 2.5$ and $\Delta R = 0.4$).
Chapter 5. Cross section at $\sqrt{s} = 14$ TeV

from simulated $t\bar{t}$ events as function of jet multiplicity, jet transverse momentum, and jet pseudo-rapidity [160]. It was found that the overall rate of extra electrons is roughly $22 \cdot 10^{-5}$ per jet and the overall rate of extra muons is roughly $70 \cdot 10^{-5}$ per jet in semi-leptonic $t\bar{t}$ events. The main sources for extra electrons are non-prompt electrons from hadronic decays inside $b$-jets and fake electrons from misidentified hadrons inside light jets. Extra muons are mainly non-prompt muons originating from hadronic decays inside $b$-jets. Extrapolation to the QCD di-jet topology and comparison with simulated events indicates this method is promising for a data-driven QCD multi-jet estimation\(^3\). Together with the conservative estimate for the amount of background to the $t\bar{t}$ signal in Section 5.6.7 it is assumed that this background is small.

5.2 Object definitions

The physics objects of interest for the selection and reconstruction of semi-leptonic $t\bar{t}$ events are electrons, muons, jets, and missing transverse energy. The more complex objects, $b$-jets and $\tau$-leptons, are disregarded since heavy flavour tagging is not used in the analysis.

Electrons

Electrons are identified and reconstructed by the egamma algorithm in the calorimeter and inner detector with the following criteria: the $p_T > 20$ GeV and $|\eta| < 2.5$. Electrons in the crack region $1.37 < |\eta| < 1.52$ of the electromagnetic calorimeter are left out. Hits in the transition radiation tracker are not mandatory and there is no requirement on $E/p$ for these ‘medium’ electrons ($isEM=0x3FF$). Isolation is required by setting a limit on the allowed maximum transverse energy in a cone of size $R = 0.20$ around the electron: $E_{T,\Delta R=0.20} < 6.0$ GeV.

Muons

Muons are identified and reconstructed in the muon spectrometer and inner detector with the STACO algorithm and have a $p_T > 20$ GeV and $|\eta| < 2.5$. Like electrons, the muons should fulfill the isolation requirement of $E_{T,\Delta R=0.20} < 6.0$ GeV. To reduce the number of falsely identified muons an additional cut $\text{bestMatch}=1$ is set, which ensures that a muon track is a combination of an inner detector track and a muon spectrometer track.

Jets

Jets are reconstructed with the ATLAS cone algorithm, using cone size $R = 0.4$, and calibrated with the H1 method. The seeds for the algorithm used are cell-based towers from the calorimeter. Jets are required to have $p_T > 20$ GeV and $|\eta| < 2.5$. Furthermore,

\(^3\)The overall expected rate for QCD di-jets is $45 \cdot 10^{-5}$ ($10 \cdot 10^{-5}$) per jet for additional electrons (muons), but the trigger and $b\bar{b} + n$-jets are not taken into account here.
5.3. Event selection

jets that have a ‘good’ electron (as defined above) in the vicinity $\Delta R_{j-e} < 0.2$ are removed, because the jet is most likely misreconstructed. For jets with a ‘good’ muon close to it, this so-called ‘overlap removal’ is not done. Note that in this case, there is a probability that non-prompt muons originating from decay inside $b$-jets are reconstructed as isolated muons.

**Missing transverse energy**

The missing transverse energy $\not{E}_T$ is determined from three components: the sum of transverse energy deposited in the topological cluster cells of the calorimeter, the correction for energy loss in the cryostat, and the contribution from muons. It includes refined calibration via the association of calorimeter cells with identified high-$p_T$ objects (electrons, photons, muons, and jets).

5.3 Event selection

The selection of semi-leptonic $t\bar{t}$ candidates from collision data will be carried out in two steps. First the trigger system selects events with an isolated lepton from raw data. Events that pass this trigger requirements are processed by offline reconstruction. Then, a set of criteria, based on the offline reconstructed physics objects is applied to these events.

Because the isolated lepton in semi-leptonic $t\bar{t}$ used for triggering and event selection is either an electron or a muon, the signal is divided into an ‘electron’ channel and a ‘muon’ channel, referred to as $t\bar{t}(e)$ and $t\bar{t}(\mu)$ respectively. The two channels are combined at a later stage to reduce the statistical uncertainty on the cross section measurement.

5.3.1 Trigger

For completeness, a brief description of the triggers involved in the $t\bar{t}$ analysis is given here. The trigger selecting electron candidate events is called e25i and the trigger selecting muon candidate events is called mu20i. The electron trigger requires an isolated electron with a transverse energy of at least 22.5 GeV, whereas the muon trigger requires a muon with a transverse momentum of more than 20 GeV. For the muon trigger, the isolation requirement is not yet implemented.

5.3.2 Preselection

The default ‘commissioning analysis’ cuts for semi-leptonic $t\bar{t}$ are:

- exactly one lepton (either electron or muon)

- $\not{E}_T > 20$ GeV

---

4The cryostat is located between the LAr barrel electromagnetic calorimeter and the tile barrel hadronic calorimeter and is about half an interaction length where hadronic showers can lose energy.
Chapter 5. Cross section at $\sqrt{s} = 14$ TeV

- at least four jets with $p_T > 20$ GeV
- three or more jets with $p_T > 40$ GeV

with the lepton, $E_T$, and jets as defined in Section 5.2.

### 5.3.3 Selection efficiencies

In Table 5.1 the selection and trigger efficiencies have been determined for semi-leptonic $t\bar{t}$ in the electron and muon channel from simulated events. The list of all Monte Carlo samples used can be found in Appendix A.2. The efficiency for each requirement is defined as the ratio of the number of events that pass the requirement versus the total number of events in that decay channel. The decay channel is determined from the Monte Carlo truth list. Each channel contains approximately 110k events, corresponding to an integrated luminosity of 1 fb$^{-1}$. The statistical uncertainties$^5$ on the efficiencies are therefore below 0.2%.

<table>
<thead>
<tr>
<th>sample</th>
<th>$\ell_{iso}$</th>
<th>$E_T$</th>
<th>4j20</th>
<th>3j40</th>
<th>sel</th>
<th>trig</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}(e)$</td>
<td>52.0%</td>
<td>91.0%</td>
<td>70.7%</td>
<td>61.9%</td>
<td>22.1%</td>
<td>52.9%</td>
<td>18.2%</td>
</tr>
<tr>
<td>$t\bar{t}(\mu)$</td>
<td>68.7%</td>
<td>91.6%</td>
<td>65.5%</td>
<td>57.3%</td>
<td>29.8%</td>
<td>59.9%</td>
<td>23.6%</td>
</tr>
</tbody>
</table>

Table 5.1: Selection and trigger efficiencies for $t\bar{t}(e)$ and $t\bar{t}(\mu)$ events.

The label “$\ell_{iso}$” refers to the selection efficiency of an isolated electron (muon) in the electron (muon) channel. “$E_T$” corresponds to the missing transverse energy requirement, and “4j20” and “3j40” to the four and three jet requirements, respectively. The combined preselection efficiency is given by “sel” and the trigger efficiency by “trig”. The total combined efficiency is given in the last column “total”.

The difference in lepton selection efficiencies $\ell_{iso}$ indicate that the reconstruction efficiency for muons is significantly larger than for electrons. At the same time, the efficiencies for the two jet selection criteria are higher in the electron channel than in the muon channel. This implies that the electrons which are missed in the electron reconstruction end up as reconstructed jets, thereby passing more easily the jet criteria. The small difference between $E_T$ efficiencies shows that the $E_T$ reconstruction performs stable and is not very sensitive to these misidentifications of electrons. The combined selection efficiencies of 22.1% and 29.8% are considerably lower than the individual efficiencies and this indicates that the four offline criteria are almost completely uncorrelated (in the fully uncorrelated case they would be 20.7% and 23.6% respectively). Like for the lepton requirement, the muon trigger efficiency is with 59.9% higher than the electron trigger efficiency of 52.9%. The main reasons for this are the difference in the $p_T$-thresholds of 20 GeV versus 22.5 GeV respectively and the higher efficiency plateau of the muon trigger at large $p_T$. Although the lepton trigger efficiencies are supposed to be highly

$^5$Approximating the uncertainty with $\Delta \epsilon_{sel} = \sqrt{\epsilon_{sel} \cdot (1 - \epsilon_{sel}) / N_{tot}}$ and $\epsilon_{sel} = N_{sel} / N_{tot}$. 

86
correlated with the offline lepton reconstruction efficiencies, due to the differences in the \( p_T \) thresholds and the acceptances of the triggers and the offline reconstruction, the combined selection efficiencies are with 18.2% and 23.6% a few percent lower than without the trigger requirements.

### 5.4 Top reconstruction

There are several strategies to test whether the events that passed the trigger and preselection criteria for semi-leptonic \( t\bar{t} \) candidates, do indeed contain top quarks. This measurement attempts to reconstruct for each event the 'hadronic' top quark of the \( t\bar{t} \) pair. The hadronic top is the top quark that decays into three jets. As shown in Figure 5.1 with \( t \rightarrow bq\bar{q}' \), one jet stems from the \( b \)-quark and the other two jets from the hadronically decaying \( W \) boson. Thus, by combining the four-momenta of these three jets, the original hadronic top quark four-momentum could in principle be obtained, and the invariant three-jet mass \( M_{jjj} \) should yield the hadronic top quark mass \( m_t \).

However, experimentally it is not a priori clear which three jets belong to the hadronic top in an event. Furthermore, additional jets may be formed due to gluon radiation, and very boosted top quarks might cause jets to merge with each other, altering the jet multiplicity of the event. To overcome this problem, without the help of \( b \)-tagging, the three jets that are taken to be from the hadronic top quark, are those jets that give the maximum transverse momentum for their combined four-momentum vector.

The main idea behind this three-jet selection is that jets from top decays are, on average, boosted in the same direction as the top quark they originate from. Since the leptonic top quark is produced back-to-back in the transverse plane with the hadronic top, it is most likely that the jet from the leptonic top will not maximise the transverse momentum when combined with two jets of the hadronic top. Also additional jets from radiation will only maximise the transverse momentum of a three-jet combination if it is boosted in the same direction as the hadronic top. However, this additional radiation is mainly soft and collinear, and it is therefore expected that the approach to reconstruct the hadronic top remains valid. To test this hypothesis with simulated \( t\bar{t} \) events, the combined four-momentum of the selected three-jet combination has been compared with the hadronic top quark using Monte Carlo truth. It turns out that in 25.3 \( \pm \) 0.3\% of the cases this approach picks a three-jet combination which is within \( \Delta R < 0.2 \) of the top quark, and is therefore considered the correct combination. Since the average jet multiplicity in semi-leptonic \( t\bar{t} \) events is almost five jets, this is always better than trying to randomly pick the correct three-jet combination.

#### 5.4.1 Hadronic top quark mass

In Figure 5.3 the invariant three-jet mass \( M_{jjj} \) is shown for the selected \( t\bar{t} \) candidate events in both the electron and muon channel. A distinction is made between three contributions: the signal from correct three-jet combinations, combinatorial background from incorrect three-jet combinations, and background from processes other than semi-
leptonic $t\bar{t}$.

![Invariant three-jet mass $M_{jjj}$ for selected $t\bar{t}$ candidate events in the (a) electron channel and (b) muon channel.](image)

**Figure 5.3:** Invariant three-jet mass $M_{jjj}$ for selected $t\bar{t}$ candidate events in the (a) electron channel and (b) muon channel.

Except for the normalisation, the distributions for both channels are alike. The signal manifestly peaks above the background at 160–170 GeV. This mass range is slightly below the input mass of 175 GeV, which could indicate a suboptimal jet energy scale calibration (a distorted detector geometry was used in the detector response simulation). The shape of the combinatorial background and background from other processes are very similar with respect to each other. The distributions increase rapidly from 50–150 GeV and decrease smoothly from 170 GeV onwards, with long tails towards high masses.

**$W^\pm$ boson mass constraint**

Two of the three jets in a hadronic top decay come from a $W^\pm$ boson. This characteristic is used to purify a sample of semi-leptonic $t\bar{t}$ candidates. Out of the three possible di-jet combinations which can be made from the three-jet combination with the largest transverse momentum, at least one is required to have an invariant mass close to the $W^\pm$ boson mass: $|M_{jj} - M_W| < 10$ GeV.

In Figure 5.4 the distribution of di-jet invariant masses $M_{jj}$ is shown for signal (light shaded) and background events (dark shaded) that passed the trigger and preselection criteria. The dashed lines indicate the limits of the $W^\pm$ boson mass constraint. Note that each events contributes three entries to the histogram, since there are three di-jet permutations per event. This results in a large combinatorial background. The $W^\pm$ boson mass peak is therefore not as pronounced as the hadronic top quark mass.

As shown in Figure 5.5, the invariant three-jet mass $M_{jjj}$ distribution for $t\bar{t}$ candidate events after applying the $M_{W}$-constraint is more pronounced. The peak is sharper and the background is much reduced.

In Table 5.2 the expected number of events that will pass the trigger and preselection criteria, together with the expected number of events that satisfy the $W^\pm$ boson mass constraint in addition are given. The $t\bar{t}$ signal is split up into the three decay channels,
and the semi-leptonic decay is further subdivided per lepton flavour. The single top events comprises the s, t, and Wt production channel. Of the processes containing a single $W^\pm$ or Z boson in association with jets, only the signal from leptonically decaying bosons is taken into account. In the diboson processes at least one boson of the two bosons decays leptonically.

The events with label ‘Signal’ are the $t\bar{t}(e)$ events for the electron channel and $t\bar{t}(\mu)$ events for the muon channel. The signal events are subdivided in ‘Peak’ and ‘Combinatorial’ events. The former corresponds to events with the correct three-jet combination, the latter with the wrong combination. The rest of the events, including $t\bar{t}$ decaying otherwise than the channel of interest, is considered as ‘Background’.

The amount of events in the muon channel is larger than in the electron channel. This is mainly caused by the fact that both the muon trigger efficiency and the muon reconstruction efficiency are higher than for electrons. However, it should also be noted that there is no overlap removal performed between muons and jets, like is done between electrons and jets. This leads to larger selection efficiencies in the muon channel espe-
Chapter 5. Cross section at $\sqrt{s} = 14$ TeV

<table>
<thead>
<tr>
<th>process</th>
<th>electron channel</th>
<th></th>
<th>muon channel</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}(e)$</td>
<td>2185.0</td>
<td>1102.9</td>
<td>7.2</td>
<td>3.8</td>
</tr>
<tr>
<td>$t\bar{t}(\mu)$</td>
<td>2.7</td>
<td>1.1</td>
<td>2833.1</td>
<td>1435.2</td>
</tr>
<tr>
<td>$t\bar{t}(\tau)$</td>
<td>167.1</td>
<td>86.2</td>
<td>242.3</td>
<td>118.0</td>
</tr>
<tr>
<td>$t\bar{t}(\ell^+\ell^-)$</td>
<td>268.0</td>
<td>87.8</td>
<td>344.4</td>
<td>119.4</td>
</tr>
<tr>
<td>$t\bar{t}(jets)$</td>
<td>10.2</td>
<td>3.1</td>
<td>40.4</td>
<td>22.5</td>
</tr>
<tr>
<td>single $t$</td>
<td>183.3</td>
<td>72.9</td>
<td>229.6</td>
<td>87.1</td>
</tr>
<tr>
<td>$W + jets$</td>
<td>741.8</td>
<td>238.6</td>
<td>1040.8</td>
<td>317.7</td>
</tr>
<tr>
<td>$Wb\bar{b} + jets$</td>
<td>45.5</td>
<td>15.1</td>
<td>58.9</td>
<td>17.2</td>
</tr>
<tr>
<td>$Z + jets$</td>
<td>113.2</td>
<td>34.2</td>
<td>78.6</td>
<td>24.0</td>
</tr>
<tr>
<td>$W^+W^-$</td>
<td>7.3</td>
<td>3.7</td>
<td>8.6</td>
<td>4.3</td>
</tr>
<tr>
<td>$WZ$</td>
<td>4.1</td>
<td>1.6</td>
<td>5.5</td>
<td>2.2</td>
</tr>
<tr>
<td>$ZZ$</td>
<td>0.7</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>Signal</td>
<td>2185.0</td>
<td>1102.9</td>
<td>2833.1</td>
<td>1435.2</td>
</tr>
<tr>
<td>– Peak</td>
<td>551.2</td>
<td>389.2</td>
<td>717.4</td>
<td>510.8</td>
</tr>
<tr>
<td>– Combinatorial</td>
<td>1633.8</td>
<td>713.7</td>
<td>2115.8</td>
<td>924.4</td>
</tr>
<tr>
<td>Background</td>
<td>1543.9</td>
<td>544.3</td>
<td>2056.8</td>
<td>716.4</td>
</tr>
<tr>
<td>$S/B$</td>
<td>1.4</td>
<td>2.0</td>
<td>1.4</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 5.2: Expected number of events passing the trigger and preselection criteria and the $M_W$-constraint for various processes and an integrated luminosity of 100 pb$^{-1}$.

pecially for events with $b$-jets, where muons are produced inside a jet via $b$-quark decay. The difference of factor four between the $t\bar{t}(jets)$ contribution to the muon channel and to the electron channel illustrates this.

In any case, for both channels the largest background comes from $W + jets$ and di-leptonic $t\bar{t}$. The reason for the latter background to pass the selection criteria is that often one of the two leptons is outside the detector acceptance and is not reconstructed. This is also the case for $Z + jets$. Single top, in particular produced via the $Wt$ channel, and $t\bar{t}(\tau)$ do contain a hadronic top and will form part of the hadronic top signal. The diboson processes do not contribute significantly due to their small cross sections.

The reduction in the number of events that pass the selection after the additional $M_W$-constraint depends on the presence of a hadronic top quark in the event. Background from $W + jets$, $Z + jets$, and $t\bar{t}(\ell^+\ell^-)$ is reduced to roughly $1/3^{rd}$, while signal from semi-leptonic $t\bar{t}$, $t\bar{t}(jets)$, and single top are only reduced by a factor $1/2$. The
$M_W$-constraint increases the signal over background ratio $S/B$ from 1.4 to 2.0 for both the electron and muon channel.

## 5.5 Cross section determination

The $t\bar{t}$ cross section $\sigma(t\bar{t})$ is measured by counting the number of events in the peak of the invariant three-jet mass distribution $N_{\text{peak}}$ and using:

$$\sigma(t\bar{t}) \times Br(t\bar{t} \to b\bar{b}qq'\ell\bar{\nu}_\ell) = \frac{1}{f_\ell} \cdot \frac{N_{\text{peak}}}{\int L \, dt} \quad (5.1)$$

The number of events in the peak are obtained by performing a maximum likelihood fit with a combination of a Gaussian and a 4th order Chebyshev polynomial to the $M_{jjj}$-distribution. The former describes the hadronic top mass signal and the latter both combinatorial background and background from other processes, as illustrated in Figure 5.6. Note that the experimental resolution is much larger than the 2.1 GeV width of the top quark mass [33], hence a Gaussian distribution should be suitable for describing the measured shape of the top quark mass peak. The fraction $f_\ell$ relates the number of events in the peak with the $t\bar{t}$ cross section. It corrects for the detector acceptance, the trigger and selection efficiencies, plus the hadronic top quark reconstruction efficiency. The value of $f_\ell$ is determined for both the electron ($f_e$) and the muon channel ($f_\mu$) from Monte Carlo simulation as will be explained in the following section. Finally, $Br(t\bar{t} \to b\bar{b}qq'\ell\nu)$ is the semi-leptonic $t\bar{t}$ branching ratio, and $\int L \, dt$ is the integrated luminosity.

![Figure 5.6](image_url)

**Figure 5.6:** (a) Example of a likelihood fit to the $M_{jjj}$-distribution using simulated muon channel data equivalent to 100 pb$^{-1}$ and (b) expected significance of the likelihood fit as function of the luminosity assuming the nominal and twice the nominal amount of background.
Chapter 5. Cross section at $\sqrt{s} = 14$ TeV

Determination of $f_\ell$

The fraction $f_\ell$ in Eq.(5.1) is determined by taking the ratio of the number of signal $t\bar{t}$ events ($N_{\text{peak}}$) and the total number of semi-leptonic $t\bar{t}$ events. But, instead of extracting $N_{\text{peak}}$ only once by using a maximum likelihood fit to the $M_{jjj}$-distribution from simulated events, the procedure is repeated 100,000 times with so-called pseudo experiments. Each such experiment represents a measurement of the $M_{jjj}$-distribution based on ‘data’ with an integrated luminosity equivalent to 100 pb$^{-1}$. The data in this case consists of data points from the original full Monte Carlo simulated sample, to which Poisson fluctuations are introduced corresponding to the statistical uncertainty on 100 pb$^{-1}$ of data. In each pseudo experiment a binned maximum likelihood fit is performed to the $M_{jjj}$-distribution according to:

$$-\ln L = -\ln \prod_{i=1}^{N_{\text{events}}} P(M_{jjj}; (\mu, \sigma), (c_0, c_1, c_2, c_3, c_4), N_{\text{sig}}, N_{\text{bkg}})$$ (5.2)

The parameters of the fit are the mean and the width of the Gaussian distribution (signal), the four constants of the Chebyshev polynomial (background), and the normalisation of these two probability density functions giving the number of signal and background events respectively.

Assuming a $t\bar{t}$ production cross section of 833 pb and a branching ratio of 10.8% for leptonic $W$ boson decay, the total number of expected $t\bar{t}(e)$ and $t\bar{t}(\mu)$ events in a sample of 100 pb$^{-1}$ integrated luminosity is approximately 12k events each$^6$. The number of events in the $M_{jjj}$-peak expected in the electron channel is $317 \pm 56$ events. This corresponds to a fraction $f_e$ of $2.6 \pm 0.5\%$. For the muon channel, the expected number of events in the peak is $494 \pm 68$ events, corresponding to a fraction $f_\mu$ of $4.1 \pm 0.6\%$. The quoted errors on the expected number of events and efficiencies include both statistical and systematic uncertainties associated with the fit. Fit uncertainties will be discussed further in Section 5.6.5.

5.6 Statistical and systematic uncertainties

In Table 5.3 the statistical and systematic uncertainties on the cross section measurement are given for both channels.

5.6.1 Luminosity

The luminosity of the LHC will be determined independently from this measurement. The uncertainty on the luminosity measurement affects the cross section measurement directly since $N_{\text{peak}} \propto L$ in Eq.(5.1).

At the LHC start-up only a rough measurement of the machine parameters will be available. The expected uncertainty on the luminosity during this phase will be of the

$^6$The normalisation depends on the cross section, the crucial quantity is therefore the fraction $f_\ell$. 

92
Table 5.3: Uncertainties on the cross-section measurement (in %).

<table>
<thead>
<tr>
<th>Source</th>
<th>Channel</th>
<th>electron</th>
<th>muon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td></td>
<td>10.5</td>
<td>8.0</td>
</tr>
<tr>
<td>Luminosity</td>
<td></td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Pile-up</td>
<td></td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>PDFs</td>
<td></td>
<td>2.5</td>
<td>2.2</td>
</tr>
<tr>
<td>Lepton ID efficiency</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Lepton trigger efficiency</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Jet Energy Scale (5%)</td>
<td></td>
<td>2.3</td>
<td>0.9</td>
</tr>
<tr>
<td>ISR/FSR</td>
<td></td>
<td>8.9</td>
<td>8.9</td>
</tr>
<tr>
<td>Shape of fit function</td>
<td></td>
<td>14.0</td>
<td>10.4</td>
</tr>
<tr>
<td>50% more W + jets</td>
<td></td>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>20% more W + jets</td>
<td></td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

order of 20–30%. A better determination of the beam profiles using specials runs of the machine will ultimately lead to a systematic uncertainty of the order of 3–5%.

Effects from pile-up (multiple hard scatterings per bunch crossing) are assumed to be negligible in the initial phase of the LHC running at low luminosity and the pile-up effects are therefore not investigated in this analysis.

### 5.6.2 Parton density functions

The parton distribution functions (PDF’s) used in the simulation\(^7\) have systematic uncertainties in their parametrisations. These uncertainties enter in the \( t\bar{t} \) cross section measurement via the factorisation principle in Eq.(1.2). The PDF uncertainties have been determined for the \( t\bar{t} \) signal using fast detector simulation with (and without) selection cuts applied.

The PDF set CTEQ6M provides NLO accuracy and allows to estimate the PDF uncertainty via 40 ‘error’ sets. Each error set corresponds to either an upward variation, \( S^+_i \), or a downward variation, \( S^-_i \), of one of the 20 parameters, \( i \), with respect to its central value (by an amount of \( \Delta \chi^2 = 100 \)). In order to reduce the computational time, the \( t\bar{t} \) signal has not been resimulated for each set. Instead, events have been generated once (with the central set) and reweighted for each error set on an event basis according to:

\[
w_{\text{evt},i}^\pm = \frac{S_i^\pm(x_1,Q^2) \cdot S_i^\pm(x_2,Q^2)}{S_0(x_1,Q^2) \cdot S_0(x_2,Q^2)}
\]

where \( S_0 \) is the central PDF set, \( S^\pm_i \) one of the forty error sets, and \( x_1, x_2, \) and \( Q^2 \) the momentum fractions and energy scale of the event. The systematic uncertainty on the

\(^7\)CTEQ6M for the \( t\bar{t} \) signal (MC@NLO) and CTEQ6L1 for all background processes.
cross section measurement arising from the PDF uncertainty is then defined as:

$$\Delta \sigma = \frac{1}{2} \sqrt{\sum_{i=1}^{20} \left[ \sigma(S_i^+) - \sigma(S_i^-) \right]^2}$$

and is estimated to be 2.5% for the $t\bar{t}(e)$ signal and 2.2% for the $t\bar{t}(\mu)$ signal. A similar study using the MRST2001 PDF sets indicates an uncertainty of 1.0% and 0.8% respectively. The uncertainty is less than for the CTEQ PDF due to the smaller variation ($\Delta \chi^2 = 50$) on the 15 parameters used for the error estimation in the MRST sets. The difference between the central values of the two different PDF sets with respect to the default CTEQ PDF indicates an uncertainty due to the PDF modelling of 2.1% and 1.8% respectively.

Note that there are some subtle issues with the reweighting procedure. First of all, the procedure is an approximation because it does not take into account the modification of the Sudakov form factors in the parton shower and the impact on the underlying event which both depend on the PDF (Chapter 2). Secondly, factorisation and renormalisation scale dependencies have not been included. And last, the PDF uncertainties on background estimation from other processes have not been investigated. These issues have been partly addressed in an earlier study on PDF uncertainties [161], however without taking into account the $t\bar{t}$ selection criteria. The selection criteria are known to reduce the PDF uncertainties [162].

### 5.6.3 Lepton identification and trigger efficiencies

The lepton identification and trigger efficiencies for electrons and muons will be determined from collision data with $Z$ boson events. For an integrated luminosity of 100 pb$^{-1}$, the uncertainties on the lepton identification and trigger efficiencies are expected to be less than 1%.

### 5.6.4 Initial and final state radiation

Because $t\bar{t}$ candidate events are selected by requiring a minimum number of jets, the selection efficiency depends on the jet multiplicity in an event. The jet multiplicity in its turn depends on the amount of radiation in an event. The uncertainty on initial and final state radiation predictions from the Monte Carlo generator translates therefore directly into an uncertainty on the selection efficiency. To assess the systematic uncertainty from this source on the cross section measurement, the ACERMC generator was used together with the parton showering from PYTHIA. The parameters for initial state radiation and final state radiation have each been varied with a factor $\frac{1}{2}$ to 2. The uncertainty is determined to be 8.9% for both electron and muon channel.

This way of estimating the uncertainty is however rather conservative. As discussed in Chapter 2, more accurate Monte Carlo techniques are available to estimate the jet multiplicity in $t\bar{t}$ events. Besides, the uncertainty on the jet multiplicity also affects the
5.6. Statistical and systematic uncertainties

hadronic top reconstruction efficiency, which is not included here. In the next chapter, this systematic uncertainty is therefore investigated in more depth.

5.6.5 Fit uncertainties

The uncertainties associated with the fitting procedure concern two sources. On one hand, the amount of signal and background events and thus the normalisation of the fit functions may vary. On the other hand, the shape of the invariant three-jet mass distribution may change due the fluctuation of data points. The former is considered the statistical uncertainty, whereas the latter is regarded the systematic uncertainty due to the shape of the fit function used.

To estimate the sensitivity of the likelihood fit to the two types of uncertainties, the $M_{jjj}$ distribution is fitted in 100,000 pseudo experiments of each 100 pb$^{-1}$ using all 8 parameters in Eq.(5.2) as free parameters. The spread in the number of fitted signal events gives the combined statistical and systematic uncertainty of 17.5% and 13.1% on the fractions $f_e$ and $f_\mu$ for the $t\bar{t}(e)$ and $t\bar{t}(\mu)$ signals respectively. Then, the fits are repeated, but now with the 6 parameters describing the shapes of the Gaussian ($\mu$ and $\sigma$) and the Chebyshev polynomial ($c_0$, $c_1$, $c_2$, $c_3$, and $c_4$) fixed to the values obtained from fitting the original full Monte Carlo sample. Hence, in this second round of pseudo experiments only $N_{\text{sig}}$ and $N_{\text{bkg}}$ are free fit parameters. Now only the statistical uncertainty follows from the width of the distribution of fitted signal events. The statistical uncertainty is determined to be 10.5% for the electron channel and 8.0% for the muon channel. The systematic uncertainty is then finally derived from the combined uncertainty by subtracting the statistical uncertainty (quadratically) and yields 14.0% (electron) and 10.4% (muon).

5.6.6 Jet energy scale

As discussed in Section 3.2.3, most likely the jet energy scale is not immediately optimal in the commissioning phase. The jet energy scale affects the jet selection efficiencies, because it influences the transverse momenta of jets, and is therefore a source of the systematic uncertainty on the cross section measurement. To see how much impact an offset in the jet energy scale has on the measurement, the transverse momenta (and energy) of all the jets present in signal and background simulation have been varied with $-5\%$, $-2\%$, $+2\%$ and $+5\%$. The changes in event selection efficiency $\epsilon_{\text{sel}}$, the hadronic top reconstruction efficiency $\epsilon_{\text{rec}}$, and the combined efficiency $\epsilon_{\text{tot}}$ due to an increase or decrease in jet energy scale have been determined. The numbers are given in Table 5.4.

Not surprisingly, the results are similar for both lepton channels, and they are visualised in Figure 5.7 for the electron channel. An increase (decrease) in jet energy scale leads to an increase (decrease) in selection efficiencies, because more (less) jets pass the two jet criteria. The hadronic top reconstruction efficiency however has the opposite behaviour: it decreases with increasing jet energy scale and the other way round. The reason is, that with a larger jet energy scale, more jets pass the selection criteria, and therefore also the number of possible three-jet combinations. The net result is that the
Chapter 5. Cross section at $\sqrt{s} = 14$ TeV

<table>
<thead>
<tr>
<th>$\Delta$JES (%)</th>
<th>electron channel</th>
<th>muon channel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\epsilon_{\text{sel}}$</td>
<td>$\epsilon_{\text{rec}}$</td>
</tr>
<tr>
<td>$-5.0$</td>
<td>16.8</td>
<td>18.6</td>
</tr>
<tr>
<td>$-2.0$</td>
<td>17.6</td>
<td>18.1</td>
</tr>
<tr>
<td>$0.0$</td>
<td>18.2</td>
<td>17.8</td>
</tr>
<tr>
<td>$+2.0$</td>
<td>18.8</td>
<td>17.4</td>
</tr>
<tr>
<td>$+5.0$</td>
<td>19.6</td>
<td>16.6</td>
</tr>
</tbody>
</table>

Table 5.4: Change in the selection efficiency $\epsilon_{\text{sel}}$, the reconstruction efficiency $\epsilon_{\text{rec}}$, and the combined efficiency $\epsilon_{\text{tot}}$ (in %) for the $t\bar{t}$ signal in the two lepton channels due to an increase or decrease in jet energy scale (JES).

The combined efficiency ($\epsilon_{\text{tot}} = \epsilon_{\text{sel}} \times \epsilon_{\text{rec}}$) is not as sensitive to the jet energy scale uncertainty as the two efficiencies separately. The systematic jet energy scale uncertainty is estimated to be 2.3% for the electron channel and 0.9% for the muon channel. These numbers are consistent with each other within their statistical uncertainties (which are rather large due to the limited amount of Monte Carlo statistics).

Figure 5.7: Change in the selection efficiency $\epsilon_{\text{sel}}$, the reconstruction efficiency $\epsilon_{\text{rec}}$, and the combined efficiency $\epsilon_{\text{tot}}$ for the $t\bar{t}(e)$ signal due to an increase or decrease in jet energy scale (JES).

There are a few comments to be made to this observation. The $W$ boson mass constraint $|M_{jj} - M_W| < 10$ GeV is kept fixed at $M_W$ while varying the jet energy scale. From Figure 5.4 it can be understood that a shift of the $W$ boson mass peak in the $M_{jj}$ distribution due to jet energy scale variation, leads to a loss of signal yield. In practice, the central value of the window will be adjusted to the $W$ boson mass peak in the $M_{jj}$ distribution from data. Furthermore, in the jet energy scale variation, the effect on the reconstruction of transverse missing energy, the $E_T$-scale, has not been taken into account.

Note that the relation between jet energy scale and $E_T$ scale is not trivial; $E_T$ and jets are reconstructed from different calorimeter input signals (topological cell clusters and towers), there are several
5.6.7 Amount of background

Monte Carlo predictions for the various backgrounds to the $t\bar{t}$ signal from processes with multiple hard jets in the final state have large uncertainties. Varying the matching parameters alone already indicates an uncertainty of order 50% on the $W$ + jets contribution in ALPGEN [159]. Furthermore, in the absence of a proper estimate of QCD multi-jet background one should reckon that the total amount of background to the $t\bar{t}$ signal could differ substantially from what is expected.

The systematic uncertainty on the cross section measurement due to a possible background underestimation has been studied. The likelihood fit was carried out using background from non-$t\bar{t}$ processes and $t\bar{t}$ combinatorics increased by 20% and 50%. Because the combinatorial $t\bar{t}$ background, the largest background to the $t\bar{t}$ signal, also increases, this is a safe estimate for to the uncertainties on the contributions from various multi-jet backgrounds as long as the $M_{jjj}$-distributions of these sources have a similar shape. As a matter of fact, more combinatorial $t\bar{t}$ background with respect to the $t\bar{t}$ signal would indicate a larger than expected average jet multiplicity.

The cross section measurement turns out to be quite insensitive to the background normalisation. An increase of the background of 20% leads to a change in the cross section of only of 0.3% in either lepton channel, while an increase of 50% results in a 1.1% (electron) and 0.8% (muon) larger cross section. The increase indirectly results in a somewhat larger statistical uncertainty, leading to a deterioration of the signal over background ratio $S/B$.

5.7 Results and conclusions

In this chapter, a maximum likelihood fit is presented to measure the semi-leptonic top quark pair production cross section. The most important uncertainties have been identified and addressed. By combining the results from the electron channel and muon channel, a higher accuracy is obtained than individually. The uncertainty on the cross section measurement using a data sample of 100 pb$^{-1}$ is estimated to be:

$$\frac{\Delta \sigma}{\sigma} = \pm 7\% \text{ (stat)} \pm 15\% \text{ (syst)} \pm 3\% \text{ (pdf)} \pm 5\% \text{ (lumi)}$$

With only a small amount of data\textsuperscript{9}, the statistical uncertainty is already less than the systematic uncertainty. The dominant sources for the systematic uncertainty are the shape of the fit function and the expected jet multiplicities in events (this will be investigated in more depth in the next chapter). The likelihood fit is however very robust against uncertainties in the amount of background events and jet energy scale variations. The total uncertainty on the cross section is almost 18% and the method does not rely on $b$-tagging.

\textsuperscript{9}At design luminosity $L = 10^{34}$ cm$^{-2}$s$^{-1}$, $\mathcal{O}(100 \text{ fb}^{-1})$ of data per year is expected at the LHC.
Chapter 5. Cross section at $\sqrt{s} = 14$ TeV

‘Cut & Count’ method

Another approach to measure the $t\bar{t}$ production cross section is the ‘cut & count’ method. The cross section is extracted by counting the numbers of events that pass the selection criteria and subtracting the number of expected (non-$t\bar{t}$) background events predicted by Monte Carlo simulation:

$$\sigma(t\bar{t}) = \frac{N_{\text{obs}} - N_{\text{bkg}}}{\int Ldt \times \epsilon_{\text{sel}}}$$

with $\epsilon_{\text{sel}}$ the selection efficiency for $t\bar{t}$ events. The accuracy that can be achieved with 100 pb$^{-1}$ of data is estimated to be:

$$\Delta \sigma/\sigma = \pm 3\% \text{ (stat)} \pm 16\% \text{ (syst)} \pm 3\% \text{ (pdf)} \pm 5\% \text{ (lumi)}$$

The statistical uncertainty is smaller than for the fit method, because in this method all $t\bar{t}$ events are considered as signal, while for the fit method only the correctly reconstructed $t\bar{t}$ events in the invariant three-jet mass peak ($N_{\text{peak}}$) are considered as signal. Hence, although both methods start with the same sample of selected events, the cut & count method determines the cross section eventually with a larger amount of $t\bar{t}$ events.

The overall systematic uncertainty is at the same level as the for the fit method. However, the individual components contribute differently. The cut & count method does not depend on a Monte Carlo prediction for the shape of the $M_{jjj}$-distribution (the largest uncertainty for the fit method). Here, the largest uncertainties come from the jet energy scale (9.7% for a 5% variation), the prediction of the background normalisation $N_{\text{bkg}}$ (9.5% for 50% more background), and the expected amount of initial and final state radiation (8.9%). These uncertainties directly affect the event selection efficiency $\epsilon_{\text{sel}}$ and thus the cross section.

Because the cut & count method and the likelihood fit methods have different dominant systematic uncertainties, they are complementary to each other. This is very valuable for cross checking results obtained from collision data.