Measuring the Top Quark Pair Production Cross Section in ATLAS

A Cross Section Analysis on Stream Test Data Using b-Tagging

Louise A. Skinnari

Master's Thesis

Department of Physics
Faculty of Engineering (Lunds Tekniska Högskola)
Lund University

July 2008

CERN-THESIS-2010-025
31/07/2008

Abstract

This thesis presents measurements of the top quark pair production cross section using the ATLAS stream test data. The cross section is determined separately for the semileptonic mode, where events in the electron plus jets channel are isolated, and the dilepton mode, where events in the electron-muon channel are considered. The cross section is determined both with and without using tagging of b-quark jets.

For the semileptonic mode, the measured cross section is $\sigma_{t\bar{t}} = 740 \pm 57$ pb without b-tagging and $\sigma_{t\bar{t}} = 788 \pm 50$ pb when using b-tagging. For the dilepton mode, the cross section is $\sigma_{t\bar{t}} = 655 \pm 73$ pb without b-tagging and $\sigma_{t\bar{t}} = 704 \pm 86$ pb when including b-tagging.

Acknowledgements

The work behind this thesis was carried out at Lawrence Berkeley National Laboratory (LBL) in Berkeley, California, during fall 2007. I want to begin by thanking my supervisor, Marjorie Shapiro, for inviting me to stay for half a year and work at LBL after my exchange studies at Berkeley, and for the continuous help and support during that time.

I also want to thank the postdocs at LBL, especially Ayana Holloway-Arce for her inspiration and for the many hours spent on discussing top quark and b-tagging pecularities, and Andrei Gaponenko for important analysis inputs.

Finally, useful feedback came from the ATLAS PhD students at LBL, in particular Lauren Tompkins who patiently answered my questions when confused by more or less ATLAS related matters.

Contents

1 Introduction **2**

2 Theoretical Overview **3**
2.1 The Standard Model of Particle Physics 3
2.1.1 Quarks and Leptons 3
2.1.2 Fundamental Interactions and Force Carriers 4
2.2 The Top Quark 6
2.2.1 Top Quark Production Mechanisms 6
2.2.2 Top Quark Decay 7
2.3 Cross Section Calculations 9

3 Experimental Apparatus **11**
3.1 The Large Hadron Collider 11
3.1.1 The LHC Accelerator 11
3.1.2 The LHC Experiments 12
3.2 The ATLAS Detector 13
3.2.1 Magnetic System 15
3.2.2 Inner Detector 15
3.2.3 Calorimeters 16
3.2.4 Muon Spectrometer 16
3.2.5 Trigger and Data Acquisition System 17
3.2.6 ATLAS Computing 18

4 Analysis Overview **21**
4.1 Cross Section Measurements Without b-Tagging 21
4.2 Cross Section Measurements Using b-Tagging 23
4.2.1 Method 23
4.3 Reconstructing the Top Mass 25

5 Event Selection **26**
5.1 Stream Test Data 26

5.2 Reconstruction and Identification . . . 27
5.2.1 Electrons and Photons . . . 27
5.2.2 Muons . . . 28
5.2.3 Jets . . . 29
5.2.4 Missing Transverse Energy . . . 30
5.3 Object Definitions . . . 30
5.3.1 Electron Definition . . . 30
5.3.2 Muon Definition . . . 31
5.3.3 Jet Definition . . . 31
5.3.4 Missing Transverse Energy Definition . . . 32
5.4 Event Selection . . . 32
5.4.1 Semileptonic Mode . . . 33
5.4.2 Dilepton Mode . . . 33
5.4.3 Event Selection Efficiencies . . . 34
5.5 Calibrations and Efficiencies . . . 34
5.5.1 Trigger Efficiency . . . 35
5.5.2 Electron Identification Efficiency . . . 35

6 b-Quark Jet Tagging 37
6.1 ATLAS b-Tagging Algorithms . . . 37
6.1.1 Transverse Impact Parameter Tagging . . . 38
6.1.2 3D Impact Parameter Tagging . . . 39
6.1.3 Secondary Vertex Tagging . . . 40
6.1.4 The ATLAS `IP3D+SV1` Tagger . . . 40
6.2 Tagging Efficiencies . . . 41
6.2.1 Jet Fractions . . . 41
6.2.2 Mistag Rate . . . 41
6.2.3 Implementation of Tagging Measurement Method . . . 43

7 Signal Acceptance 44
7.1 Acceptance in the Semileptonic Mode . . . 44
7.2 Acceptance in the Dilepton Mode . . . 46

8 Background Estimation 48
8.1 Single Gauge Bosons . . . 48
8.2 Single Top . . . 49
8.3 Gauge Boson Pairs . . . 50

9 Results 51
9.1 Semileptonic Mode Analysis . . . 51
9.1.1 Cross Section Without b-Tagging . . . 51

9.1.2 Cross Section Using b-Tagging 51
9.2 Dilepton Mode Analysis 53
9.2.1 Cross Section Without b-Tagging 53
9.2.2 Cross Section Using b-Tagging 55
9.2.3 Systematic Uncertainties 56
9.3 Reconstructing the Top Mass 59
9.3.1 Invariant W-Boson Mass 59
9.3.2 Top Quark Mass 59

10 Conclusions **61**

A Monte Carlo Samples **63**

B Summary in Swedish **65**

List of Figures

2.1 Schematic figure showing the three generations of quarks and leptons. The generations differ only by the masses of the particles, with masses increasing to the right. 4
2.2 Leading order Feynman diagrams for $t\bar{t}$ pair production mechanisms in hadron collisions [8]. 7
2.3 Feynamn diagrams for the three single top production processes [9]. . . . 8

3.1 The ATLAS detector [12]. 14
3.2 Schematic figure of the ATLAS trigger levels [4]. 17

4.1 The CDF original top discovery [6]. The boxed regions show the predicted number of events without top quarks, the circles show the measured number of events without b-tagging and the black marks show the measured number of events after b-tagging. The excess in the number of measured events is ascribed to top quarks. 22

5.1 Number of events in the stream test data before each cut for the semileptonic mode. 34
5.2 Number of events in the stream test data before each cut in the dilepton mode. 35

6.1 Definition of the transverse impact parameter and its sign [33]. 38
6.2 Transverse (a) and longitudinal (b) impact parameter distributions for b-jets and u-jets [33]. 39

7.1 Graphical representation of the effect of the selection cuts on the number of events in the MC@NLO $t\bar{t}$ sample for the semileptonic decay mode. . 45
7.2 Graphical representation of the effect of the selection cuts on the number of events in the MC@NLO $t\bar{t}$ sample for the dilepton mode. 47

9.1 Normalized jet multiplicities for stream test data and its different components without b-tagging for semileptonic event candidates. 52

9.2 Jet multiplicity distributions for events passing the preselection cuts (semileptonic events) with zero, at least one, or at least two b-tagged jets. 53
9.3 Normalized jet multiplicities for stream test data and its different components when requiring at least one (a) or at least two (b) b-jets for the semileptonic mode. 54
9.4 Normalized jet multiplicities for stream test data and its different components without b-tagging for fully leptonic event candidates. 56
9.5 Jet multiplicities for events in the dilepton mode without b-tagging and with at least one or at least two b-tagged jets. 57
9.6 Normalized jet multiplicities for stream test data and its different components when requiring at least one (a) or at least two (b) b-jets in the dilepton mode. 57
9.7 Invariant mass of the hadronic W-boson for events with four or five jets of which two are b-tagged. There are three entries per event for events with five jets. 59
9.8 Reconstructed top mass for events with four or five jets of which two are b-tagged. For each hadronic W-boson dijet combination, the three-jet combination (two untagged jets plus one b-jet) resulting in highest sum p_T is chosen. 60

List of Tables

6.1 (True) jet fractions (%) in events with four or five jets for the $t\bar{t}$ sample (semileptonic mode). 42
6.2 (True) jet fractions (%) in events with two, three or four jets for the $t\bar{t}$ sample (dilepton mode). 42

7.1 The acceptance of the event selection (excluding trigger requirements) for MC@NLO $t\bar{t}$ signal events in the semileptonic mode. The errors are statistical only. 45
7.2 The acceptance of the event selection (excluding trigger requirements) for MC@NLO $t\bar{t}$ signal events in the dilepton mode, with errors being statistical only. 46

8.1 Summary of the MC samples used to study the backgrounds to the $t\bar{t}$ decay. 49

9.1 Estimation of the number of $t\bar{t}$ events in the stream test data for the semileptonic mode. The quoted errors are statistical only. 52
9.2 Signal to background ratios and single top suppression for events with zero and one jets, and for events with four or more jets, for the semileptonic mode. The uncertainties are statistical only. 54
9.3 The $t\bar{t}$ cross section and tagging efficiencies calculated for the semileptonic mode for both the looser and tighter weight cuts. The errors are statistical only. 54
9.4 Estimation of the number of $t\bar{t}$ events in the stream test data for the dilepton mode, where the errors are statistical only. 55
9.5 Signal to background ratios and single top suppression for events with zero and one jets, and for events with two or more jets, for the dilepton mode. The uncertainties are statistical only. 58
9.6 The $t\bar{t}$ cross section and tagging efficiencies calculated for the dilepton mode for both the looser and tighter weight cuts. The errors are statistical only. 58

10.1 Summarized results for the $t\bar{t}$ production cross section. 61

Chapter 1

Introduction

High energy physics explores the fundamentals of our universe and aims at answering the basic questions about nature. It studies the tiny elementary particles together with the largest structures of the universe. The current best description of elementary particles and their interactions is encorporated in what is called the Standard Model. The Standard Model has been experimentally verified over and over, however, it has shortcomings. The theory does not include gravity and other recently observed phenomena, such as dark matter and dark energy, remain unexplained. With the upcoming start of the Large Hadron Collider (LHC) at CERN, the field of elementary particle physics is yet again facing new tests. The LHC will explore the high energy frontier and the near future will most certainly deliver exciting new discoveries.

One of the elementary particles in the Standard Model is the top quark, the heaviest of them all. Since it was discovered in 1995, it has been studied extensively to determine its mass, its cross section etc. Due to the high energy reached by the LHC, top quarks will there be produced in large quantities, enabling precision studies of its properties. Furthermore, top quarks will constitute an important background process to many searches for "new physics" and they consequently need to be thoroughly understood.

At the LHC, top quarks will mainly be produced in pairs. The analysis carried out in this thesis studies this top pair production. The cross section, which essentially determines the likelihood of a top pair being produced, is measured. Each produced top quark decays into a W-boson and a b-quark. The cross section is determined both with and without identifying the b-quarks (b-jet tagging). There are different final states for the decaying top quark pair, depending on how the W-bosons decay. In this analysis, the semileptonic mode, where one W decays into a lepton-neutrino pair and the other W decays hadronically to jets, and the dilepton mode, where both Ws decay to a lepton-neutrino pair, are studied. Finally, the top mass is reconstructed from its decay products to verify the top quark hypothesis.

Chapter 2

Theoretical Overview

2.1 The Standard Model of Particle Physics

The Standard Model (SM) is our current best representation of the physics phenomena observed in nature. It is a quantum field theory describing the properties of all known elementary particles and three of the four fundamental interactions. The SM has been empirically verified for decades, however, it is not a complete theory. One reason for its incompleteness is the inability to include gravity. Additionally, there are many unanswered questions about the universe that need explanations. Many possible extensions to the SM have been proposed. These extensions are generically known as "Beyond the Standard Model" (BSM) physics.

The observed elementary particles are divided into three classes: Leptons, quarks and force carriers. Leptons and quarks are spin-1/2 fermions whereas the force carriers are spin-1 gauge bosons. There are four known fundamental forces in our universe: The electromagnetic and weak forces (which together form the electroweak interaction), the strong force, and gravity. The SM includes both the electroweak and the strong interactions, however, as already mentioned, it fails to incorporate the gravitational interaction.

The different constituents of the SM are described below in more detail. First, a brief overview of quarks and leptons is given. Thereafter, the fundamental interactions and force carriers are described. This also includes some discussion of the CKM-matrix, responsible for mixing of quarks, and the Higgs mechanism, believed to be the generator of quark masses.

2.1.1 Quarks and Leptons

Quarks and leptons are spin-1/2 particles. They are divided into three generations (see Figure 2.1), where the generations differ only by the masses of the ingoing particles. Each generation contains one electrically charged lepton (electron e, muon μ, tauon τ),

	I	II	III
Leptons	$\begin{pmatrix} e \\ \nu_e \end{pmatrix}$	$\begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}$	$\begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}$
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}$	$\begin{pmatrix} c \\ s \end{pmatrix}$	$\begin{pmatrix} t \\ b \end{pmatrix}$

Figure 2.1: Schematic figure showing the three generations of quarks and leptons. The generations differ only by the masses of the particles, with masses increasing to the right.

one neutral lepton (neutrinos ν_e, ν_μ, ν_τ), one up-like quark (up u, charm c, top t) and one down-like quark (down d, strange s, bottom b) [1].

Quarks have electric charge which enables them to participate in electromagnetic interactions. The up-type quarks carry an electric charge of $+2/3$ while the down-type quarks have an electric charge of -1/3. Quarks also carry color charge (red, green or blue) and they are hence affected by the strong force. Leptons, on the other hand, are color-neutral and uncapable of interacting via the strong force. Electrons, muons and tauons carry an electric charge of $+1$ and can therefore interact through both the weak and electromagnetic interactions whereas the neutrinos, which are electrically neutral, only interact weakly.

In addition to the elementary particles mentioned above, each particle has its own antiparticle associated with it. An electrically charged particle has an antiparticle with the exact same properties as its corresponding particle except that it has opposite charge and opposite magnetic moment relative to the spin direction [2]. Also the electrically neutral neutrinos have antiparticles. It is still debated, however, whether neutrinos are their own antiparticles (then called *Majorana neutrinos*) or not (*Dirac neutrinos*) [3].

Quarks are the fundamental building blocks of composite particles. The composite particles are divided into baryons (which have net quark number three) and mesons (with net quark number zero). Classical examples of baryons are the proton, with two u-quarks and one d-quark, and the neutron, with one u-quark and two d-quarks. An example of the mesons is the pion, which consists of one u-quark and one $\bar{d}$-quark.

2.1.2 Fundamental Interactions and Force Carriers

In addition to describing the elementary particles observed in nature, the SM also incorporates theories describing the fundamental interactions. In the SM, two major theories are combined, the electroweak theory and quantum chromodynamics (QCD).

The electroweak theory is a unification of the processes behind electromagnetic and weak interactions. QCD is a quantum field theory responsible for the strong force.

Electromagnetic processes are mediated by the photon (γ) according to the theory of quantum electrodynamics (QED), whereas the force carriers for weak interaction are gauge bosons (W^+, W^- and Z^0). The electroweak interaction neither conserves quark nor lepton flavor. The strong force is mediated by gluons and the interactions between quarks and gluons are described by QCD. Each gluon has an effective color charge resulting from combining a color and an anticolor charge. There are eight possibilities for creating different effective color charges and hence eight different gluons (g_1,...,g_8). Because the gluons carry color charge, they are able to interact with each other. As a result, QCD processes are more complicated than electroweak processes.

The CKM Matrix

According to the SM, quark masses and mixings of quarks are related. They originate from interactions between the quarks and the Higgs field (see below). The mixings between the three quark generations are parametrised by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [3]. The CKM matrix elements determine the coupling constants of W-bosons to quarks and relate the initial and final state flavors. The CKM-matrix is shown in Equ. 2.1:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.1}$$

The Higgs Mechanism

The SM combination of the electroweak theory and QCD is mathematically structured by symmetry groups as SU(3)×SU(2)×U(1). The SU(3) group describes QCD and the SU(2)×U(1) group describes the electroweak interaction. The electroweak symmetry group is spontaneously broken by the proposed existance of a Higgs field with a non-zero vacuum expectation value. The existance of a Higgs field explains why the gauge bosons (W^+, W^-, Z^0) are massive while the photon (γ) remains massless in the SM [5].

According to theory regarding the Higgs mechanism, quark masses are generated when the Higgs field acquires a vacuum expectation value. Also lepton masses are generated in the electroweak theory by the Higgs mechanism. The Higgs mechanism includes couplings with the yet unobserved neutral scalar boson H^0, the Higgs boson. Experimentally observing the Higgs boson and understanding electroweak symmetry breaking is one of the main goals of the LHC [4].

2.2 The Top Quark

The top quark was discovered in 1995 at the Fermilab Tevatron collider by the CDF and D0 collaborations [6]. The Tevatron is a hadron collider, accelerating and colliding protons with antiprotons. The two experiments associated with it, CDF and D0, are currently the only experiments where top quarks have been directly observed.

The top quark is a third generation up-type quark with electric charge +2/3. It is the heaviest of all known elementary particles, with a current world average mass of 172.5 GeV [3], nearly as heavy as a gold nucleus.

Studying top quarks is important for several reasons. With its large mass, the top quark is the only fermion on the electroweak scale [5]. It is therefore of great interest for the studies of electroweak symmetry breaking. Accurately measuring the top quark mass also helps put constraints on the (SM) Higgs boson mass. Additionally, top quarks will constitute a significant background process to many BSM searches, including supersymmetry searches.

2.2.1 Top Quark Production Mechanisms

Top Quark Pair Production

In hadron collisions, $p\bar{p}$ at the Tevatron and pp at the LHC, top quarks are mainly produced in pairs of one top quark and one antitop quark ($t\bar{t}$). The top quark pair production occurs through two channels: Quark-antiquark annihilation ($q\bar{q} \rightarrow t\bar{t}$) and gluon coupling ($gg \rightarrow t\bar{t}$), both mediated by the strong interaction [3]. At the Tevatron, $q\bar{q}$ annihilation is the primary process whereas at the LHC, gluon coupling will dominate. Leading order Feynman diagrams for $t\bar{t}$ pair production mechanisms are shown in Figure 2.2.

At the Tevatron, the center-of-mass energy ($\sqrt{s}$) is 1.96 TeV. Corresponding top pair production cross section is 6.7 pb (cross sections are discussed in Section 2.3). At the LHC, the center-of-mass energy will be higher (14 TeV) and the total $t\bar{t}$ production cross section is expected to be 833 pb [7]. Top quarks will therefore be produced in much greater quantities at the LHC than at the Tevatron. A cross section of 833 pb corresponds to about 8 million $t\bar{t}$ pairs being produced per year at the luminosity the LHC is expected to achieve during its initial few years of running.

Single Top Production

In addition to being produced in pairs through the strong interaction, top quarks can also be produced in single top channels via the weak interaction. Single top production can occur through three subprocesses [9]:

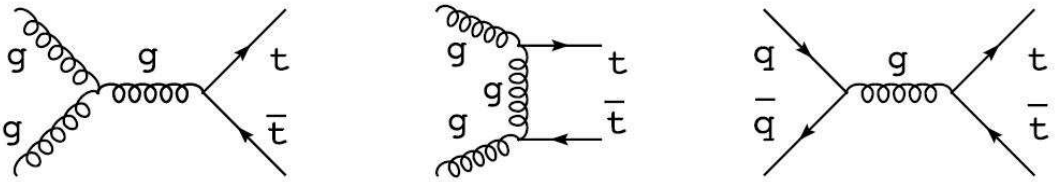


Figure 2.2: Leading order Feynman diagrams for $t\bar{t}$ pair production mechanisms in hadron collisions [8].

- t-channel: The t-channel refers to the process where a space-like W-boson is exchanged. It is also known as W-gluon fusion since the final state b-quark results from a gluon splitting to a $b\bar{b}$ pair. The t-channel is the largest source of single top production, with a cross section of 81.3 pb.
- s-channel: In the s-channel, a time-like W-boson is produced and decays to a t and a b-quark. This process has the lowest cross section of the three single top channels, only 3.3 pb.
- Wt: In the Wt process, a real W-boson is produced together with the top quark. The cross section of the Wt process is 26.7 pb.

Feynman diagrams for the three subprocesses can be seen in Figure 2.3.

Identifying top quarks originating from electroweak single top channels is a much harder task than the identification of top quarks from QCD $t\bar{t}$ production. This is due to both a less characteristic signature of the decay and to higher backgrounds. In this analysis, top pair production is considered, however, single top production is an important background process. Single produced top quarks must therefore be studied as part of the background estimation.

2.2.2 Top Quark Decay

A top quark decays predominantly into a W-boson and a b-quark, regardless of whether it was produced through a QCD $t\bar{t}$ process or through an electroweak single top channel. The reason why almost all top quarks decay as $t \to Wb$ is that the other decay modes ($t \to Ws$ and $t \to Wd$) are heavily suppressed in the CKM matrix (see Section 2.1.2). The latest estimates from the Particle Data Group [3] of the concerned CKM matrix elements are:

$$V_{td} = 0.00814$$
$$V_{ts} = 0.04161$$
$$V_{tb} = 0.99910$$

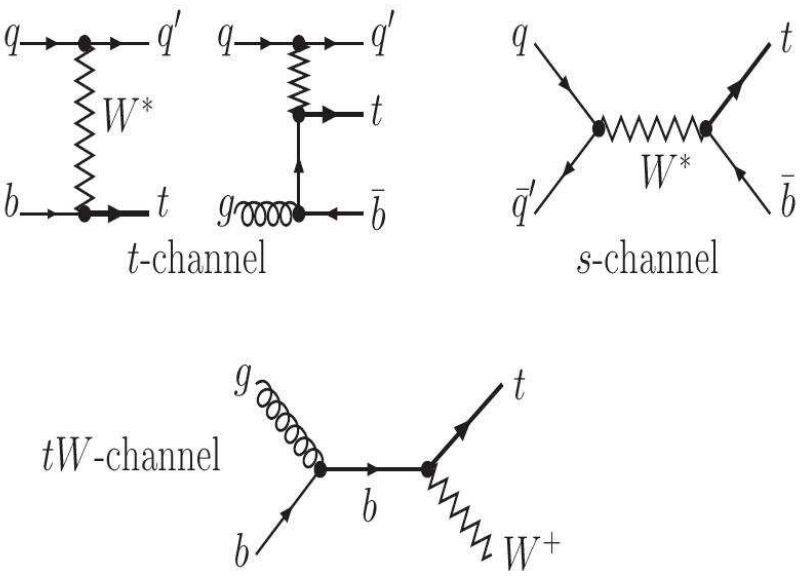


Figure 2.3: Feynamn diagrams for the three single top production processes [9].

For this analysis, all top quarks are assumed to decay via the $t \to Wb$ process (i.e. assuming $V_{tb} = 1$).

The W-boson originating from each top quark can decay in two different ways. It can either decay leptonically to a lepton and a neutrino, or hadronically to a quark-antiquark pair. The final states of a leptonically decaying W are $e\nu_e$, $\mu\nu_\mu$ and $\tau\nu_\tau$. The final states are equally probable and hence have equal branching ratios. A W-boson can decay hadronically to the two lightest generations of quarks ($W \to u\bar{d}$, $W \to c\bar{s}$). Since the quarks carry color charge, a hadronic final state has a three times larger branching ratio (BR) than a leptonic: $\mathrm{BR}(W \to q\bar{q}') = \frac{3}{9}$ and $\mathrm{BR}(W \to l\nu_l) = \frac{1}{9}$.

Since each $t\bar{t}$ pair contains two top quarks (one top quark and one antitop quark, however, the signatures for these are the same), two W-bosons are produced in each $t\bar{t}$ decay. There are essentially three possible final states for the $t\bar{t}$ decay, depending on how the W-bosons decay [8]:

- Fully hadronic: In the fully hadronic decay mode (or the all-jets mode) both W-bosons decay to a quark-antiquark pair. The final state of this mode consists of six jets of which two are b-jets originating from the top decay and the other four are from the decaying Ws.

  From the branching ratios of the W decay above, the branching ratio for the fully hadronic mode is calculated as:

  $\mathrm{BR(all\ jets)} = (\frac{3}{9} \cdot \frac{3}{9}) \cdot 4 = \frac{36}{81} = \frac{4}{9}$

Despite $\frac{4}{9}$ of all $t\bar{t}$ pairs decaying to the fully hadronic mode, the mode is experimentally hard to analyze. The absense of a high-p_T lepton makes it complicated to distinguish the fully hadronic mode from the SM QCD multijet production. The fully hadronic decay mode is not studied in this analysis.

- Fully leptonic: Contrary to the fully hadronic mode, the fully leptonic mode (or the dilepton mode) has a fairly clean signature. Here, both W-bosons decay to a lepton and a neutrino. The final state thus contains two high-p_T leptons, two neutrinos and two b-jets. The decay rate is lower for the dilepton mode than for the other modes:

  $\text{BR}(W \to ee/\mu\mu/\tau\tau) = \frac{1}{9} \cdot \frac{1}{9} = \frac{1}{81}$

  $\text{BR}(W \to e\mu/e\tau/\mu\tau) = (\frac{1}{9} \cdot \frac{1}{9}) \cdot 2 = \frac{2}{81}$

  In total, only $\frac{1}{9}$ of $t\bar{t}$ pairs have both W-bosons decaying leptonically.

- Semileptonic: In the semileptonic mode, one of the W-bosons decays leptonically while the other decays hadronically. The final state is distinguished by one high-p_T lepton, one neutrino and four jets (of which two are b-jets from the top decay and the other two originate from the hadronically decaying W).

  The branching ratio is calculated using the branching ratios for the W decay above:

  $\text{BR}(e/\mu/\tau + \text{jets}) = (\frac{1}{9} \cdot \frac{3}{9} \cdot 2) \cdot 2 = \frac{12}{81} = \frac{4}{27}$

  The branching ratio for each channel (e+jets, μ+jets or τ+jets) is thus $\frac{4}{27}$, resulting in a total branching ratio for the semileptonic mode of $\frac{4}{9}$.

2.3 Cross Section Calculations

The production cross section describes the likelihood of an interaction to take place. It is measured in barn (b), where 1 b = 10^{-28} m^2. The cross section is closely related to the number of produced events and the beam luminosity. The rate of events (dN/dt) originating from a specific interaction type is related to the interaction cross section (σ) and the instantaneous beam luminosity (L [cm^{-2} s^{-1}]) by:

$$\frac{dN}{dt} = L \cdot \sigma \tag{2.2}$$

The integrated luminosity ($\mathcal{L}$ [cm^{-2}]) is defined by integrating the instantaneous luminosity with respect to time, $\mathcal{L} = \int L dt$. This results in a total number of produced

events (N) from a particular interaction type of:

$$N = \mathcal{L} \cdot \sigma \tag{2.3}$$

The above formula needs modification in order to be used for cross section calculations. In an anlysis one can only count the number of observed events, not the total number of produced events. Included in the number of observed events are also the background events. The fraction of the number of observed events to the number of produced events is the acceptance ($\mathcal{A}$). The number of background events (N_{bkg}) is estimated separately and subtracted from the number of observed events (N_{obs}). Reformulating Equ. 2.3 results in the following cross section formula:

$$\sigma = \frac{N_{obs} - N_{bkg}}{\mathcal{A} \cdot \mathcal{L}} \tag{2.4}$$

Chapter 3

Experimental Apparatus

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a particle accelerator, soon due to start operating at CERN (European Organization for Nuclear Research) outside Geneva. The LHC has been built in a circular tunnel about 100 meters under ground, on the border between France and Switerland. The 27 kilometer in circumference tunnel was previously used by the Large Electron-Positron Collider (LEP).

The LHC will accelerate and collide two counter circulating beams of 7 TeV protons at a total final center-of-mass energy of 14 TeV [4]. The energies reached by the LHC have never been reached in a controlled environment before and the discovery potential is hugh. The first collisions are, at the time of writing, scheduled to take place during fall 2008.

3.1.1 The LHC Accelerator

The two proton beams at the LHC will be created in chain of machines, each step with increasing energy. The first step in this process is stripping electrons off hydrogen atoms to create protons. From the linear accelerator Linac2, 50 MeV protons continue to the Proton Synchrotron Booster (PSB) where they are accelerated to 1.4 GeV. From the PSB, the beam is injected into the Proton Synchrotron (PS) where the protons reach an energy of 26 GeV. Finally, the Super Proton Synchrotron (SPS) accelerates the protons to 450 GeV after which they are injected into the main LHC ring. The two counter circulating beams are then accelerated in separate beam pipes. The beams are governed by a strong magnetic field, created using superconducting electromagnets. The peak magnetic dipole field strength is 8.3 Tesla (T) [10].

The protons in each beam are concentrated in bunches. The bunches, 2808 in total, each contains 10^{11} particles. The bunches are at 25 ns apart, corresponding to a beam crossing frequency of 40 MHz. The design luminosity of the LHC is 10^{34} cm^{-2} s^{-2},

however, during the inital running period the luminosity will be lower. In each of the beam crossings, approximately 20 collisions will occur. Due to the high collision rate, pile-up of events is a scenario that requires consideration.

3.1.2 The LHC Experiments

Six different detectors, or experiments, have been constructed to study the collisions at the LHC. The counter circulating beams will collide at four interaction points around the LHC ring. Four of the experiments (ATLAS, CMS, LHCb and ALICE) are large detectors, each situated at one of the interaction points. The other two experiments (LHCf and TOTEM) are smaller and situated close to already existing interaction points [11]. The different experiments are briefly summarized below.

- ATLAS (A Toroidal LHC ApparatuS) is a large general purpose detector. It is discussed in detail in Section 3.2.
- CMS (Compact Muon Solenoid), like ATLAS, is a large general purpose detector, designed to enable studies of whatever might show up in the high energy collisions. It has essentially the same physics goals and comparable discovery potential as ATLAS. What distinguishes ATLAS and CMS is the design of the detectors.
- The LHCb (Large Hadron Collider beauty) is an experiment specialized in studying the physics of the bottom quark, or the beauty quark as it is occasionally referred to.
- ALICE (A Large Ion Collider Experiment), a detector that studies, as the name implies, heavy ion collisions. The LHC will be able to accelerate heavy ions (Pb nuclei) in addition to protons, at a center-of-mass energy of 5.5 TeV per nucleon. The physics of these Pb-Pb nuclei collisions, including potential observations of the quark-gluon plasma, will be studied in ALICE.
- LHCf (Large Hadron Collider forward), a smaller experiment that will be installed close to the ATLAS interaction point. It is intended to study cosmic ray physics by measuring the number and energy of neutral pions in the forward regions of ATLAS.
- Another smaller experiment, TOTEM (TOTal Elastic and diffractive cross section Measurement), will be installed close to the CMS interaction point. It is dedicated to measuring the total cross section, scattering and diffractive processes at the LHC.

3.2 The ATLAS Detector

The ATLAS detector is a general purpose detector, designed to measure the broadest range of signals possible. The high energy and luminosity of the LHC enables a large variety of physics studies. This includes detailed studies of already known physics objects as well as exploring the high energy frontier. The physics goal of the ATLAS experiment include [4]:

- Understanding the mechanism behind electroweak symmetry breaking (see Section 2.1.2), the most important physics goal. This automatically leads to a major focus on searches for the Higgs boson.
- Searches for other phenomena related to symmetry breaking, such as supersymmetry and supersymmetric particles.
- Searches for new heavy gauge bosons.
- Searches for evidence for composite quarks and leptons.
- Conducting precision measurements of the W-boson and top quark masses.
- Investigating CP violation in B decays.

ATLAS Coordinate System

Throughout this thesis, the standard ATLAS coordinate system is employed [4]:

- The z-axis is defined by the beam direction and the x and y coordinates lie in the plane transverse to the beam direction.
- The x-axis points from the interaction point to the center of the LHC ring.
- The y-axis points upwards from the beam.

From the rectangular coordinates, spherical (r,ϕ,θ) and cylindrical (r,ϕ,z) coordinates are defined as usual. The pseudorapidity η is defined in terms of the azimuthal angle θ:

$$\eta = -\ln\ (\ \tan(\frac{\theta}{2}))$$

The distance ΔR will be used frequently and is defined as:

$$\Delta R = \sqrt{\Delta^2\eta + \Delta^2\phi}$$

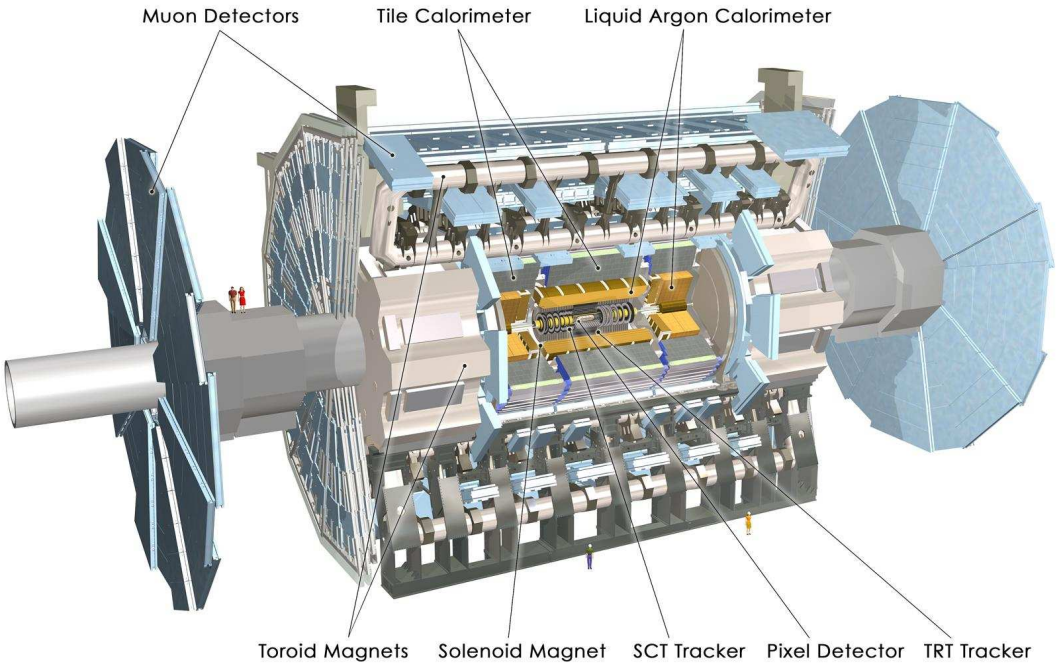


Figure 3.1: The ATLAS detector [12].

Detector Overview

The 46 m long, 25 m high and 25 m wide ATLAS detector is composed of four subsystems: Inner detector system, electromagnetic and hadronic calorimeters, full coverage muon detection and a magnetic system (Figure 3.2). The inner detector measures the momentum of charged particles whereas the calorimeters measure their energies. In addition to the detector components, three more systems are used to handle the vast flow of data: The trigger system, the data acquisition system and the computing system. The most important design critera of the ATLAS detector are [4]:

- Electromagnetic and hadronic calorimetry for accurate electron, photon, jet and missing transverse energy measurements.
- Efficient tracking, enabling high-p_T lepton momentum measurements and electron, photon, tauon and heavy flavor identification at high luminosity and full event reconstruction at low luminosity.
- High-precision muon momentum measurements and possibilities of using the muon spectrometer alone at high luminosity.

- Large pseudorapidity η acceptance with close to full coverage in azimuthal angle θ.
- Triggering and measurements of low-p_T particles.

3.2.1 Magnetic System

The ATLAS magnetic system consists of an inner superconducting Central Solenoid (CS) and superconducting air-core toroids. The CS surrounds the inner detector and provides a magnetic field of 2 T with a peak field strength of 2.6 T. The toroids, one barrel and two end-caps, each contain eight coils symmetrically arranged outside the calorimeters at a radius of 20 m. The peak magnetic fields are 3.9 T and 4.1 T for the barrel and the end-cap toroids respectively. The CS is located inside the electromagnetic calorimeter. In order to minimize its impact on the measurements in the calorimeter, it is designed with minimum coil material [4].

As a result of the layout of the magnetic fields, charged particle tracks are bent in the detector. The tracks are bent in the r-ϕ plane as the particles transverse the CS and in the r-z plane while they pass the toroids.

3.2.2 Inner Detector

The inner detector is designed to provide precision tracking while simultaneously minimizing multiple scattering of the particles transversing. The inner detector is a system of subdetectors located closest to the interaction point (the vertex). It consists of high resolution detectors immediately outside the vertex and continuous tracking detectors at the outer radii to meet the requirements on momentum and vertex resolution, enabling high precision measurements. By reconstructing tracks and vertices with high efficiency, the inner detector, together with the calorimeter and muon system, is responsible for the recognition of electrons, photons and muons. The acceptance of the inner detector covers the pseudorapidity region $|\eta| < 2.5$ [13].

The density of tracks from charged particles is expected to be very high at the LHC. Fine-granularity detectors are therefore necessary to allow precision measurements of the charged particle trajectories. For this purpose ATLAS uses two different semiconductor tracking detectors with silicon microstrip and pixel technologies. Closest to the beam vertex are the silicon pixel detectors, which offer the highest granularity. The number of layers must be limited both to reduce the amount of induced material but also due to their high cost. The main contribution of the pixel detector is its precise measurements of track trajectories.

Outside the pixel detectors is the Semiconductor Tracker (SCT) which measures particle momenta. The third subdetector of the inner detector is the Transition Radiation

Tracker (TRT). It provides possibilities of continuous tracking using less material and at a lower cost than the pixel and the SCT.

The inner detector is of great importance for the identification of short-lived particles such as b-quarks and is therefore a crucial part of the detector for many measurements, including those discussed in this thesis.

3.2.3 Calorimeters

The ATLAS calorimeter system consists of electromagnetic and hadronic calorimeters. Its main purpose is to carry out accurate measurements of the energy and position of electrons and photons, the energy and direction of jets and measuring the missing transverse energy of the event [14].

The hadronic calorimeter is divided into a barrel (covering the pseudorapidity range $|\eta| < 1.7$) and two end-caps (covering $1.5 < |\eta| < 3.2$). The electromagnetic calorimeter covers the range $|\eta| < 3.2$. There are also forward calorimeters covering $3.1 < |\eta| < 4.9$. The electromagnetic calorimeter uses a lead/liquid-argon (LAr) technique whereas the hadronic calorimeters use plastic scintillator plates (tiles) in an iron absorber at small pseudorapidities and LAr at larger rapidities.

One challenge for the calorimeters is the handling of pile-up events. Pile-up is a result of the high luminosity at the LHC where about 20 collisions are produced for each bunch crossing (every 25 ns). To minimize the impact of pile-up on the detector performance, fast detector response (¡ 50 ns) is required in addition to fine detector granularity.

3.2.4 Muon Spectrometer

The ATLAS detector has a high resolution muon spectrometer to take advantage of the robust signatures that high-momentum muons offer. It is located outside the calorimeters and defines the overall dimensions of the ATLAS detector. The muon spectrometer is based on magnetic deflection of muon tracks in large superconducting toroids together with stand-alone triggering (trigger chambers) and high precision tracking chambers [4].

The precision tracking chambers are built using Monitored Drift Tubes (MDTs) in the barrel region ($|\eta| \leq 1.0$) and in the end-caps out to a pseudorapidity $|\eta| \leq 2.0$. The inner end-cap regions ($2.0 \leq |\eta| \leq 2.7$), close to the interaction point where rates are higher, are instead covered using Cathode Strip Chambers (CSCs) which have higher granularity. In the barrel region, the chambers are arranged in three cylindrical layers at radii of approximately 5 m, 7.5 m and 10 m. The end-cap regions consist of four disks arranged vertically out to about $z = \pm 23$ m [4].

The trigger chambers cover the pseudorapidity range $|\eta| \leq 2.4$. For the triggering, Resistive Plate Chambers (RPCs) are used in the barrel regions while Thin Gap Chambers (TGCs) are used in the end-caps.

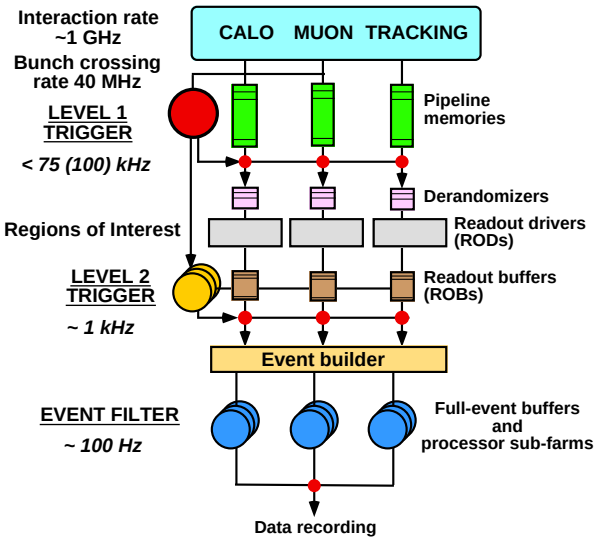


Figure 3.2: Schematic figure of the ATLAS trigger levels [4].

3.2.5 Trigger and Data Acquisition System

The trigger and data acquisition system in ATLAS is based on three separate levels of online selection. The trigger system is essential for reducing the enormous data flow and to select potentially interesting events from the huge background present in the LHC collisions. Each trigger level applies additional selection criteria to narrow down the number of accepted events.

The initial bunch crossing rate in the LHC is 40 MHz but it must be reduced to about 100 Hz for permanent storage. This puts extreme requirements on the triggers to both reject non-interesting minimum bias events with a high rejection factor and simultaneously maintain maximum efficiency for selecting the rare physics processes one wishes to study.

The ATLAS trigger system is divided into three subsystems: The first level tigger (LVL1), the second level trigger (LVL2) and the Event Filter (EF). The LVL2 trigger and the EF together constitute the High Level Trigger (HLT). A schematic figure of the trigger system is shown in Figure 3.2.

First Level Trigger

The first event selection in ATLAS is governed by the LVL1 trigger. The trigger decision is based on reduced-granularity information from different ATLAS subdetectors. The muon trigger chambers are used to identify high-p_T muons. Basic information from the electromagnetic and hadronic calorimeters are used to identify high-p_T electrons,

photons, jets and tauons decaying into hadrons. Furthermore, large missing energy and total transverse energy are used as trigger signatures. At the LVL1 trigger level, most physics requirements can be met using considerably simple selection criteria. The trigger implementation can, however, be programmed to select more complicated event signatures.

One requirement on the LVL1 trigger is the ability to uniquely identify the bunch crossing of interest. Due to a bunch crossing interval of merely 25 ns, this is non-trivial. During the time it takes for the LVL1 trigger decision to be formed and distributed, all information from the detectors needs to be temporarly stored in pipeline memories. In order not to exceed the maximum pipeline memory storage, the LVL1 trigger needs to take no longer than 2.5 μs from the corresponding bunch crossing to the LVL1 accept. To achieve this, the LVL1 trigger is implemented as a system of hardware processes.

Events that are selected by the LVL1 trigger are first read out from the front-end electronics into readout drivers (RODs) and then into readout buffers (ROBs). The events are kept in the ROBs until a decision is made by the LVL2 trigger whether to keep or reject the event [4].

Second Level Trigger

As opposed to the hardware implemented LVL1 trigger, the LVL2 trigger is, together with the EF, implemented in software. The data used by the LVL2 trigger is provided by the LVL1 trigger as Region-of-Interest (RoI) information. This includes information on position and transverse momentum of the candidate trigger objects and missing transverse energy. Using the RoI information, the LVL2 trigger reads the data necessary to form the LVL2 decision from the ROBs, where it was originally stored by the LVL1 trigger.

Event Filter

The last step in the online selection process is made by the EF. The events selected by the EF in this final stage will be written to storage and accessible for subsequent offline analysis. The total event rate will now have been reduced by the three trigger levels from an original rate of 40 MHz to 100 Hz, which is the output rate from the EF. This corresponds to a data rate of 100 MB/s [4].

3.2.6 ATLAS Computing

ATLAS uses an advanced computing system to handle and distribute the enormous amount of data that will be produced in the LHC collisions. The yearly data volume is on the order PB (10^{15} B). The overlying computing system is based on the GRID technology, a way of sharing computer power and data storage capacity over the internet.

Event data from the ATLAS detector will primarly be processed at CERN in the Tier-0 facility. After being archived at Tier-0 at CERN, it is copied to various Tier-1 facilities located around the world. From the Tier-1 facilities, the data can be further processed before being passed on for analyses at the Tier-2 centers.

ATLAS software uses an object-oriented approach and is mainly based on the C++ language. Software development is built on a common framework, Athena [15]. All processing of ATLAS data takes place within the Athena framework, from the software based HLT to the simulation of the detector response to various physics processes, and data analysis.

The data flowing from the trigger system is byte stream data. After the EF, the last stage of the trigger system, events are in the RAW data format and stored as Raw Data Objects (RDOs). After the event data has been run through reconstruction, it is referred to as Event Summary Data (ESD). From this stage onwards, it has an object-oriented representation. In principal, all information can be accessed from the ESDs and there is no longer a need for the RDOs accept for re-reconstructing data and for some calibrations. The size of the ESDs are about 500 kB per event and they are stored in so called POOL ROOT files.

From the ESD, the Analysis Object Data (AOD) is derived. It is a reduced event representation where each event corresponds to about 100 kB. Also AODs are stored in POOL ROOT files. The AODs contain information about physics objects and other elements which are of interest for the analysis. For the work conducted in this thesis, AODs are used exclusively. Other data formats for stripped event data include, among others, TAG (Tag data) and DPD (Derived Physics Data).

Offline Software

The ATLAS offline software is implemented to further process the events from the trigger and data acquisition system and to deliver the processed events to users. Additionally, it provides a framework of tools for analyzing the information. The different components of the ATLAS offline software system can be structured as [15]:

- The Athena Framework: A common event processing framework
- Event generation and simulation
- Event selection and reconstruction
- Calibration and Alignment
- Software Infrastructure

Monte Carlo Production Chain

A variety of different Monte Carlo (MC) samples are used for the analysis. Producing MC events requires a full chain of steps: From generation of events, to simulation, digitization and reconstruction, after which the MC data are stored as ESDs in the same way as "real" data. From the ESDs, AODs can be created for analysis purposes.

Events are generated within the Athena framework. The generation involves producing particle four-vectors from specific physics processes. Several event generators exist and can be used for this purpose. MC samples used in this thesis are generated using Pythia [16], Herwig [17] and MC@NLO [18].

After the event generation, events are further processed through simulation. Simulation of events refers to passing generated events through a GEANT4 simulation of the ATLAS detector in order to produce GEANT4 hits. These hits include information about each particle, where in the detector it went through and how much energy it deposited [19].

The next step in the MC production chain is the digitization of events. Digitization starts from the GEANT4 hits resulting from simulation and produce digits, such as times and voltages, corresponding to how the detector responds when particles traverse in the collisions. The outcome of digitization is Raw Data (RDO), and the MC data now have the same format as "real" data.

Finally, digitized events are run through the reconstruction software. During reconstruction, the digits are reconstructed into tracks and energy deposits in the ESD format. The reconstruction step in the MC chain is the same as will take place with "real" data. As mentioned above, the reduced data objects AODs can be created from the ESDs.

Chapter 4

Analysis Overview

The purpose of the analysis conducted in this thesis is to study the top quark pair production cross section using the ATLAS stream test data (see Section 5.1) [20]. The method is similar to what was used by the CDF Collaboration for their original top quark discovery in 1995 [6].

An important part of the work is to study the $t\bar{t}$ decay by identifying b-quarks from the decaying W-bosons , b-quark tagging. At CDF, tagging jets originating from b-quark fragmentation was neccessary in order to reduce backgrounds and to identify the $t\bar{t}$ signal events. The CDF top discovery and the need for b-tagging can be seen in Figure 4.1.

Part of the work for this thesis is closely connected to the top analysis documented in the ATLAS note "A pre-commissioning $t\bar{t}$ cross section measurement at ATLAS" [21]. The cross section calculations without b-tagging follows the method employed for the analysis in the above ATLAS note. For the semileptonic mode, the calculations are similar. This thesis also discusses the dilepton mode and additionally measures the $t\bar{t}$ cross section using b-tagging with a different method. Because of higher signal-to-background ratio at the LHC, $t\bar{t}$ decays can be identified without b-tagging. However, b-tagging is still an efficient method for reducing backgrounds and purifying the sample. Here, the $t\bar{t}$ cross section is measured for the semileptonic mode and for the dilepton mode, respectively, both with and without using b-tagging.

4.1 Cross Section Measurements Without b-Tagging

The measurement of the $t\bar{t}$ production cross section without applying b-tagging is conducted using a counting method. Starting from Equ. 2.4, the idea is to count the number of observed $t\bar{t}$ event candidates. The background processes need to be thoroughly analyzed in order to estimate the number of background events (see Chapter 8). From here on, the number of observed $t\bar{t}$ event candidates subtracted by the estimated number of

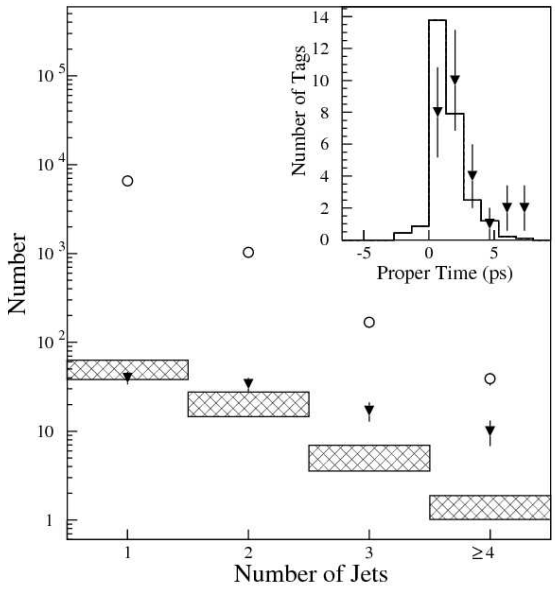


Figure 4.1: The CDF original top discovery [6]. The boxed regions show the predicted number of events without top quarks, the circles show the measured number of events without b-tagging and the black marks show the measured number of events after b-tagging. The excess in the number of measured events is ascribed to top quarks.

background events is $N_{t\bar{t}}$:

$$N_{t\bar{t}} = N_{obs} - N_{bkg} \tag{4.1}$$

The acceptance of event selection is estimated from a MC@NLO generated $t\bar{t}$ sample. The luminosity of the stream test data has already been calculated elsewhere and the result is here simple quoted.

When determining the cross section, one needs to take the efficiency of the trigger $\epsilon_{trigger}$ into account (see Section 5.5.1). Furthermore, the analysis acceptance is corrected by the electron identification efficiency ϵ_{reco} (discussed in Section 5.5.2). As discussed in Chapter 7, when the MC@NLO $t\bar{t}$ sample used to calculate the acceptance was created, it was filtered at generator level for a p_T ¿, 5 MeV lepton with a filter efficiency ϵ_{filter} of 0.554. Since the acceptance is calculated with respect to that inclusive $t\bar{t}$ sample, the $t\bar{t}$ cross section is here corrected by dividing by the filter efficiency.

This results in [21]:

$$\sigma_{t\bar{t}} = \frac{N_{t\bar{t}}}{\mathcal{L}\epsilon_{trigger}(\mathcal{A}\epsilon_{reco})\epsilon_{filter}} \tag{4.2}$$

4.2 Cross Section Measurements Using b-Tagging

The $t\bar{t}$ production cross section when applying b-tagging is measured using an event counting method. It is based on work presented by Henri Bachacou for the ATLAS Flavour Tagging Group [22]. The method allows for a simultaneous measurement of the b-tagging efficiency and the $t\bar{t}$ cross section (and to some extent the c-quark jet tagging efficiency). Precise measurements of the tagging performance is essential for its use in determining the cross section. Since each decaying $t\bar{t}$ pair contains two b-quarks in the final state, counting events with a certain number of b-quarks gives a method for measuring the b-tagging efficiency.

4.2.1 Method

Ideally, all selected $t\bar{t}$ events contain two b-quark jets and up to four light jets in the detector acceptance (zero for the dilepton mode, two for the semileptonic mode and four for the fully hadronic mode). However, not all jets fall within the detector acceptance and additionally, there are jets from Initial State Radiation (ISR) and Final State Radition (FSR). As a result, the event topology is in reality more complicated.

Events decaying to the semileptonic mode ideally have four jets in the final state, of which two are b-jets. For this mode, events with one, two or three tagged jets and four or five jets in total are therefore counted. For the dilepton mode, which ideally has events with two b-jets only in the final state, events with one or two tagged jets and two, three or four jets in total are instead counted. For both modes, the number of

background events is estimated using Monte Carlo and subtracted from the number of selected events.

If the selected $t\bar{t}$ sample always contained exactly two b-jets and only jets originating from b-quarks were tagged, the calculation of the b-tagging efficiency would be greatly simplified. In this case, the probability for tagging both b-jets is ϵ_b^2, where ϵ_b is the b-tagging efficiency. The probability for tagging one of the two b-jets is $2\epsilon_b(1\text{-}\epsilon_b)$. The number of single tagged (N_1) and double tagged (N_2) events are functions of the total number of $t\bar{t}$ events (N) and the b-tagging efficiency ϵ_b:

$$\begin{aligned} N_2 &= N\epsilon_b^2 \\ N_1 &= 2N\epsilon_b(1-\epsilon_b) \end{aligned} \tag{4.3}$$

Solving this simple system of equations for ϵ_b gives:

$$\epsilon_b = \frac{2N_2}{N_1 + 2N_2} \tag{4.4}$$

However, as briefly discussed above, the situation is not as simple as this. Various effects arise from jets from ISR/FSR or,for the semileptonic mode, from a hadronically decaying W, b-jets from the $t\bar{t}$ decay that fall outside the accepted region and a small number of b-jets produced through gluon radiation instead of via the $t\bar{t}$ decay [22]. To take these contributions into account, the actual flavor content of the sample of selected events is estimated from a Monte Carlo $t\bar{t}$ sample. This gives the jet fractions, F_{ijk}, of selected events before b-tagging with i b-jets, j c-jets and k light jets where i,j,k each range from 0 and up.

The expected number of events with n tagged jets, ¡N_n¿, can now be written as a sum over all possible combinations of i b-jets, j c-jets and k light jets as a function of the different tagging efficiencies:

$$\begin{aligned} &< N_n >= \mathcal{L}\sigma_{t\bar{t}}\epsilon_{trigger}(A\epsilon_{reco})\epsilon_{filter} \\ &\textstyle\sum_{i,j,k} F_{ijk} \sum_{i',j',k'} (A_i^{i'}\epsilon_b^{i'}(1-\epsilon_b)^{i-i'})(A_j^{j'}\epsilon_c^{j'}(1-\epsilon_c)^{j-j'})(A_k^{k'}\epsilon_l^{k'}(1-\epsilon_l)^{k-k'}) \end{aligned} \tag{4.5}$$

where $\mathcal{L}$ is the integrated luminosity, $\sigma_{t\bar{t}}$ is the $t\bar{t}$ cross section, $\epsilon_{trigger}$ the trigger efficiency, (A ϵ_{reco}) the acceptance corrected by the electron identification efficiency and ϵ_{filter} the Monte Carlo filter efficiency. The indices i',j',k' indicate the number of tagged b, c and light jets ($i' + j' + k' = n$). Finally, the factors $A_i^{i'}$ are defined as:

$$A_i^{i'} = \frac{i!}{i'!(i-i')!} \tag{4.6}$$

corresponding to the number of arrangements possible. From this, a likelihood function is formed as a product of Poisson distributions:

$$L = \prod_n Poisson(N_n, < N_n >) \tag{4.7}$$

where N_n is the counted number of events with n tags and $< N_n >$ the expected number of events with n tags. For the semileptonic mode, events with n=1,2,3 are taken into account whereas for the dilepton mode only events with n=1,2 are considered. Also considering events with zero tags is desirable for the dilepton mode but these events are too contaminated with background to be useful.

The likelihood fit is maximized with respect to the floating variables: For the semileptonic mode, $\sigma_{t\bar{t}}$ and both ϵ_b and ϵ_c are allowed to float since there are three equations (one for each of one, two and three tagged events). For the dilepton mode where only single and double tagged events are considered, only $\sigma_{t\bar{t}}$ and ϵ_b are allowed to float. For ϵ_c, the value resulting from the estimation in the semileptonic mode is instead used. For both modes, the mistag rate (i.e. the light jet tagging efficiency) is held fixed. It has to be determined elsewhere, see Section 6.2.2 for details.

4.3 Reconstructing the Top Mass

Precisely measuring the top quark mass is an extensive analysis of its own. Here, no attempts of conducting such a top mass analysis is made. It would require, among others, detailed studies of the jet and the b-jet energy scales. For the analysis in this thesis, a simple top mass reconstruction is performed as a validation of the $t\bar{t}$ event selection. The dilepton mode final state contains two undetected neutrinos which give rise to a joint missing transverse energy in the detector. For this reason, the dilepton mode has limited use in reconstructing the top mass and here the top mass (and the invariant mass of the hadronically decaying W) is reconstructed for the semileptonic mode only.

Chapter 5

Event Selection

This chapter describes the event selection used for the analysis. First, the ATLAS stream test data [23], which constitutes the "data", is discussed in detail. In addition to the data sample, various Monte Carlo samples are used to estimate the background, see Chapter 8. A brief overview of the reconstruction and identification of objects is given, with a main purpose of specifying the reconstruction algorithms used. The next section defines the physics objects of interest for the analysis, summarizing the various selection cuts. Also the event selection procedure and the event selection cuts are specified. Finally, calibrations and efficiencies that are used but calculated elsewhere are summarized.

5.1 Stream Test Data

The stream test data was created as an attempt of making Monte Carlo (MC) datasets that are as similar to "real" data as possible. The datasets are mixtures of various physics processes and contain no MC truth information about the objects.

Events that are accepted by the ATLAS trigger system are written to different streams depending on whether the triggering object was an electron, a muon, high transverse momentum etc. However, if the event satisfy the requirements for e.g. both the electron and the muon trigger, there are two possible streaming scenarios: Inclusive or exclusive data streaming [23]. In the case of inclusive streaming, the event will be written to both streams (in this case the electron and the muon stream). On the other hand, for exclusive streaming, the event will be written to a separate overlap stream only. Hence, when analyzing any exclusive stream one always has to include the overlap stream.

The stream test data was created with the two different streaming scenarios in mind. However, for storing the detector data only one of the two streaming alternatives will be implemented by ATLAS [20]. In total, the stream test data corresponds to 18 pb^{-1} of

data, equivalent to ten runs of data taking, each corresponding to 30 minutes at a luminosity of 10^{33} cm^{-2}s^{-1} [23]. For inclusive and exclusive streaming respectively, five raw data streams were defined: Electrons, Muons, Jets, Photons, and Tau and $\not\!\!E_T$. Including the additional overlap stream for exclusive streaming results in 11 streams in total for the stream test data.

An important feature of the stream test data is that it was simulated using release 11.0.42.X and reconstructed with release 12.0.6.X [20]. As a result of the different software releases for simulation and reconstruction, small calibration corrections need to be made (see Section 5.5). Additionally, for the same reason, minor discrepancies exist in the trigger decision. This is taken into account by requiring a reconstructed electron to match a triggering electron (see Section 5.4).

For the analysis carried out in this thesis, only the inclusive electron stream is used. When generating events for the stream test data, simulated online "dead-time" and luminosity blocks (in which luminosity information is stored) of bad data were included. The resulting luminosity of the sample is therefore not 18 pb^{-1} and needs to be calculated using the luminosity database. Taking the dead-time and file losses into account, the luminosity of the inclusive electron samples (samples 1a in Appendix A) is 15.03 pb^{-}1. Further removing the bad luminosity blocks (four luminosity block are marked "BAD" in the database), the final luminosity for the data used is calculated to be [21]:

$$\mathcal{L} = 14.81 \text{ pb}^{-1} \tag{5.1}$$

5.2 Reconstruction and Identification

5.2.1 Electrons and Photons

Electrons and photons are identified mainly using the inner detector and the electromagnetic calorimeter. Due to a significant amount of material in the inner detector, many electrons loose part of their energy already at this early stage of transversing the detector.

Reconstruction and identification of electrons and photons are similar and use the same algorithms. At AOD level, electrons and photons are distinguished by a track matching in the electromagnetic cluster. Photons are, however, not used for this analysis. For electrons, the reconstruction and identification procedures are described in detail below.

Electron Reconstruction

Two algorithms are used for reconstructing electrons [24]. The standard algorithm focuses on isolated high-p_T electrons. The other algorithm is mainly used for low-p_T electrons and electrons in jets. The standard algorithm is initialized by a cluster in the

electromagnetic calorimeter whereas the alternative algorithm is initialized by a track in the inner detector. Both algorithms result in the same Electron object and overlap removal is applied to handle the cases where both algorithms have reconstructed the same object.

For analysis purpose, electrons reconstructed by a particular algorithm can be selected using the `author` parameter. Electrons reconstucted by the standard (cluster) algorithm are selected by choosing the `author` to be "egamma". Soft electrons from the track based algorithm are instead selected by choosing `author` "softe".

Electron Identification

There are two methods for identifying electrons [24]. The identification aims at distinguishing between the electron hypothesis and the background hypothesis with processes giving similar detector signatures (mainly pions and jets). The standard method is used for isolated electrons whereas the alternative method is optimized for electrons originating from a b-quark where additional particles from the decay of the B-hadron or from fragmentation are likely to be near the electron. The electron identification methods are independent of which electron reconstruction algorithm that was used.

A variety of discriminating variables from the calorimeters are used to identify isolated electrons (i.e. the standard identification method) [25]. The calorimeter cuts help distinguish electrons from charged hadrons. The identification can be further improved using information from the inner detector, such as tracks pointing to the electromagnetic cluster and energy-momentum matching. This helps reduce background from photons and pions [24]. Also the alternative identification method uses discriminating variables from the calorimeters and the inner detector.

For this analysis only isolated electrons are considered. Isolated electrons, identified by the standard identification method, are selected using the `isEM` flag. If a cut is passed, a bit is set in the `isEM` flag. The identification can be made looser or tighter depending on which flags are required to be set.

5.2.2 Muons

Muon identification in ATLAS is enabled by the inner detector, the calorimeters and the muon spectrometer. Reconstruction of muons is primarly performed in the muon spectrometer.

Two muon spectrometer track reconstructions are implemented in ATLAS: Muonboy and MOORE. There are also inner detector tracking algorithms: NewTracking (the default), iPatRec and xKalman. For details on the algoritms see [26]. In an analysis, often a combined tracking algorithm is of interest. The combined algorithms result in high efficiency, good fake rejection and optimized momentum resolution. The two most commonly used combined tracking algorithms, STACO and MUID, uses inner detector

tracks and muon spectrometer tracks for the muon reconstruction. Finally, for low-p_T muons, the MuGirl and the MuTag algorithms can be used.

STACO

The first muon track reconstruction method used in ATLAS is STACO (STAtistical COmbination) [4]. STACO is based on statistically merging the independent inner detector measurements with the muon spectrometer track. The merging uses parameters of the reconstructed tracks and their covariance matrices.

At AOD level, STACO muons are found in the `StacoMuonCollection`. STACO standalone uses muons found by MuonBoy extrapolated back to the primary vertex. For combined muons, STACO uses Muonboy for the muon spectrometer reconstruction and the xKalman package for the inner detector track reconstruction. STACO low-p_T muons are found using the MuTag algorithm.

MUID

MUID is another muon track reconstruction method employed by ATLAS [4]. Instead of statistically merging independent tracks from the inner detector and the muon spectrometer, MUID constructs a global track fit. The fit uses hits that were found by the inner detector and the muon spectrometer separately, and used by the independent reconstructions. The tracks are matched by forming a chi-square from the parameter differences and summed covariances.

At AOD level, MUID muons are found in the `MuidMuonCollection`. MUID combined muons uses MOORE for the muon spectrometer reconstruction and the iPatRec package for the inner detector reconstruction. For MUID standalone, muon tracks from MOORE are used and extrapolated to the primary vertex. For low-p_T muons, the MuGirl algorithm is employed.

5.2.3 Jets

Jets, formed as a cone of particles produced by the hadronization of a quark or a gluon, can be identified by several algorithms in ATLAS. The algorithms differ in the input used to the jet finding, the jet finder algorithm itself and the main parameter of the jet finder [27]. The different types used for the algorithms are:

- Input type: Calorimeter towers at the electromagnetic scale ("tower") and topological clusters ("topo").
- Jet finding algorithm: Seeded cone algorithm ("cone") and inclusive k_T algorithm ("kt").

- Main parameter: The main parameter of the jet finder is dependent on which jet algorithm is used. For the cone algorithm, the cone radius (radius = 0.4 or 0.7) multiplied by 10 is used as parameter. For the k_T algorithm, the D parameter (D = 0.4 or 0.6) times 10 is used.

The jet finders, cone and k_T, are different methods for reconstructing jets [28]. Cone is a geometric algorithm where the jet constituents lie in a cone of given radius, where the cone axis is required to be aligned with the jet 4-momentum. The cone algorithm can be implemented as a seeded or as a seedless algorithm. The seeded algorithm, which is used by default, searches for cones around the highest-p_T constituent only. The k_T algorithm instead groups the closest constituents recursively into a jet.

At AOD level, different jet collections are labeled by input type, jet finder and main parameter. Two examples are Cone4TowerParticleJets and Kt6TopoParticleJets (see [29]).

5.2.4 Missing Transverse Energy

The missing transverse energy $\not{E}_T$ is important to measure for two reasons [4]. Firstly, it is an important signal for new physics. Secondly, it is necessary in order to reconstruct invariant mass distributions for particles with decays involving neutrinos, such as the top quark for instance.

The missing transverse energy is calculated from the energy deposited in all calorimeter cells and from muons in the muon spectrometer [30]. Part of the energy is lost in the cryostat and for this a correction is applied. $\not{E}_T$ reconstruction from calorimeter cells uses a refined calibration from physics objects where each cell is defined to belong to either zero or one physics object. The $\not{E}_T$ quantity recommended for physics analysis, labeled `MET_RefFinal`, uses this refined calibration. The name "missing transverse energy" is confusing because energy indicates it being a scalar when the $\not{E}_T$ is really calculated as a 2-vector with x- and y-components.

5.3 Object Definitions

Certain criteria need to be specified for how one chooses to select physics objects. In this section, selection cuts for different objects used in the analysis are defined.

5.3.1 Electron Definition

The definition of an electron is an object from an `ElectronContainer` with StoreGate key (used in ATLAS to retrieve data objects) `ElectronCollection`. Only electrons that have been reconstructed using the standard cluster algorithm are considered (`AuthorEgamma`). An electron is required to pass all electron identifications cuts except

the TRT ratio cut which is not used (`(isEM&0x7FF) == 0`). Two kinematic cuts are applied: The electron is required to have a transverse momentum greater than 25 GeV and a pseudorapidity less than 2.4 (exluding the calorimeter crack regions). To summarize, the electron selection cuts are:

- `AuthorEgamma`
- `(isEM&0x7FF) == 0`
- $p_T > 25$ GeV
- $|\eta| < 2.4$ excluding the calorimeter crack regions $|\eta| \in [1.37, 1.52]$

5.3.2 Muon Definition

A muon is an object from a MuonContainer with StoreGate key `StacoMuonCollection`. The `StacoMuonCollection` contains a combination of muons from different reconstruction algorithms, however, only high-p_T muons are selected (`author = highPt`). The pseudorapidity is required to be less than 2.4 and for the transverse momentum, a farily loose cut value of 15 GeV is applied. Additionally, the muon is required to be isolated with an isolation transverse energy less than 6 GeV in a cone of $\Delta R \leq 0.2$. The muon is also required to be separated from closed jet by more than 0.3 in ΔR . Hence, the muon selection cuts are as follows:

- `author = highPt`
- $p_T > 15$ GeV
- $|\eta| < 2.4$
- isolation $E_T < 6$ GeV in a cone of $\Delta R \leq 0.2$
- separation from jets $\Delta R(\mu, jet) > 0.3$

5.3.3 Jet Definition

A jet is defined as an object from a ParticleJetContainer with StoreGate key `Cone4TowerParticleJets`. Thus, jets that are studied have been reconstructed using the seeded cone algorithm, with cone radius 0.4, from calorimeter tower signals. Selected jets have a transverse momentum greater than 25 GeV and a pseudorapidity less than 2.5. In addition, overlap removal between electrons and jets is employed to avoid double counting an already selected electron as a jet. Jets are consequently required to be seperated from an electron (as defined above) by more than 0.3 in ΔR . To summarize:

- $p_T > 25$ GeV
- $\eta < 2.5$
- electron-jet separation: $\Delta R(e, jet) > 0.3$

b-Jet Definition

b-Jets are jets originating from b-quark fragmentation. In this analysis, the current ATLAS default tagger `1P3D+SV1` [31] is used. A jet is defined as a b-jet if its likelihood weight w is greater than 3.0. Increasing the weight cut would be optimal in order to reduce the mistag rate, however, because of low statistics in the stream test data a looser cut value is primarly chosen. Some studies in this thesis are for reference conducted using also a tighter cut value of 6.5. The loose and tight b-jet selections are:

- Loose b-jet selection: $w \geq 3.0$
- Tight b-jet selection: $w \geq 6.5$

5.3.4 Missing Transverse Energy Definition

The missing transverse energy $\not\!\!E_T$ is defined by the `MissingET` object with StoreGate key `MET_RefFinal`. See Section 5.2.4 for details on missing transverse energy.

5.4 Event Selection

To measure the $t\bar{t}$ production cross section, one needs to reconstruct the $t\bar{t}$ event candidates from its decay products. The event selection is common to both the analysis without b-tagging and the analysis using b-tagging. For both the semileptonic and the dilepton mode, at least one of the Ws will decay into a lepton-neutrino pair. To simplify the analysis by considering a single trigger stream, it is required that the event has an electron that satisfies the trigger requirements. Consequently, a first event selection is based on an inclusive sample of $W \rightarrow e\nu$ events.

A variety of selection cuts are used to define the inclusive W sample. Since events with an electron are considered, all selected events are required to pass the LVL2 inclusive 25 GeV electron trigger `L2_e25i` [20]. An event must also have one and exactly one electron (as defined in Section 5.3). In addition to passing the trigger, the reconstructed electron is required to match the triggering electron passing the `L2_e25i` cuts. A match is defined as the distance ΔR between the reconstructed electron and the triggering electron being less than 0.2. This requirement on matching is necessary in order to use a tag and probe method for determining the trigger efficiency (see Section 5.5.1) [21].

Additional selection cuts for the inclusive W sample are used for the semileptonic mode. The semileptonic mode and the dilepton mode are distinguished due to the dilepton mode containing two decaying W-bosons. Hence, the dilepton mode have two undetected neutrinos contributing to the missing transverse energy in the detector.

5.4.1 Semileptonic Mode

For the semileptonic mode, one of the W-bosons from the $t\bar{t}$ pair decays leptonically, here to an electron and a neutrino, and the other hadronically, as described in Section 2.2. The final state of this decay mode therefore consists of one electron, four jets of which two are b-jets and missing transverse energy from the undetected neutrino.

The event selection starts from an inclusive W sample. The undetected neutrino from the W decay results in missing transverse energy $\not\!\!E_T$ in the detector. In addition to the trigger requirement, a minimum $\not\!\!E_T$ of 25 GeV is required. Furthermore, a transverse W mass (m_T) cut of 45 GeV is used, where the transverse mass is defined as:

$$m_T \equiv (E_T(e) + \not\!\!E_T\ )^2 - (\overrightarrow{p_T}(e) + \not\!\!E_T\ )^2 \tag{5.2}$$

At least four jets are required in the decay signature in addition to the inclusive W selection. To summarize, the semileptonic event selection cuts are:

- Events pass the `L2_e25i` trigger and trigger electron match ($\Delta R < 0.2$)
- Exactly one electron
- $\not\!\!E_T > 25$ GeV
- $m_T > 45$ GeV
- $\mathrm{N}_{jets} \geq 4$

5.4.2 Dilepton Mode

One-ninth of $t\bar{t}$ decays are fully leptonic, with both W-bosons decaying into a lepton and a neutrino. The dilepton mode provides a very clean sample of $t\bar{t}$ events. Although the dilepton mode has limited use in measuring the top mass due to the two neutrinos constituting a joint missing transverse energy, it is valuable for measuring the $t\bar{t}$ cross section. Here, events with one electron and one muon are considered.

Also for the dilepton mode, the event selection is based on an inclusive W selection. However, due to the ambiguity regarding the joint missing transverse energy from the two undetected neutrinos, neither the $\not\!\!E_T$ cut nor the transverse mass cut is used for the dilepton mode. One and exactly one muon (as defined in Section 5.3) is required. For the dilepton mode, at least two jets are required, arising from the b-quark hadronization. To summarize, the event selection cuts for the dilepton mode are:

- Events pass the `L2_e25i` trigger and trigger electron match ($\Delta R < 0.2$)
- Exactly one electron
- Exactly one muon
- $N_{jets} \geq 2$

5.4.3 Event Selection Efficiencies

This section gives a brief overview of the number of events before each selection cut. The number of events before each selection cut is shown in Figure 5.1 for the semileptonic mode and in Figure 5.2 for the dilepton mode.

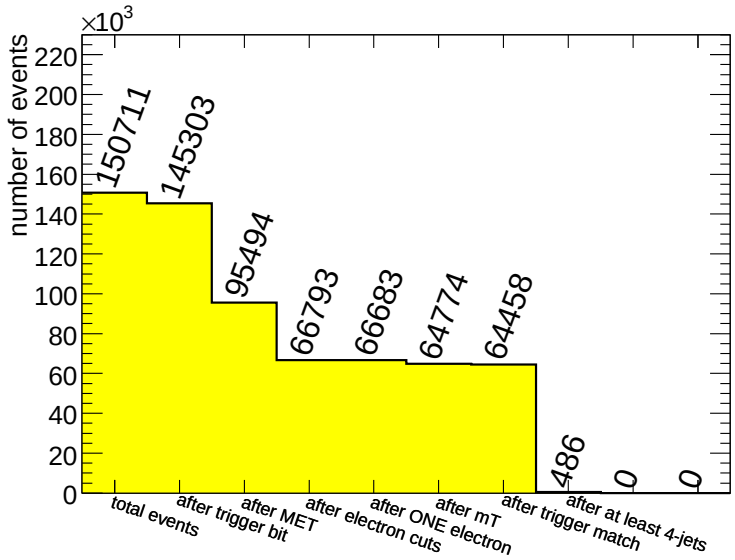


Figure 5.1: Number of events in the stream test data before each cut for the semileptonic mode.

5.5 Calibrations and Efficiencies

As mentioned in Section 5.1, the stream test data was simulated and reconstructed using different software releases. Consequently, the calibration during reconstruction was not optimal [21]. Moreover, the Monte Carlo samples that are used to calculate the acceptance were simulated using release 12 (as opposed to the stream test data which was simulated using release 11 but reconstructed with release 12). The release 12

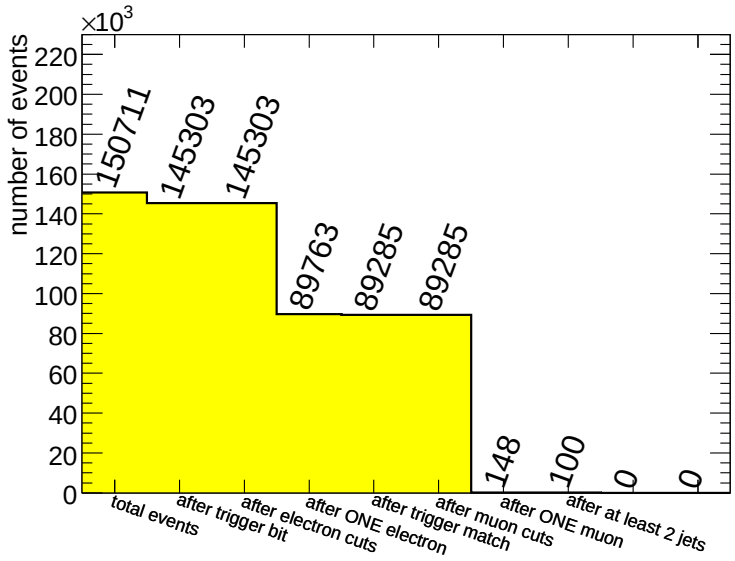


Figure 5.2: Number of events in the stream test data before each cut in the dilepton mode.

simulation uses more detector material than the release 11 simulation. To account for these differences, corrections are neccessary. The stream test data is treated as "real" data and corrections are derived and applied to the release 12 Monte Carlo rather than the stream test data.

5.5.1 Trigger Efficiency

The trigger efficiency of the `L2_e25i` trigger with respect to reconstruction was measured by Andre Bach for the ATLAS Note "A pre-commissioning $t\bar{t}$ cross section measurement at ATLAS" [21]. The measurement was carried out using a tag and probe method with electrons in the $Z \rightarrow ee$ peak. The result is here simply quoted. For electrons passing the selection cuts specified in Section 5.3, the trigger efficiency with respect to reconstruction is:

$$\epsilon_{trigger} = 0.9896 \pm 0.0011 \tag{5.3}$$

where the quoted error is statistical only.

5.5.2 Electron Identification Efficiency

Because of the differences between the release 11 and release 12 simulations, a correction is applied to the electron identification efficiency. This correction was calculated by Ayana Holloway-Arce for the ATLAS Note [21]. The calculation involved a method

similar to the tag and probe method used to measure the trigger efficiency. Also here, only the result is quoted. The electron identification efficiency ratio (data)/(Monte Carlo) is:

$$\epsilon_{reco} = 1.0183 \pm 0.0075$$

with the quoted errors being statistical only. The release 12 Monte Carlo thus have a lower electron identification efficiency than the release 11 simulated stream test data as a consequence of the increment of detector material in the later software release.

Chapter 6

b-Quark Jet Tagging

Identifying jets originating from b-quark fragmentation is essential for reducing background in many searches for new phenomena as well as for some precision measurements. It is also an important tool for the analysis event selection. In addition to its application of selecting events containing top quarks, it is useful for identifying Standard Model or supersymmetric Higgs bosons since these often, according to theory, couple to heavy objects.

Several methods for tagging b-quark jets have been developed by ATLAS. The most important ATLAS subdetector for tagging flavored jets is the inner detector. In this chapter, a brief overview is first given of the ATLAS b-tagging methods. Thereafter, the discussion is focused on the specific b-tagging algorithm used for this analysis.

In order to apply b-tagging to an analysis, detailed knowledge of its performance is required. To determine the tagging efficiencies, an event counting method is used. As described in Section 4.2, the method gives an estimate of the tagging efficiencies using data (here, data refers to the stream test data) as opposed to Monte Carlo information, while simultaneously determining the cross section. To accurately measure the tagging efficiencies from data, the fractions of differently flavored jets are needed. An estimate of this using Monte Carlo information is given in Section 6.2.1.

6.1 ATLAS b-Tagging Algorithms

Jets from b-quarks are identified by expoiting the long lifetime (approximately $1.5{\cdot}10^{-12}$ s [32]) and high mass of B-hadrons [33],[34]. Decaying B-hadrons have tracks with large impact parameters, with respect to the primary vertex, that form displaced secondary vertices. Algorithms that take use of decaying B-hadrons and their long lifetimes are referred to as space b-taggers. There are also methods for searching for soft leptons (electrons or muons) from semileptonic b-decays in the hadronic calorimeter. In this analysis, space b-tagging is used.

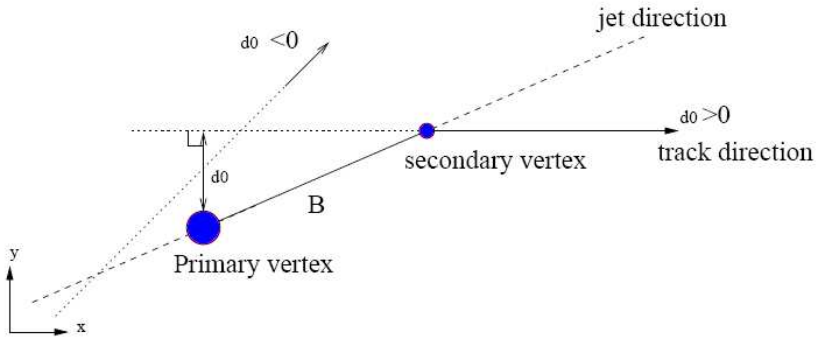


Figure 6.1: Definition of the transverse impact parameter and its sign [33].

A few different methods are used for space b-tagging [33]. The first method uses the transverse impact parameter a_0, illustrated in Figure 6.1. It is defined as the closest distance between the primary vertex and track helix in the plane transverse to the track. The sign of the impact parameter is determined by the jet direction. There is a difference between the impact parameter distributions for b-jets and light jets. Tracks from light jets have a symmetric transverse impact parameter distribution whereas b-jets have a long tail at positive values resulting from the long lifetimes of the B-hadrons.

Also the longitudinal impact parameter z_0 can be used for b-tagging. It is defined by the trajectory point with smallest transverse impact parameter. Similarly to the case for the transverse impact parameter, jets from b-quarks have longitudinal impact parameter distributions with a long positive tail as opposed to the symmetric distributions for light jets. The longitudinal impact parameter does, however, have a smaller discriminating power than the transverse impact parameter due to worse resolution.

From the distributions of the impact parameters, likelihood functions are constructed to distinguish between tracks originiating from the primary interactions (small $|a_0|$ and $|z_0|$) and from a decaying B-hadron (large $|a_0|$ and $|z_0|$). These likelihoods of tracks inside a jet are combined to a joint likelihood for the jet being a b-jet or a light jet.

In addition to using the impact parameters, one can take advantage of displaced secondary vertices to identify b-jets: Secondary vertex tagging. This is possible when there are two or more tracks within a jet with significant impact parameters.

6.1.1 Transverse Impact Parameter Tagging

The transverse impact parameter tagging uses, as indicated by the name, the transverse impact parameter of tracks from decaying B-hadrons. From here on it is denoted `IP2D` (2D Impact Parameter). The impact parameter is defined to be positive if the track crosses the jet axis in front of the primary vertex and negative if it crosses behind it

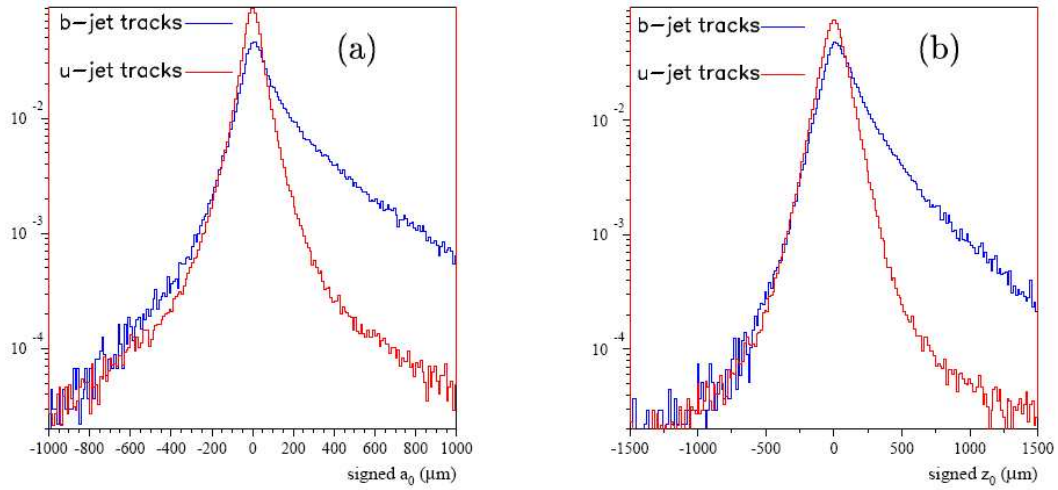


Figure 6.2: Transverse (a) and longitudinal (b) impact parameter distributions for b-jets and u-jets [33].

[33]. One further defines the significance S of the impact parameter as:

$$S_{a_0} = \frac{a_0}{\sigma_{a_0}} \tag{6.1}$$

where a_0 is the transverse impact parameter and σ_{a_0} the error on a_0. The likelihood functions $P_b(S_{a_0})$ and $P_u(S_{a_0})$ for a track originating from a b-jet or a u-jet is then used to compute a track weight w_t:

$$w_t = \frac{P_b(S_{a_0})}{P_u(S_{a_0})} \tag{6.2}$$

The weight of a jet w_{jet} is finally defined by summing the logarithms of the weights of different good tracks in the jet:

$$w_{jet} = \sum_{i \in jet} \ln w_t^i \tag{6.3}$$

An example of the distributions of this weight parameter for b-jets and u-jets can be seen in Figure 6.2.

6.1.2 3D Impact Parameter Tagging

The resolution on the longitudinal impact parameter z_0 is worse than on the transverse impact parameter a_0. The tagging performance can, however, still be improved using the longitudinal significance in combination with the transverse. The resulting algorithm is

labelled IP3D (2D+1D Impact Parameter). The jet weight is now a function of both the longitudinal and the transverse significance [33]:

$$w_t = \frac{P_b(S_{a_0}, S_{z_0})}{P_u(S_{a_0}, S_{z_0})} \tag{6.4}$$

6.1.3 Secondary Vertex Tagging

Identifying secondary vertices gives important information for the flavor tagging. First, a primary vertex is reconstructed, whose accuracy is crucial for the method. For the secondary vertex reconstruction, the method identifies pairs of tracks that form two-track vertices to first reject "bad" tracks originating from secondary interactions in the beam pipe and decay products from K^0 and Δ particles. Thereafter, the two-track vertices are combined to form a single effective secondary vertex. Because of B- and D-hadrons decaying into cascades of particles, the effective secondary vertex corresponds to a mean position between their decay points [33].

The secondary vertex tagging is mainly used in combination with an impact parameter tagger. To combine the two, one needs some discriminating variables for the secondary vertex information which is not strongly correlated with the impact parameter significances. Various variables can be used for this purpose, e.g. the fraction of jet energy in the secondary vertex, the number of two-track secondary vertices etc. Combining an impact parameter tagger and a secondary vertex tagger generally improves the tagging performance. The method using secondary vertices is highly dependent on the quality of the software though.

6.1.4 The ATLAS IP3D+SV1 Tagger

For the analysis in this thesis, the current default ATLAS tagging algorithm IP3D+SV1 is used [31]. It is a combined 3D impact parameter tagger and a secondary vertex tagger. Different likelihood weight cut values (w), constructed from the results of the IP3D impact parameter and the SV1 secondary vertex taggers, can be chosen in order to make a tight or a loose selection. Jets originating from b-quarks have higher likelihood weights whereas jets from light quarks have lower likelihood wieghts. Here, two likelihood weight cut values are used and studied:

- $w = 3.0$: A loose weight cut value which corresponds to a higher tagging efficiency. When using a looser cut value, one also has a higher mistag rate. However, because of low statistics in the stream test data this may be preferred.

- $w = 6.5$: A tight cut value corresponding to a lower b-jet tagging efficiency and a lower mistag rate. Using an increased weight cut is optimal in order to reduce the mistag rate (compared to that of $w = 3.0$), but it results in lower statistics.

Tagging efficiencies and cross sections are calculated using both likelihood values separately for comparison.

6.2 Tagging Efficiencies

Detailed knowledge about the flavor tagging performance is crucial for its use in determining the $t\bar{t}$ cross section. For this analysis, the b-tagging efficiency is estimated from data (i.e. the stream test data). Understanding the performance using real data is preferrable since Monte Carlo estimates only give approximate detector and performance descriptions. An upper limit on the mistag rate, i.e. the probability of tagging light jets as b-quark jets, is also given.

As discussed in Section 4.2, the b-tagging efficiency (and to some extent the c-tagging efficiency) can be determined simultaneously with the $t\bar{t}$ cross section. To utilize the method, the jet flavor fractions of the sample are needed. Additionally, the mistag rate must be estimated prior to the cross section and heavy flavor tagging efficiency measurements.

6.2.1 Jet Fractions

The jet fractions needed to determine the b-tagging efficiency are estimated using Monte Carlo truth from a $t\bar{t}$ sample (sample 2a in Appendix A). The fractions are calculated after making all kinematic cuts (as defined in Section 5.4) but before making any selection based on b-tagging. They are furthermore determined separately for the semileptonic mode (Table 6.1) and the dilepton mode (Table 6.2). For the semileptonic mode, events with four or five jets are considered. For the dilepton mode, which ideally contains two and not four jets in the final state, events with two, three or four jets are considered. The jet fractions are, however, Monte Carlo estimates and since the MC is not perfect, these fractions will have systematic uncertainties.

6.2.2 Mistag Rate

The mistag rate (or the light jet tagging efficiency as it is occasionally referred to in this thesis) is difficult to measure from data. Carrying out a detailed measurement of the mistag rate lies beyond the scope of this thesis. Effort was put in to including the mistag rate estimation in the likelihood calculations, however, without success. One alternative would be to use dijet events to get an upper limit on the mistag rate, assuming that all tagged jets are mistagged light jets. Also this was studied but due to significant heavy flavor contributions, the resulting upper limit is too high to be useful. For now, the mistag rate is instead estimated using Monte Carlo truth information about the fractions of tagged light flavored jets.

Jet flavor combination	Fraction of events
2 b-jets, 2 l-jets	32.89
2 b-jets, 1 c-jet, 1 l-jet	14.73
2 b-jets, 3 l-jets	14.43
2 b-jets, 1 c-jet, 2 l-jets	8.85
1 b-jet, 3 l-jets	11.46
1 b-jet, 1 c-jet, 2 l-jets	7.08
1 b-jet, 1 c-jet, 3 l-jets	2.47
1 b-jet, 4 l-jets	3.77

Table 6.1: (True) jet fractions (%) in events with four or five jets for the $t\bar{t}$ sample (semileptonic mode).

Jet flavor combination	Fraction of events
2 b-jets	29.61
1 b-jet, 1 c-jet	1.18
1 b-jet, 1 l-jet	17.77
2 l-jets	1.37
2 b-jets, 2 l-jet	8.14
1 b-jet, 3 l-jets	3.15
2 b-jets, 1 l-jet	23.12
1 b-jet, 1 c-jet, 1 l-jet	1.53
1 b-jet, 2 l-jets	8.32

Table 6.2: (True) jet fractions (%) in events with two, three or four jets for the $t\bar{t}$ sample (dilepton mode).

For the two likelihood weights $w = 3.0$ and 6.5 the following mistag rates (ϵ_l) are used:

- $w = 3.0$ use $\epsilon_l = 4$ %
- $w = 6.5$ use $\epsilon_l = 1$ %

It should be mentioned that the b-tagging efficiency measurement is not particularly sensitive to the exact value of the mistag rate. The c-tagging efficiency is, however, correlated with the mistag rate.

6.2.3 Implementation of Tagging Measurement Method

This section gives an overview of the implementation of the method for calculating the $b-$ (and $c-$) tagging efficiency and simultaneously extracting the $t\bar{t}$ cross section. The method starts from the likelihood formula in Equ. 4.7 Maximizing the likelihood function is equivalent to minimizing the negative logarithm of the likelikhood itself. This minimization of the likelihood function is carried out in ROOT [35], using Minuit [36]. The function that is actually minimized is (-2) times the logarithm of the likelihood function. The factor 2 is included so that Minuit calculates the errors correctly.

Before starting the minimization, a few variables must be specified, such as the mode (semileptonic or dilepton), the weight cut (3.0 or 6.5), which sample to use and whether background substraction should be applied or not. For the semileptonic mode, three parameters are allowed to float: ϵ_b, ϵ_c and $\sigma_{t\bar{t}}$, whereas for the dilepton mode, ϵ_c is kept fixed at the value resulting from the minimization in the semileptonic mode. The efficiencies are restricted to float between -0.05 and 1.0 whereas the cross section can range from 0 to 2000 pb.

The data used for the minimization is read by ROOT from output files from runs over AODs. If background subtraction is applied, the backgrounds are read from the various runs over different Monte Carlo samples and normalized and scaled as specified in Section 8.1. The likelihood function is then first calculated using the start values for efficiencies and cross section to determine the expected number of events. After each turn in the minimization loop, the likelihood is re-calculated until Minuit converges.

Chapter 7

Signal Acceptance

In this chapter the acceptance of the selection of $t\bar{t}$ signal events is calculated for the semileptonic mode and the dilepton mode, respectively. The acceptance of the event selection is calculated from a MC@NLO generated $t\bar{t}$ sample (sample 2a in Appendix A). When this sample was created, it was filtered at generator level for a lepton with p_T ¿ 5 MeV. All efficiencies are therefore calculated with respect to that $t\bar{t}$ inclusive lepton sample.

7.1 Acceptance in the Semileptonic Mode

Without Using b-Tagging

The selection efficiencies of different steps in selecting $t\bar{t}$ signal events are here calculated for the semileptonic mode. They are listed in Table 7.1 and can be seen graphically in Figure 7.1. The final selection efficiency, including the jet multiplicity cut ($N_{jet} \geq 4$), for the semilptonic mode gives the $t\bar{t}$ signal acceptance before b-tagging:

$$\mathcal{A} = 0.05228 \pm 0.00226 \tag{7.1}$$

where the error is statistical only.

Including b-Tagging

The signal acceptance when also including b-tagging is calculated from the event selection efficiencies in Table 7.1, but without employing the last step ($N_{jets} \geq 4$). Instead, the efficiency for selecting events with four or five jets is used (since only four and five jet events are considered in the b-tagging method). This efficiency is calculated to be 0.3592 ± 0.0141, resulting in an analysis acceptance of:

$$\mathcal{A}_{btag} = 0.04604 \pm 0.00248 \tag{7.2}$$

with the quoted error being statistical only.

Selection cut	Acceptance (relative to previous cut)
generator (single lepton, $p_T > 5$ MeV)	0.554 ± 0
$\not\!\!E_T > 25$ GeV	0.8782 ± 0.0019
electron cuts	0.1961 ± 0.0071
exactly one electron	0.9659 ± 0.0045
W $m_T > 45$ GeV	0.7705 ± 0.0050
$N_{jets} \geq 4$	0.4079 ± 0.0092

Table 7.1: The acceptance of the event selection (excluding trigger requirements) for MC@NLO $t\bar{t}$ signal events in the semileptonic mode. The errors are statistical only.

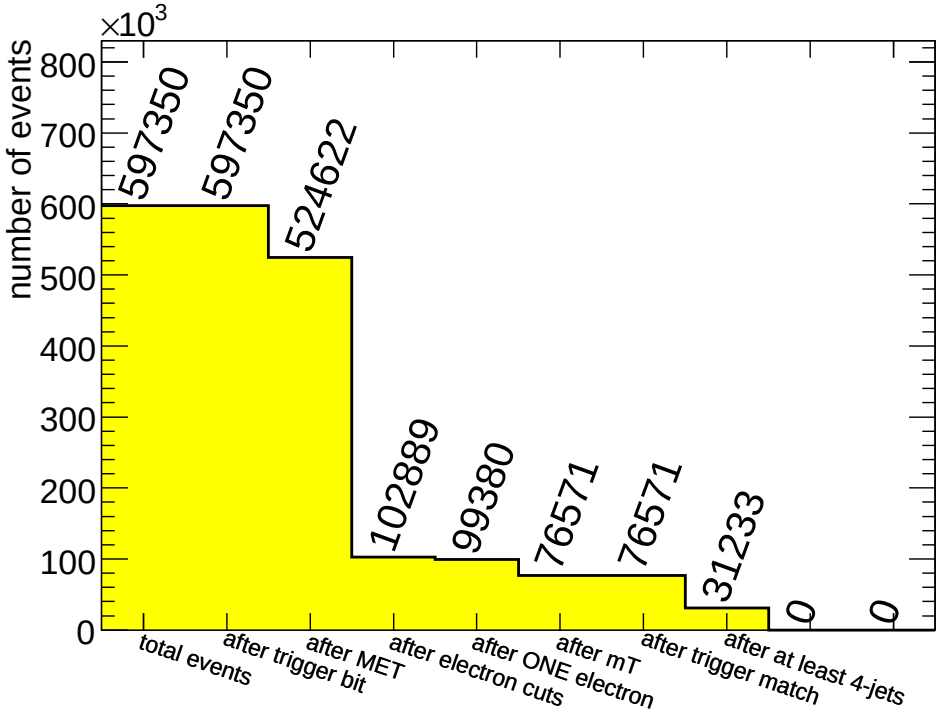


Figure 7.1: Graphical representation of the effect of the selection cuts on the number of events in the MC@NLO $t\bar{t}$ sample for the semileptonic decay mode.

Selection cut	Acceptance (relative to previous cut)
Generator (single lepton, $p_T > 5$ MeV)	0.554 ± 0
electron cuts	1.0000 ± 0
exactly one electron	0.1946 ± 0.0067
muon cuts	1.0000 ± 0
exactly one muon	0.1078 ± 0.0272
$N_{jets} \geq 2$	0.8009 ± 0.0137

Table 7.2: The acceptance of the event selection (excluding trigger requirements) for MC@NLO $t\bar{t}$ signal events in the dilepton mode, with errors being statistical only.

7.2 Acceptance in the Dilepton Mode

Here, the selection efficiencies of the different steps in selecting $t\bar{t}$ signal events are calculated for the dilepton mode. They are summarized in Table 7.2 and can be seen graphically in Figure 7.2. The resulting acceptance for the MC@NLO $t\bar{t}$ sample in the dilepton mode (including the jet multiplicity cut of $N_{jet} \geq 2$) before b-tagging is calculated to be:

$$\mathcal{A} = 0.01680 \pm 0.00429 \tag{7.3}$$

The quoted error is the statistcal error.

Including b-Tagging

The signal acceptance when also including b-tagging is calculated in a similar way as for the semileptonic mode. The calculation starts from the event selection efficiencies in Table 7.2, but without applying the last step ($N_{jets} \geq 2$). Instead, the efficiency for selecting events with two, three or four jets is used: 0.7642 ± 0.0195. This results in an analysis acceptance of:

$$\mathcal{A}_{btag} = 0.01603 \pm 0.00411 \tag{7.4}$$

where again the error is statistcal only.

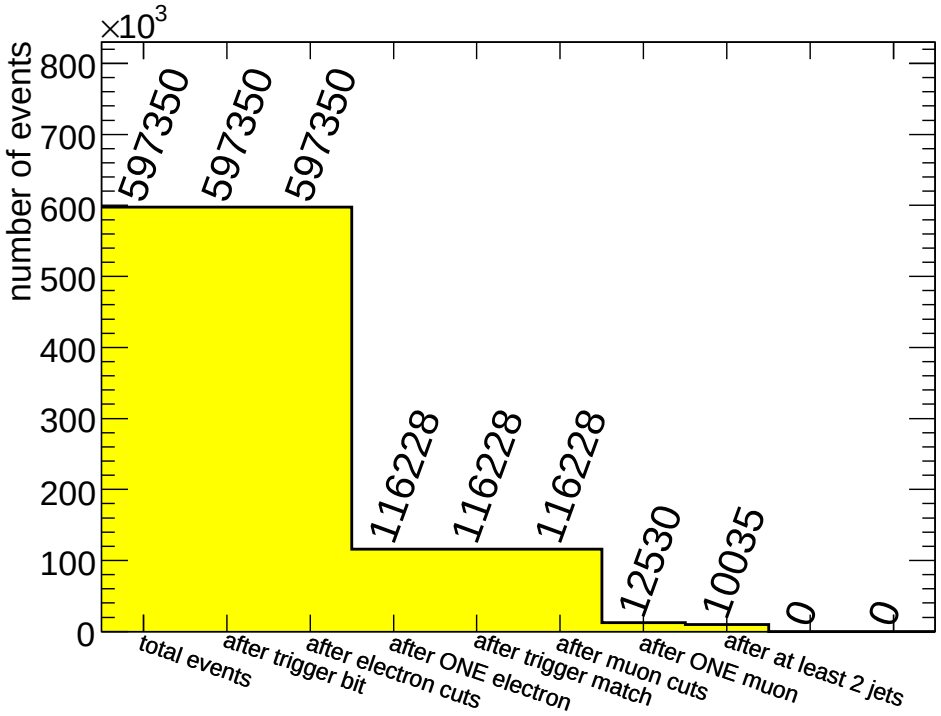


Figure 7.2: Graphical representation of the effect of the selection cuts on the number of events in the MC@NLO $t\bar{t}$ sample for the dilepton mode.

Chapter 8

Background Estimation

In addition to the $t\bar{t}$ pair production mechanism there are many other processes contributing to the selected top sample. These background events need to be carefully studied. There are two different types of backgrounds: Events that contain a real electron passing the electron trigger and events with a "fake" electron, resulting from detector effects.

In real data, jets faking electrons is expected to be a significant background. However, due to the high jet rejection (of order 10^3), simulating a sufficient amount of fake electron events is computationally prohibited. As a result, only a few of these events exist in the stream test data and were shown to be negligible [21].

The background events containing a real electron are, on the other hand, not negligible. To estimate the background, various MC generated samples are used. The backgrounds are divided into single gauge bosons, single top and gauge boson pairs. The MC samples used to estimate the background to the decaying $t\bar{t}$ pair are summarized in Table 8.1 and described in detail in the following sections. The datasets were produced by ATLAS for the computing system commissioning (CSC) effort.

8.1 Single Gauge Bosons

The main contributing background originate from single boson processes with either a W-boson decaying into a lepton plus a neutrino or a Z-boson decaying into two leptons:

$$W \to e\nu \quad W \to \mu\nu \quad W \to \tau\nu$$
$$Z \to ee \quad Z \to \mu\mu \quad Z \to \tau\tau$$

The electroweak samples used here were generated using the PYTHIA MC generator [16] (samples 3a- 3f in Appendix A). All samples were generated with release 11

Sample	MC cross section [pb]	Filter efficiency	$\mathcal{L}$ [pb^{-1}] for 10k events	Input events
$t\bar{t}$	833	0.54	22.2	597350
$W \to e\nu$	17440	0.625	0.917	440050
$W \to \mu\nu$	17440	0.69	0.831	32503
$W \to \tau\nu$	17170	0.198	2.9	158100
$Z \to ee$	1675	0.855	7.0	481050
$Z \to \mu\mu$	1675	0.894	6.7	246596
$Z \to \tau\tau$	1639	0.047	130	176550
Wt	26.7	1	375	48350
s-channel	3.3	1	3030	48300
t-channel	81.3	1	123	44450
WW	70	0.35	408.16	50000
ZZ	11	0.19	4784.7	47800
WZ	27	0.29	1277.1	45900

Table 8.1: Summary of the MC samples used to study the backgrounds to the $t\bar{t}$ decay.

(PYTHIA version 6.323) and simulated with release 12. The total rates and jet multiplicities vary significantly between different MC generators. The uncertainties on the total rates are not relevant for this analysis since the rates are normalized to the stream test data (see below). The uncertainty on the jet multiplicity is, however, of importance and constitute the primary systematic uncertainty. For this thesis, no studies of these MC systematics are conducted. It is simply noted that the PYTHIA generated jet multiplicities should really be regarded as MC estimates and nothing more.

The backgrounds are normalized to the stream test data luminosity using their theoretical cross sections. For the semileptonic mode, most events in the zero and one jet bins originate from single bosons. This is not the case for the dilepton mode where other processes such as dibosons play a substantial role. For the semileptonic mode, only the relative cross sections for W and Z production, which are theoretically well understood, are used. The total rates of the single bosons are thereafter scaled to matched the number of zero and one jet events in the stream test data. The other background estimates are subtracted from the stream test data prior to the single boson scaling.

8.2 Single Top

Top quarks produced through single top channels constitute a significant background to $t\bar{t}$ events. The fraction of single produced to pair produced top quarks, $N_{singletop}/N_{t\bar{t}}$, is on the order 10 % for the semileptonic mode in events with at least four jets. There are

three different processes behind single top production: Wt, t-channel, s-channel. See Section 2.2.1 for details on the mechanisms behind single top production. The single top processes are generated and simulated in release 12 using the AcerMC Monte Carlo generator [37] (samples 4a, 4b and 4c in Appendix A).

8.3 Gauge Boson Pairs

A small yet contributing background process arises from gauge boson pairs, here also referred to as dibosons. There are three different combinations of gauge boson pairs: WW, WZ and ZZ. They are all generated in release 11 using the Herwig Monte Carlo generator [17] and simulated in release 12 (samples 4d, 4e and 4f in Appendix A).

Chapter 9

Results

This chapter summarizes the results from the tagging efficiency calculations and the $t\bar{t}$ production cross section measurements. The $t\bar{t}$ cross section is measured in two channels: The semileptonic and the dilepton channel. For each decay channel, the cross section is determined with and without applying b-jet tagging.

9.1 Semileptonic Mode Analysis

9.1.1 Cross Section Without b-Tagging

This section gives the results from the measurement of the $t\bar{t}$ pair production cross section without using b-tagging. All background processes are normalized to the stream test data luminosity. In addition, as described in Section 8.1, the single bosons are scaled to match the number of zero and one jet events in the stream test data (after removing the background events from single top and dibosons). Table 9.1 shows the number of zero and one jet events and the number of events with at least four jets as measured for the stream test data and as estimated for the different background processes after normalization and scaling. Jet multiplicities for the stream test data and the normalized background components are shown in Figure 9.1.

Using the cross section formula in Equ. 4.2, the $t\bar{t}$ pair production cross section is calculated to be:

$$\sigma_{t\bar{t}} = 740 \pm 57 \text{ pb} \tag{9.1}$$

where the errors quoted are statistical only.

9.1.2 Cross Section Using b-Tagging

This section presents the results from the calculations of the tagging efficiencies and the $t\bar{t}$ production cross section when applying b-tagging. To show the effect of b-tagging on

Physical process	Number of 0+1 jet events	Number of 4+ jet events
Stream test data	61721 ± 248	486 ± 22
Backgrounds		
Wt	4.97 ± 0.20	9.62 ± 0.28
s-channel	1.57 ± 0.04	0.19 ± 0.01
t-channel	36.08 ± 0.99	14.14 ± 0.62
WW	38.25 ± 6.18	0.80 ± 0.89
WZ	0.77 ± 0.88	0.05 ± 0.22
ZZ	9.28 ± 3.05	0.22 ± 0.47
$W \rightarrow e\nu$	60362.00 ± 245.69	128.03 ± 11.31
$W \rightarrow \tau\nu$	1143.46 ± 33.82	3.99 ± 2.00
$Z \rightarrow ee$	75.43 ± 1.93	9.03 ± 0.67
Data excess over all background		319.81 ± 24.82

Table 9.1: Estimation of the number of $t\bar{t}$ events in the stream test data for the semileptonic mode. The quoted errors are statistical only.

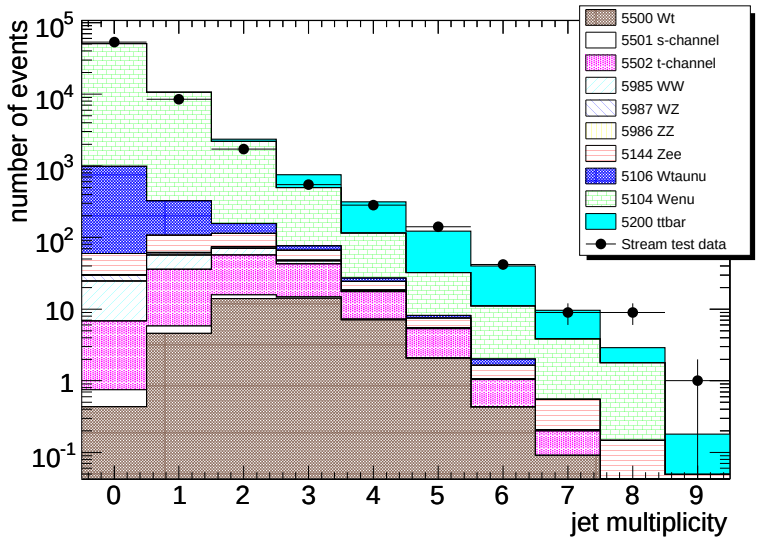
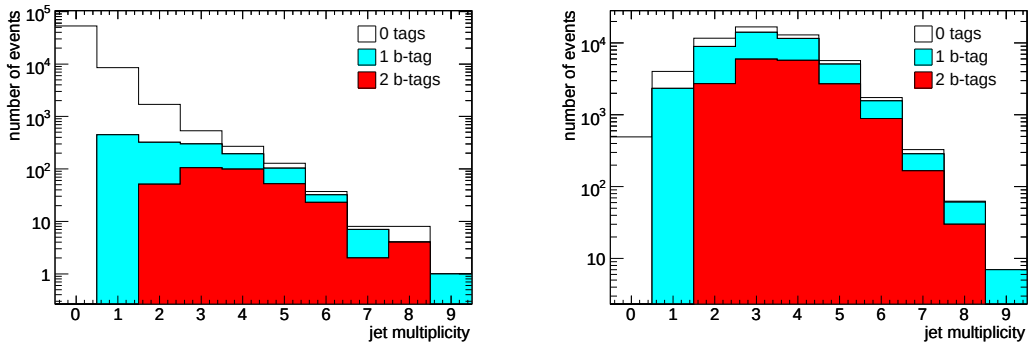


Figure 9.1: Normalized jet multiplicities for stream test data and its different components without b-tagging for semileptonic event candidates.

$t\bar{t}$ jet multiplicities, jet multiplicity distributions are plotted in Figure 9.2 for stream test data and for the MC@NLO $t\bar{t}$ sample, respectively.

All background processes are normalized to the stream test data luminosity also when using b-tagging. Additionally, the single bosons are again scaled to match the number of events in the zero and one jet events in the stream test data (after removing the background events from single top and dibosons but before b-tagging). The composition of data events with at least one or at least two b-jets are shown in Figure 9.3 for the stream test data and the normalized background components.



(a) Stream test data

(b) MC@NLO $t\bar{t}$

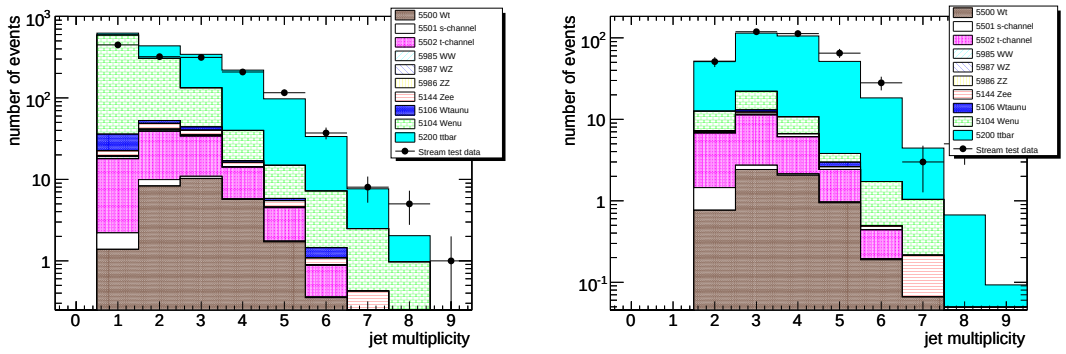
Figure 9.2: Jet multiplicity distributions for events passing the preselection cuts (semileptonic events) with zero, at least one, or at least two b-tagged jets.

To show the effect of applying b-tagging, signal to background ratios and single top suppression are calculated. These are shown in Table 9.2. The $t\bar{t}$ pair production cross section is determined simultaneously with the tagging efficiencies. The minimization is carried out using both the looser and the tighter weight cut values ($w = 3.0$ respective 6.5). The results are shown in Table 9.1.2.

9.2 Dilepton Mode Analysis

9.2.1 Cross Section Without b-Tagging

This section summarizes the results from the measurement of the $t\bar{t}$ pair production cross section in the dilepton mode without b-tagging. All background processes are normalized to the stream test data luminosity. In addition, as described in Section 8.1, the single bosons are scaled to match the number of zero and one jet events in the



(a) Number of b-jets ≥ 1 (b) Number of b-jets ≥ 2

Figure 9.3: Normalized jet multiplicities for stream test data and its different components when requiring at least one (a) or at least two (b) b-jets for the semileptonic mode.

	0 tags	1 tag	2 tags
0+1 jet events			
N_{stream}/N_{bkg}	1.00 ± 0.005	0.76 ± 0.07	-
$N_{t\bar{t}}/N_{singletop}$	1.51 ± 0.02	1.88 ± 0.02	-
≥4 jet events			
N_{stream}/N_{bkg}	2.93 ± 0.02	5.71 ± 0.01	12.29 ± 0.01
$N_{t\bar{t}}/N_{singletop}$	13.50 ± 0.002	14.90 ± 0.002	18.04 ± 0.003

Table 9.2: Signal to background ratios and single top suppression for events with zero and one jets, and for events with four or more jets, for the semileptonic mode. The uncertainties are statistical only.

	w=3.0	w=6.5
ϵ_b (%)	80.4 ± 4.2	71.2 ± 4.1
ϵ_c (%)	34.7 ± 13.0	13.3 ± 7.7
$\sigma_{t\bar{t}}$ (pb)	788 ± 50	798 ± 55

Table 9.3: The $t\bar{t}$ cross section and tagging efficiencies calculated for the semileptonic mode for both the looser and tighter weight cuts. The errors are statistical only.

Physical process	Number of 0+1 jet events	Number of 2+ jet events
Stream test data	48 ± 6.93	100 ± 10
Backgrounds		
Wt	0.26 ± 0.05	0.77 ± 0.08
s-channel	0.05 ± 0.01	0.04 ± 0.01
t-channel	1.32 ± 0.19	1.38 ± 0.19
WW	7.12 ± 0.23	0.48 ± 0.06
WZ	0.09 ± 0.01	0.05 ± 0.01
ZZ	0.75 ± 0.04	0.14 ± 0.02
$W \to e\nu$	13.58 ± 3.69	1.47 ± 1.21
$W \to \mu\nu$	5.48 ± 5.48	0 ± 0
$W \to \tau\nu$	1.29 ± 0.65	0.32 ± 0.32
$Z \to ee$	1.62 ± 0.27	0.75 ± 0.18
$Z \to \mu\mu$	9.95 ± 0.94	0.90 ± 0.28
$Z \to \tau\tau$	27.08 ± 0.42	2.79 ± 0.13
Data excess over all background		90.92 ± 10.09

Table 9.4: Estimation of the number of $t\bar{t}$ events in the stream test data for the dilepton mode, where the errors are statistical only.

stream test data (after removing the background events from single top and dibosons). Table 9.4 shows the number of zero and one jet events and the number of events with at least two jets as measured for the stream test data and as estimated for the different background processes after normalization and scaling. The jet multiplicities for the stream test data and the normalized background components are shown in Figure 9.4.

Using the cross section formula in Equ. 4.2, the $t\bar{t}$ pair production cross section for the dilepton mode is calculated to be:

$$\sigma_{t\bar{t}} = 655 \pm 73 \text{ pb} \tag{9.2}$$

with errors being statistical only.

9.2.2 Cross Section Using b-Tagging

The results from the calculations of the tagging efficiencies and the $t\bar{t}$ production cross section when applying b-tagging are here given. First, to show the effect of b-tagging on $t\bar{t}$ jet multiplicities, jet multiplicity distributions are plotted in Figure 9.5 for stream test data and for the MC@NLO $t\bar{t}$ sample, respectively.

All background processes are normalized to the stream test data luminosity. Contrary to the semileptonic mode, the single bosons are no longer responsible for the

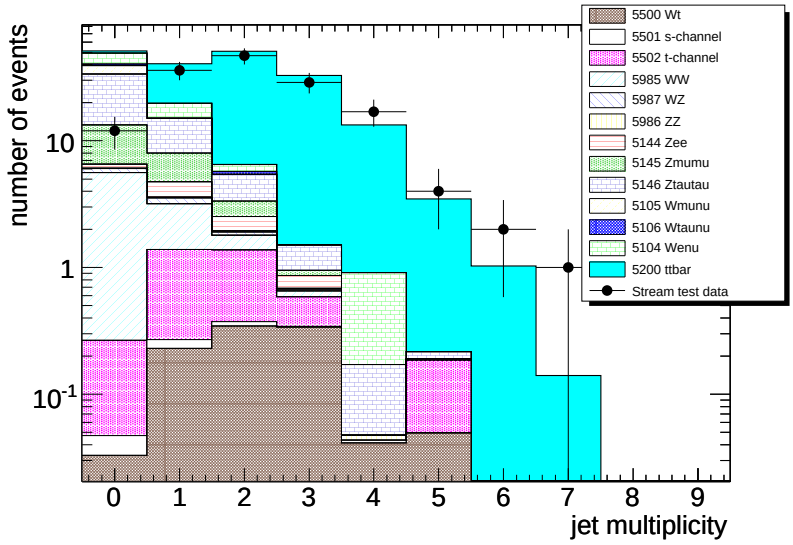


Figure 9.4: Normalized jet multiplicities for stream test data and its different components without b-tagging for fully leptonic event candidates.

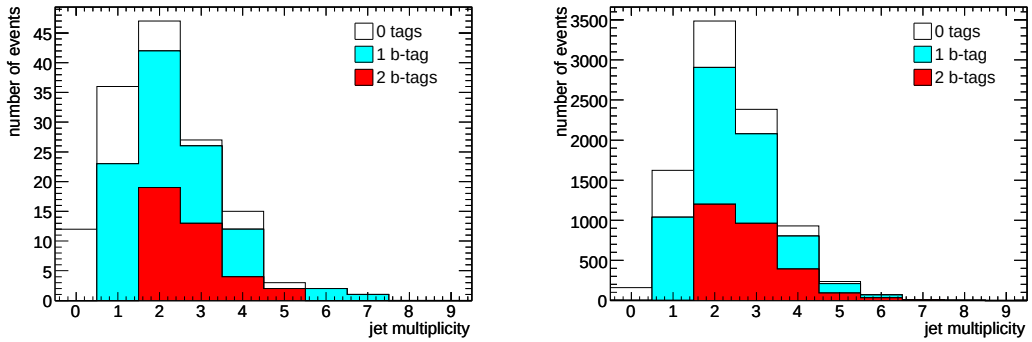
majority of the events in the zero and one jet bins. Therefore, no scaling is applied to the single bosons for the dilepton mode. The composition of data events with at least one or at least two b-jets are shown in Figure 9.6 for the stream test data and the normalized background components.

To show the effect of employing b-tagging for the dilepton mode, signal to background ratios and single top suppression are calculated. These are listed in Table 9.5. The values can be compared to corresponding numbers for the semileptonic mode (listed in Table 9.2). As can be seen, the dilepton mode results in significantly higher signal to background ratios and also the single top suppression is higher.

The $t\bar{t}$ pair production cross section in the dilepton mode is determined simultaneously with the tagging efficiencies. The minimization is carried out using both the looser and the tighter weight cut values ($w = 3.0$ and $w = 6.5$). For the dilepton mode, only ϵ_b and $\sigma_{t\bar{t}}$ are allowed to float. For ϵ_c, the value from the semileptonic mode minimization is used. The results are shown in Table 9.6.

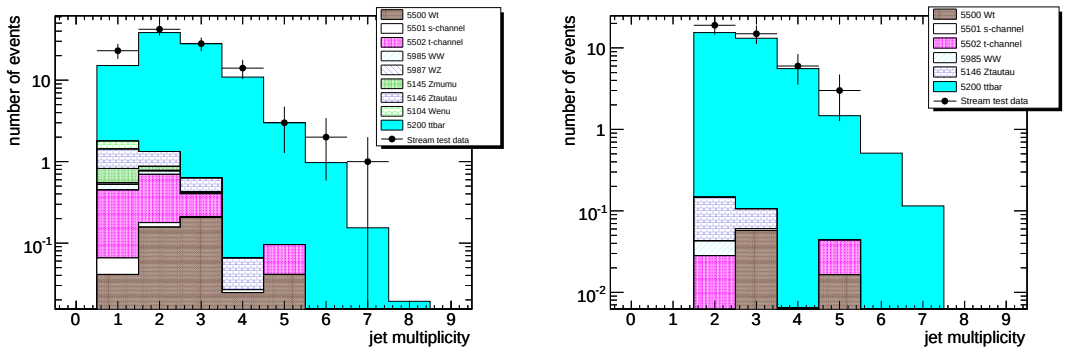
9.2.3 Systematic Uncertainties

Various systematic uncertainties affect the cross section results. These ought to be analyzed for completeness of the analysis, however, a detailed study of the systematics lies beyond the scope of this thesis. Here, a brief discussion of effects contributing to the systematic uncertainties is given.



(a) Stream test data (b) MC@NLO $t\bar{t}$

Figure 9.5: Jet multiplicities for events in the dilepton mode without b-tagging and with at least one or at least two b-tagged jets.



(a) Number of b-jets ≥ 1 (b) Number of b-jets ≥ 2

Figure 9.6: Normalized jet multiplicities for stream test data and its different components when requiring at least one (a) or at least two (b) b-jets in the dilepton mode.

	0 tag	1 tag	2 tag
0+1 jet events			
N_{stream}/N_{bkg}	0.69 ± 0.24	12.44 ± 0.02	-
$N_{t\bar{t}}/N_{singletop}$	13.77 ± 0.009	29.67 ± 0.008	-
2+ jet events			
N_{stream}/N_{bkg}	10.94 ± 0.01	37.56 ± 0.004	123.43 ± 0.002
$N_{t\bar{t}}/N_{singletop}$	42.08 ± 0.002	64.96 ± 0.002	271.70 ± 0.001

Table 9.5: Signal to background ratios and single top suppression for events with zero and one jets, and for events with two or more jets, for the dilepton mode. The uncertainties are statistical only.

	w=3.0	w=6.5
ϵ_b (%)	76.7 ± 6.7	61.7 ± 8.1
$\sigma_{t\bar{t}}$ (pb)	704 ± 86	733 ± 108

Table 9.6: The $t\bar{t}$ cross section and tagging efficiencies calculated for the dilepton mode for both the looser and tighter weight cuts. The errors are statistical only.

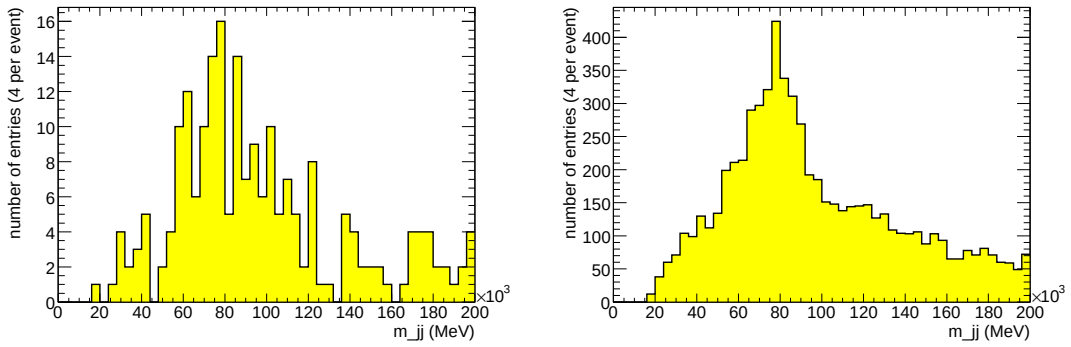
One of the most important systematics for the analysis without b-tagging is the jet energy scale. Uncertainties on ISR/FSR affect the signal aceptance. Other systematics arise from the electromagnetic and the missing transverse energy scales although these are smaller. Another small systematic uncertainty is the one associated with the trigger efficiency, resulting from the difference in the software releases used for the stream test data. Important for the cross section measurements are the uncertainties on backgrounds, in particular on the single bosons. The single boson jet multiplicites vary significantly between different Monte Carlo generators which results in a significant contribution to the systematic uncertainty [21].

Additional systematic uncertainties arise when applying b-tagging. The jet fractions used for the measurements are determined usíng truth information from imperfect Monte Carlo generated samples, which gives rise to systematics. Uncertainties also exist on the definitions of jet flavors. In addition to the uncertainty on the jet energy scale, also the b-jet energy scale contributes to the systematics when using b-tagging. Systematics are also associated with the mistag rate, although these were shown to be small for the cross section value and the b-tagging efficiency.

9.3 Reconstructing the Top Mass

9.3.1 Invariant W-Boson Mass

The invariant mass of the hadronically decaying W-boson and corresponding reconstructed top mass are determined to validate the $t\bar{t}$ event selection for the semileptonic mode. Here, events with four or five jets of which two are tagged as b-jets are selected. For events with four jets, the hadronic W-boson mass is calculated from the two untagged jets. For events with five jets there are three possible combinations of the hadronic W-boson mass and the mass is here recontructed with three entries per event. The distribution of the invariant mass is shown in Figure 9.7 for stream test data and MC@NLO $t\bar{t}$, respectively. The figure shows consistency between the stream test data $t\bar{t}$ selected sample and the MC@NLO $t\bar{t}$ sample. The stream test data is, however, statistics limited.



(a) Stream test data

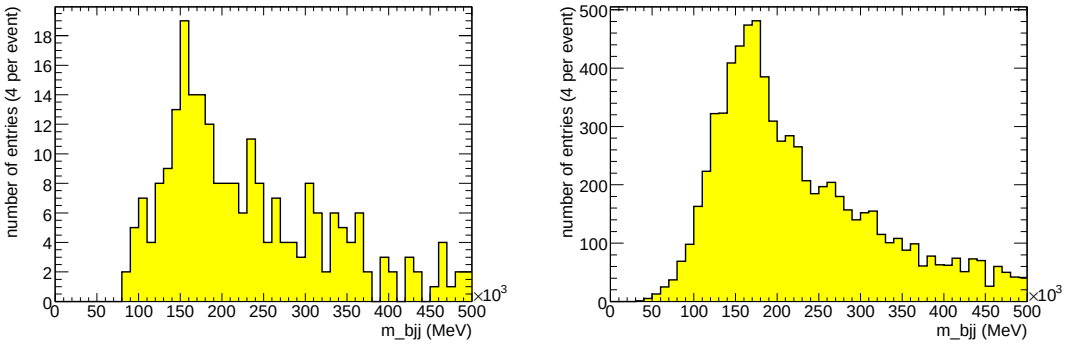
(b) MC@NLO $t\bar{t}$

Figure 9.7: Invariant mass of the hadronic W-boson for events with four or five jets of which two are b-tagged. There are three entries per event for events with five jets.

9.3.2 Top Quark Mass

The reconstructed top mass is determined by choosing the three-jet combination of the dijets constituting the hadronic W-boson together with the b-jet resulting in highest total transverse momentum. As for the W-boson invariant mass, there are one entry per event for events with four jets and three entries per event for five-jet events. The distribution of the reconstructed top mass is shown in Figure 9.8 for stream test data and MC@NLO $t\bar{t}$. The two again appear to be consistent, although the stream test

data is statistics limited. The invariant mass and reconstructed top mass distributions can be used for internal jet calibrations. Since statistics are low for the stream test data, one alternative is to also include events with one b-tagged jet, although this would also increase the number of background events.



(a) Stream test data

(b) MC@NLO $t\bar{t}$

Figure 9.8: Reconstructed top mass for events with four or five jets of which two are b-tagged. For each hadronic W-boson dijet combination, the three-jet combination (two untagged jets plus one b-jet) resulting in highest sum p_T is chosen.

Chapter 10

Conclusions

The analysis in this thesis measures the $t\bar{t}$ production cross section $\sigma_{t\bar{t}}$ for the semileptonic mode and for the dilepton mode. The ATLAS stream test data constitutes the data sample and various Monte Carlo samples are used to estimate the backgrounds. The analysis in the semileptonic mode considers events in the electron plus jets channel whereas the dilepton mode analysis studies events in the electron-muon channel.

The $t\bar{t}$ cross section is measured separately with and without applying b-tagging. The measurements without b-tagging uses an event counting method, where events with at least two or four jets are considered for the dilepton mode and the semileptonic mode, respectively. The analysis using b-tagging counts the number of tagged jets in two-, three- and four-jet events for the dilepton mode and in four- and five-jet events for the semileptonic mode. The method allows for a simultaneous measurement of the cross section and the b-tagging efficiency, assuming one knows the mistag rate. When employing b-tagging, two different weight cut values are used: A loose cut value of 3.0 and a tighter cut value of 6.5. The different cross section results are summarized in Table 10.1.

Measuring the cross section separately in the two channels, with and without using b-tagging, allows for comparisons and consistency checks. As can be seen in Table 10.1, the cross section is somewhat higher for the semileptonic mode than the dilepton mode. The cross section results within the two modes with and without using b-tagging are all consistent and give similar results. The cross section measurements have different

Decay mode	$\sigma_{t\bar{t}}$ without b-tagging	$\sigma_{t\bar{t}}$ with b-tagging (w=3.0)	$\sigma_{t\bar{t}}$ with b-tagging (w=6.5)
Semileptonic	740 ± 57 pb	788 ± 50 pb	798 ± 55 pb
Dilepton	655 ± 73 pb	704 ± 86 pb	733 ± 108 pb

Table 10.1: Summarized results for the $t\bar{t}$ production cross section.

systematics, which makes these type of comparisons even more valuable.

Theoretically, the $t\bar{t}$ cross section is predicted to be 833 pb at ATLAS. That cross section value is, however, not necessarily a "correct" value for the data used in this analysis. On the contrary, the stream test data is known to only account for a limited number of physics processes and it is despite its similarities with real data, still Monte Carlo generated data.

The dilepton mode measurements in particular are additionally somewhat restricted by the limited statistics of the stream test data. This will not be an issue with real data. Despite limitied statistics, it has, however, been shown that ATLAS has great prospects for detailed $t\bar{t}$ studies already with early data.

Appendix A

Monte Carlo Samples

A variety of different Monte Carlo samples were used in the analysis. The datasets and their provenance are listed below.

1. "Data" sample

   (a) `streamtest.004*.inclEle.merge.AOD.v12000605`

   recon `streamtest.004*.inclEle.recon.(AOD|ESD).v12000604`

   This data was created from an admixture of RDOs, which included different generators (even for the same process).

2. Samples for $t\bar{t}$ signal

   (a) MC@NLO : `trig1_misal1_mc12.005200.T1_McAtNlo_Jimmy.recon.AOD.v12000604`

   evgen 12.0.4.1, filtered for 5 MeV lepton (efficiency = 0.554).

   digit 12.0.6.1

3. Samples for electroweak backgrounds

   (a) $W \to e\nu$ PYTHIA v11: `trig1_misal1_csc11.005104.PythiaWenu.recon.AOD.v12000601`

   evgen 11.0.42.9, filtered for a 10 GeV lepton (efficiency 0.625)

   digit 12.0.31.7, 1 mm range cut (corrected in analysis [38]).

   (b) $Z \to ee$: `trig1_misal1_mc12.005144.PythiaZee.recon.AOD.v12000604`

   evgen 12.0.6.4, filtered for a 10 GeV lepton (efficiency 0.855)

   digit 12.0.6.4

   (c) $W \to \tau\nu$: `trig1_misal1_csc11.005106.PythiaWtaunu.recon.AOD.v12000601`

   evgen 11.0.5.13, filtered for a 5 MeV lepton (efficiency 0.198)

   digit 12.0.31.6, 1 mm range cut (corrected in analysis [38]).

(d) $W \to \mu\nu$: `trig1_misal1_csc11.005105.PythiaWmunu.recon.AOD.v12000601`
evgen 11.0.42.2, filtered for a 5 MeV lepton (efficiency 0.69)
digit 12.0.31.3

(e) $Z \to \mu\mu$: `trig1_misal1_csc11_V1.005145.PythiaZmumu.recon.AOD.v12000601`
evgen 11.0.42.17, filtered for a 5 GeV lepton with $|\eta| < 2.8$ (efficiency 0.894)
digit 12.0.31.6, 1 mm range cut (corrected in analysis [38]).

(f) $Z \to \tau\tau$: `trig1_misal1_csc11.005146.PythiaZtautau.recon.AOD.v12000601`
evgen 11.0.5.13, filtered for a 5 GeV lepton with $|\eta| < 2.8$ (efficiency 0.047)
digit 12.0.31.6, 1 mm range cut (corrected in analysis [38]).

4. Samples for single top and dibosons

(a) t-channel : `trig1_misal1_mc12.005502.AcerMC_tchan.merge.AOD.v12000605`
evgen 12.0.4.2
digit 12.0.31.8
recon 12.0.6.1

(b) Wt: `trig1_misal1_mc12.005500.AcerMC_Wt.merge.AOD.v12000605`
evgen 12.0.4.2
digit 12.0.31.8
recon 12.0.6.1

(c) s-channel : `trig1_misal1_mc12.005501.AcerMC_schan.merge.AOD.v12000605`
evgen 12.0.4.2
digit 12.0.31.8
recon 12.0.6.1

(d) WW: `trig1_misal1_csc11.005985.WW_Herwig.merge.AOD.v12000605`
evgen 11.0.5.11, filtered for a 10 GeV lepton (efficiency 0.35)
digit 12.0.31.7
recon 12.0.6.1

(e) WZ: `trig1_misal1_csc11.005987.WZ_Herwig.merge.AOD.v12000605`
evgen 11.0.5.11, filtered for a 10 GeV lepton (efficiency 0.29)
digit 12.0.31.7
recon 12.0.6.1

(f) ZZ: `trig1_misal1_csc11.005986.ZZ_Herwig.merge.AOD.v12000605`
evgen 11.0.5.11, filtered for a 10 GeV lepton (efficiency 0.19)
digit 12.0.31.7
recon 12.0.6.1

Appendix B

Summary in Swedish

Elementarpartikelfysiken studerar universums allra minsta beståndsdelar såväl som dess största strukturer, allt för att söka svar på de fundamentala frågorna kring materiens uppbyggnad och funktion. Den idag bästa beskrivningen av elementarpartiklarna och de krafter som verkar dem emellan är den så kallade standardmodellen. Elementarpartiklarna delas i standardmodellen in i tre grupper: Kvarkar, leptoner och kraftbärare. Kvarkarna, sex till antalet, är de partiklar som bygger upp hadronerna (däribland protonen och neutronen). Leptonerna, också de sex till antalet, inkluderar exempelvis elektronen. Kvarkar och leptoner delas in i tre familjer, där varje familj består av en elektriskt laddad lepton (elektron e, myon μ, tauon τ), en elektriskt neutral lepton (neutriner ν_e, ν_μ, ν_τ), en kvark av upp-typ (upp u, charm c, topp t) samt en kvark av ner-typ (ner d, sär s, botten b).

Kraftbärarna, även kallade bosoner, förmedlar de fundamental krafterna. Det finns fyra fundamentala krafter varav tre är inkluderade i standardmodellen. Beskrivna i standardmodellen är den elektromagnetiska kraften som verkar mellan elektriskt laddade partiklar och bland annat binder elektroner till den laddade kärnan i atomen, den svaga kraften som styr radioaktiva sönderfall, samt den starka kraften som håller samman protoner och neutroner i atomkärnan. Den elektromagnetiska kraften förmedlas av fotonen, den svaga kraften av W- och Z-bosonerna och den starka kraften av gluoner. Utöver dessa tre krafter finns även gravitationskraften som inte är förenlig med standardmodellen (vilket dessutom är en av den i övrigt så väl verifierade standardmodellens svagheter).

En av kvarkarna i standardmodellen är toppkvarken. Det är den tyngsta kvarken, med en massa nästan lika stor som en guldatomkärna. Toppkvarken upptäcktes så sent som 1995 av forskare vid Tevatron-acceleratorn i USA. Den är därför fortfarande av stort experimentellt intresse, både för studier av dess egna egenskaper men också då den är en viktig bakgrundsprocess till nya analyser.

Utanför Genéve på gränsen mellan Frankrike och Schweiz ligger det internationella partikelfysiklaboratoriet CERN. På CERN håller en ny partikelaccelerator på att färdig-

ställas som beräknas starta under hösten 2008, LHC (Large Hadron Collider). LHC kommer att accelerera och kollidera protoner vid en högre energi än vad som någonsin tidigare kunnat ske. En av detektorerna som ska studera kollisionerna vid LHC är den 46 meter långa och 25 meter höga ATLAS detektorn.

Det här examensarbetet studerar toppkvarken och dess egenskaper. Arbetet är en del av en förstudie vid ATLAS experimentet. Då acceleratorn LHC ännu inte startats görs studierna och beräkningarna på Monte Carlo simulerad data, som har precis samma utseende som "riktig" data. Vid LHC kommer toppkvarkar främst att produceras i par och det huvudsakliga syftet med detta arbete är att bestämma tvärsnittet för produktionen av toppkvarkspar. Tvärsnittet är i princip ett mått på sannolikheten att ett toppkvarkspar ska bildas.

För att kunna bestämma tvärsnittet måste toppkvarkarna som bildas identifieras i kollisionerna samt de olika bakgrundsprocesserna uppskattas. En toppkvark sönderfaller nästan uteslutande till en W-boson och en b-kvark. W-bosonen kan därpå sönderfalla till antingen en lepton och en neutrino eller till två kvarkar. I den här analysen studeras två olika sönderfallskanaler. I den första kanalen sönderfaller en av de två bildade W-bosonerna till en elektron och en neutrino medan den andra W-bosonen sönderfaller till två kvarkar. I den andra kanalen som studeras sönderfaller en av W-bosonerna till ett elektron-neutrino par och den andra till ett myon-neutrino par. Vidare bestäms tvärsnittet både med och utan att även identifiera b-kvarkarna från de producerade toppkvarkarna. Att identifiera b-kvarkarna är ett effektivt hjälpmedel för att reducera bakgrundshändelser men det kräver istället noga studier av effektiviteten med vilken b-kvarksidentifikationen sker. Avslutningsvis görs även en enkel uppskattning av toppkvarkens massa för att verifiera att det faktiskt är toppkvarkar som studeras.

Bibliography

[1] B.R. Martin and G. Shaw, *Particle Physics*, 2nd Edition, John Wiley & Sons (1992).

[2] W.N. Cottingham and D.A. Greenwood, *An Introduction to the Standard Model of Particle Physics*, Cambridge University Press (1998).

[3] W.-M. Yao et al., *Review of Particle Physics*, Journal of Physics, G 33, 1 (2006).

[4] ATLAS Collaboration, *Detector and Physics Performance, Technical Design Report Volume I*, CERN/LHCC/99-14 (1999).

[5] ATLAS Collaboration, *Detector and Physics Performance, Technical Design Report Volume II*, CERN/LHCC/99-15 (1999).

[6] CDF Collaboration, F. Abe et al., *Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. 74, 2626-2631 (1995);
D0 Collaboration, S. Abachi et al., *Observation of the Top Quark*, Phys. Rev. Lett. 74, 2632-2637 (1995).

[7] ATLAS-TWiki, *CSC Notes List, Recommendations for the Notes*, https://twiki.cern.ch/twiki/bin/view/Atlas/CSCNotesList#Recommendations_for_the_Notes

[8] W. Verkerke and I. V. Vulpen, *Commissioning Atlas Using Top-Quark Pair Production*, ATL-COM-PHYS-2007-023 (2007).

[9] M. Mohammadi Najatabadi, *Single Top Production at LHC*, arXiv:hep-ex/0605034v2 (2006).

[10] CERN Geneva, *LHC: The Guide*, CERN-Brochure-2008-001-Eng (2008).

[11] CERN - European Organization for Nuclear Research: http://www.cern.ch

[12] The ATLAS Experiment: http://www.atlas.ch

[13] ATLAS Collaboration, *Inner Detector, Technical Design Report, Volume I*, ATLAS TDR 4, CERN/LHCC/97-16 (1997).

[14] ATLAS Collaboration, *Calorimeter Performace, Technical Design Report*, CERN/LHCC/96-40 (1997).

[15] ATLAS Collaboration, *ATLAS Computing, Technical Design Report*, ATLAS-TDR-017, CERN-LHCC-2005-022 (2005).

[16] T. Sjostrand, L. Lönnblad, S. Mrenna and P. Skands, *PYTHIA 6.3 Physics and Manual*, arXiv:hep-ph/0308153 (2003).

[17] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour and B.R. Webber, *HERWIG 6.5*, arXiv: [hep-ph/0011363]; hep-ph/0210213, JHEP 0101 (2001) 010.

[18] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, arXiv: hep-ph/0204244, JHEP 0206 (2002) 029;
S. Frixione, P. Nason and B.R. Webber, *Matching NLO QCD and Parton Showers in Heavy Flavour Production*, arXiv: hep-ph/0305252, JHEP 0308 (2003) 007.

[19] ATLAS-TWiki, *The ATLAS Computing Workbook*,
https://twiki.cern.ch/twiki/bin/view/Atlas/WorkBook

[20] ATLAS TWiki, *Streaming Test For Users*,
https://twiki.cern.ch/twiki/bin/view/Atlas/StreamingTestForUsers

[21] A. Bach, A. Gaponenko, I. Hinchcliffe, A. Holloway-Arce, P. Loscutoff, M. Shapiro, L. Skinnari, L. Tompkins and J. Virzi, *A Pre-Commissioning $t\bar{t}$ Cross Section Measurement at ATLAS*, ATL-PHYS-INT-2008-019 (2008).

[22] The ATLAS Collaboration, *Flavour Tagging Calibration with $t\bar{t}$ Events in ATLAS*, ATL-PHYS-2008-000, Draft 4.0, (2008),
https://twiki.cern.ch/twiki/pub/Atlas/AbstractTop/ttbarBtagCSCnote_draft4.pdf

[23] Data Streaming Group, *Data Streaming in ATLAS, Report of the Data Streaming Group*, ATL-GEN-INT-2007-002 (2007).

[24] ATLAS TWiki, *Electron Reconstruction*,
https://twiki.cern.ch/twiki/bin/view/Atlas/ElectronReconstruction

[25] ATLAS TWiki, *EGamma AOD*,
https://twiki.cern.ch/twiki/bin/view/Atlas/EgammaAOD

[26] ATLAS TWiki, *Muon Reco Pedia*, https://twiki.cern.ch/twiki/bin/view/Atlas/MuonRecoPedia

[27] ATLAS TWiki, *Jet Analysis*, https://twiki.cern.ch/twiki/bin/view/Atlas/JetAnalysis

[28] ATLAS TWiki, *Jet Algorithms*, https://twiki.cern.ch/twiki/bin/view/Atlas/JetAlgorithms

[29] ATLAS TWiki, *Physics Analysis Workbook AOD Objects Rel 12*, https://twiki.cern.ch/twiki/bin/view/Atlas/PhysicsAnalysisWorkBookAODObjectsRel12

[30] ATLAS TWiki, *Introduction to Hadronic Calibration*, https://twiki.cern.ch/twiki/bin/view/Atlas/IntroductionToHadronicCalibration

[31] ATLAS-TWiki, *b-Tagging Basics*, https://twiki.cern.ch/twiki/bin/view/Atlas/BTaggingBasics

[32] *Online b-Tagging Selection for the ATLAS Experiment at the LHC*, DAQ-CONF-2005-007.

[33] S. Corréard, V. Kostioukhine, J. Levêque, A. Rozanov and J.B. de Vivie, *b-Tagging with DC1 Data*, ATL-PHYS-2004-006.

[34] R. Hawkings, *The ATLAS Inner Detector and Flavour Tagging*, Eur Phys J C 34, s01, s109-s116 (2004).

[35] The ROOT webpage: http://root.cern.ch

[36] F. James, *Minuit, Function Minimization and Error Analysis, Reference Manual*, http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html, CERN.

[37] B. P. Kersevan and E. Richter-Was, *The Monte Carlo Event Generator AcerMC 2.0 with Interfaces to Pythia 6.2 and Herwig 6.5*, arXiv: hep-ph/0405247 (2004).

[38] ATLAS TWiki, *News For Physics Users*, https://twiki.cern.ch/twiki/bin/view/Atlas/NewsForPhysicsUsers